Robot Deictics: How Gesture and Context Shape Referential Communication

Allison Sauppé and Bilge Mutlu
Department of Computer Sciences, University of Wisconsin–Madison
1210 West Dayton Street, Madison, WI 53706 USA
asauppe@cs.wisc.edu, bilge@cs.wisc.edu

ABSTRACT

As robots collaborate with humans in increasingly diverse environments, they will need to effectively refer to objects of joint interest and adapt their references to various physical, environmental, and task conditions. Humans use a broad range of *deictic gestures*—gestures that direct attention to collocated objects, persons, or spaces—that include pointing, touching, and exhibiting to help their listeners understand their references. These gestures offer varying levels of support under different conditions, making some gestures more or less suitable for different settings. While these gestures offer a rich space for designing communicative behaviors for robots, a better understanding of how different deictic gestures affect communication under different conditions is critical for achieving effective human-robot interaction. In this paper, we seek to build such an understanding by implementing six deictic gestures on a humanlike robot and evaluating their communicative effectiveness in six diverse settings that represent physical, environmental, and task conditions under which robots are expected to employ deictic communication. Our results show that gestures which come into physical contact with the object offer the highest overall communicative accuracy and that specific settings benefit from the use of particular types of gestures. Our results highlight the rich design space for deictic gestures and inform how robots might adapt their gestures to the specific physical, environmental, and task conditions.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems—*human factors, software psychology*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*evaluation/ methodology, user-centered design*

General Terms

Design, Human Factors

1. INTRODUCTION

As robots begin to assist humans in increasingly diverse environments and tasks, communication will become one challenge robots will face in working with their human partners. Uncontrollable and unpredictable environmental effects, such as noise, lighting, and visual obstructions, will require robots to adapt their gestures to effectively refer to objects of joint interest despite these distractions [19]. Even in an environment where such distractions are absent, gestures are still desirable channels of communication, as they can be used to augment or replace complex verbal descriptions [9, 20].

HRI'14, March 3–6, 2014, Bielefeld, Germany.

ACM 978-1-4503-2658-2/14/03 ...$15.00.
http://dx.doi.org/10.1145/2559636.2559657.

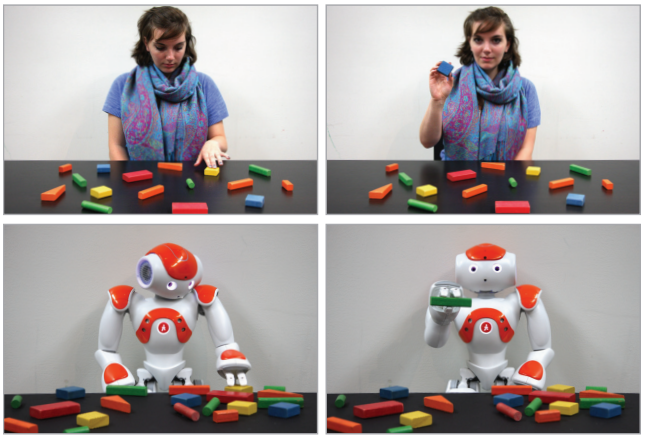

Figure 1: Examples of human deictic gestures *touching* and *exhibiting* and their equivalent implementations on the NAO robot.

Non-verbal behaviors provide valuable information when combined with partially articulated speech, enabling humans to use a combination of pronouns and gestures to communicate information [8]. Those gestures which rely on knowledge of the environment to provide context for their interpretation—such as pointing, presenting, and exhibiting—are referred to as *deictic gestures* [8]. While deictic gestures are often referred to as "pointing gestures", this term encompasses a larger range of hand gestures beyond pointing [7]. For example, in infant-caretaker relations, humans might use touching instead of pointing to more concretely refer an object to an infant [17]. Similarly, instructors might hold up a piece they are asking their student to find [7]. Clark's work highlighted a number of deictic gestures used in human interaction and how they might be used in conjunction with speech to achieve communicative goals [7]. Not only do each of these gestures take a form other than pointing, but each is chosen to accommodate the setting—a particular environmental context that could affect verbal communication—and communicative goals of the participants [7, 17].

Robot behavior designers have recognized the importance of deictic gestures for robots, implementing robots that point while giving directions [24], achieving common spatial ground [5], and directing the sorting of items [25]. While pointing is a common deictic gesture that is implemented on robots [4, 28], a model encompassing the full range of deictic gestures available for robots has yet to be

developed. Even in human interaction literature, whether gestures are chosen based on a particular setting and why they might be preferred to pointing is suggested in prior work for some gestures [7, 17], but is unknown across many gestures. An understanding of which deictic gestures are best suited for a given setting will allow robots to become more effective communicators not only in their mimicking of human choices of gesture, but also in completing or replacing utterances when bringing attention to objects of joint interest [9, 20], which could be useful in particular settings.

To further understand this design space for robot deictics, we designed a study of human deictic gestures in a variety of settings. Using a combination of gestures and settings collected from literature, we implemented six gestures—pointing, presenting, exhibiting, touching, grouping, and sweeping—for six settings on a NAO robot, which were contextualized in a wooden block identification task. Examples of human deictic gestures and their implementation on a robot can be seen in Figure 1. For each gesture-setting combination, the participant identified the blocks referred to and evaluated the robot's gesture in regards to how humanlike and effective it was. From our results, we provide recommendations for gestures for each setting, as well as speculate on which properties of gestures can help predict their effectiveness in helping the listener identify the object.

2. BACKGROUND

Humans frequently employ deictic gestures during tasks to augment or replace their speech, particularly when their current setting requires carefully choosing a gesture to use. We review prior work on human deictic gestures and why specific gestures might be used in a particular setting. Additionally, we discuss current work concerning the implementation of deictic gestures on robots.

2.1. Human Deictic Gestures

Deictic gestures are often used to augment or replace verbal descriptions of the object being referred to, also called the *referent* [16]. The importance of deictic gestures in communication is shown in pre-verbal children, who will use deictic gestures as a way of communicating with their caretakers prior to their ability to form utterances to describe their wants and needs [6]. Once humans have the ability to verbally communicate, the use of deictic gestures increases and becomes more nuanced, serving to support more complex utterances [13]. At this point, deictic gestures are used to decrease cognitive burden, allowing for complex verbal descriptors to be eliminated in favor of a deictic gesture toward the referent [9]. The replacement of fully articulated speech with a combination of partially articulated speech and deictic gestures reduces cognitive load for the speaker, by requiring less processing to form an utterance, and the listener, by requiring less processing to interpret the utterance. For example, a speaker might replace the description of an object with "this" and a deictic gesture that indicates the referent. Gestures might even fully replace utterances in settings such as a noisy factory environment [10].

Traditionally thought of as "pointing gestures", deictic gestures are comprised of a more diverse set of gestures that are used to draw attention to an object. Caretakers often use touch to more concretely focus the attention of infants on an object [17], while students or instructors in an instructional block building task might hold up a piece to implicitly confirm whether it's correct [7]. Prior research has demonstrated that the speaker's use of gestures affects information recall and rapport in listeners [4, 12].

Work by Clark demonstrated that deictic gestures are much broader than pointing [7]. However, while the use of these additional deictic gestures has been mentioned in relation to other research [1, 3, 7, 17], a more thorough understanding of the breadth of deictic gestures and why they are chosen for particular settings is needed.

2.2. Robots and Deictic Gestures

Prior work in human-robot interaction recognizes the need for robots to gesture naturally in order to communicate in a more humanlike fashion. Much of this work has focused on enabling robots to use deictic gestures to enhance task outcomes, such as the robot's use of gestures improving user performance in manipulation tasks [4, 24, 25]. In general, robots use deictic gestures similarly to humans to help bring attention to objects of joint interest and achieve common spatial ground [5]. When the environment may make using deictic gestures difficult or impossible, robots are also able to use perspective-taking to ensure that their deictic gestures are used in ways that are interpretable by the listener [29]. To ensure gestures are used appropriately, research has focused on enabling robots to use pointing gestures in socially appropriate ways [18].

Deictics are often thought of as referring to an object, but can also be used to refer to a region of space, such as the opposite end of a room. Prior work in robot deictics has shown that referring to a region of space—which is often more difficult to verbalize than an object—results in only a marginally worse accuracy rate than referring to an object [27]. Robots are able to use visual differences in spaces in combination with deictics to help listeners identify the correct region in the space [11]. St. Clair et al. demonstrated that a robot using a combination of deictic gestures and gaze to refer to a space resulted in higher accuracy than using just one modality [27]. Additionally, how the gesture was implemented and executed was significant, with human pointing using a bent arm producing significantly worse accuracy than pointing with a straight arm.

Although HRI research has successfully implemented human pointing behaviors in numerous applications, there is still much to understand about what other deictic gestures robot behavior designers should consider using, what properties are most effective, and how the particular setting should shape gesture choice.

3. UNDERSTANDING DEICTIC GESTURES AND SETTINGS

Understanding both the types of deictic gestures available to robots and the settings in which they may be appropriate is necessary for a comprehensive study of the interplay between setting, communicative goals, and gesture choice. In this section, we describe six gestures and six settings that robots are envisioned to encounter and that might effect gesture choice.

3.1. Deictics

We describe six deictic gestures that we focus on in this work: pointing, presenting, touching, exhibiting, sweeping, and grouping. These gestures combine results from prior work to inform our understanding. Examples of each gesture can be seen in Figure 2.

Pointing – Pointing is often considered the prototypical deictic gesture, being universally understood across cultures [15], ages [23], and even species [21]. A pointing gesture uses an extended index finger with the hand rotated so that the palm faces toward or perpendicular to the ground to direct attention. The hand does not come into physical contact with the referent. A pointing referent may be a single object, a region of space, or no specific object or region [20]. Prior work has already explored implementing human pointing gestures on robots, revealing that pointing gestures which point away from the body, rather than across the body, are more accurate at communicating the referent [27].

Presenting – Presenting uses a similar style to pointing in that the speaker gestures toward the referent without coming into contact. However, where as pointing leaves only the index finger extended, presenting extends all fingers and points the palm of the hand upwards. This gesture is often interpreted as inviting, encouraging

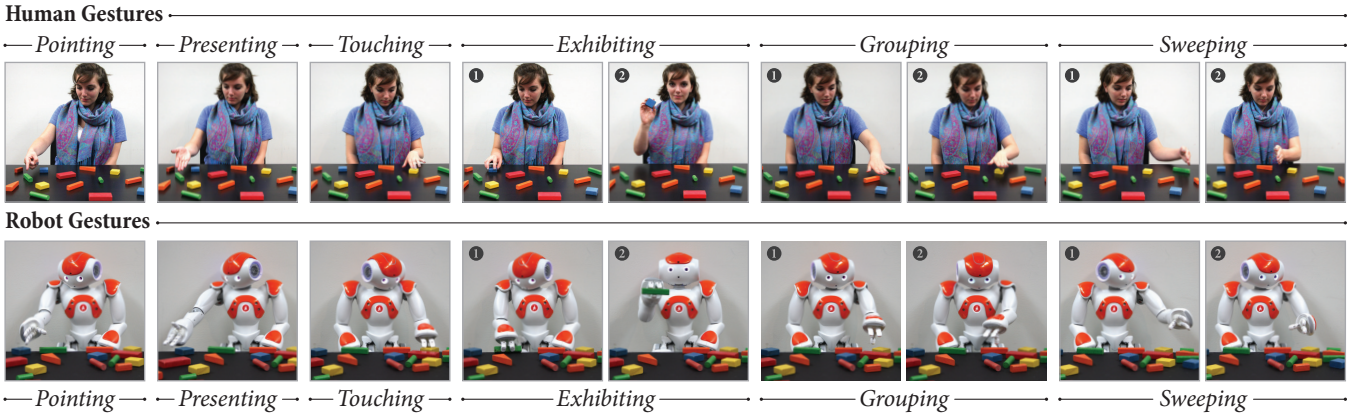

Figure 2: Instances of a human performer and the NAO robot demonstrating the deictic gestures studied in this work.

the listener to, for example, pick up the referent [3, 14]. Presenting gestures are also used by speakers to indicate that they are ready to receive an object that they previously requested [3].

Touching – Touching is used in similar settings as pointing and presenting; however, touching removes ambiguity in that the speaker's hand comes into direct physical contact with the referent. This absence of ambiguity makes touching ideal in situations where verbal communication is absent or impaired. For example, touching seems to be a preferred deictic gesture for mothers communicating with non-verbal infants, since verbal capabilities are required for understanding pointing gestures [17]. Touching may also be a preferred deictic gesture in a factory or on a noisy shop floor. Touching may also be used to refer to an object in situations where constructing an accurate verbal description to augment a pointing gesture is difficult. For example, it may be difficult to verbally differentiate between or provide an accurate pointing gesture for similar objects in close proximity to one another. Additionally, if certain properties of an object cannot be described concisely, touching the object may be more cost-effective than attempting to describe the object.

Exhibiting – Exhibiting is a natural extension of touching where the object is grasped and lifted so that it can be observed by others [7]. This gesture might be used when joint attention is obstructed due to referent location, making other gestures unusable. For example, objects concealed by other objects may prevent the listener from seeing the object, requiring an exhibiting gesture.

Grouping – Grouping offers a gesture similar to presenting in that the fingers are extended with the palm facing down. Instead of referring to a single block, however, grouping takes advantage of the larger area covered by the hand to reference those objects located underneath the hand. The speaker may use a circular hand motion—still in the grouping gesture—around the area they wish to indicate in cases where an area instead of an object is the referent. This gesture has also been implemented in interactive tabletop and wall touchscreen displays to highlight a group of objects [30].

Sweeping – Similar to grouping, sweeping references one or more objects in a given area. A speaker utilizing sweeping will place their hand, with fingers extended, perpendicular to and above the surface to indicate a beginning boundary for referenced objects. The gesture then sweeps across additional referenced objects [1].

3.2. Settings

While humans employ a variety of deictic gestures to direct attention to an object, each gesture has unique functional properties that might diminish its effectiveness in some settings. As robots start working alongside humans, it is expected that they will encounter similar settings as humans. In this section, we describe six settings that we believe can impact which gestures a robot should choose.

Distance from Referrer – The accuracy of a gesture may diminish when the distance between referrer and referent is larger, as listeners may make greater interpolations regarding where the speaker's hand is gesturing. In extremes, objects are located immediately in front of or substantially far away (e.g., the opposite end of a table) from the referrer. While some robots might be capable of reaching locations that would not be available to humans, there will always be situations where the robot will need to reference a distant referent.

Clustered Objects – Varied amounts of space exist between objects laid out on a table. This can vary from objects clustered together very closely to objects spread far apart. This setting also mimics the possibility of having one versus many objects, with one object effectively obtained by spreading objects far apart.

Noise – Many environments in which robots are expected to work, such as warehouses and assembly lines, can be noisy. Since deictic gestures are often accompanied by speech that can elaborate on the purpose of the gesture, noise might make some gestures more difficult to understand.

No Visibility – Often times, objects to which the robot wishes to draw attention may be in the referrer's line of sight but may not be visible to the listener. For example, objects may be located in a container or behind a structure or object. In these cases, deictic gesturing may indicate to the listener that some object in the general area of the gesture is located outside their line of sight.

Ambiguity – During assembly tasks, pieces which initially look similar may differ in small ways, such as screws that have slightly different lengths and widths. These pieces may be difficult to differentiate verbally due to these subtle differences. Lack of adequate vocabulary may also hinder verbal differentiation and may also place significant cognitive burden on both the speaker and the listener.

Neutral – Those cases where there may not be any environmental factors affecting communication results in a neutral setting. Here, a diverse set of objects is nearby the referent with ample space between each object and in clear view of all involved parties.

These settings serve as a representative sample of the situations robots are expected to encounter, particularly in joint tasks with humans, making them appropriate contextualizations for better understanding how the affects of gestures change across settings.

4. IMPLEMENTATION

We chose to contextualize our implementation in an object referencing task, where the robot would refer to one or more wooden building blocks distributed on a workspace. The use of wooden building blocks in this task was inspired by Shah et al. [26]. In the context of our wooden building block task, we created two workspaces of blocks to accommodate our six settings and designed each of the gestures in every setting.

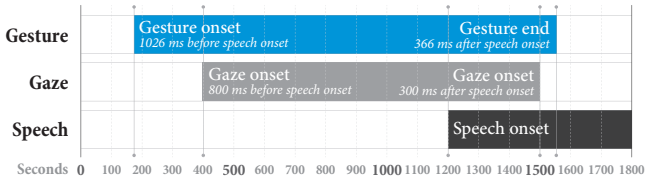

Figure 3: A model of the gesture-contingent gaze behavior implemented in our study. Start and end times are relative to the onset of speech.

4.1. Gesture Design

We implemented our behaviors on the NAO robot. The NAO features six degrees of freedom in each arm: shoulder pitch, should roll, elbow yaw, elbow roll, wrist yaw, and finger pitch. The technical capabilities of the NAO enabled us to create accurate reproductions of each gesture in a variety of settings. We implemented the gestures on the NAO through *puppeteering*, a technique in which a designer manually guides the robot in executing a gesture while a motion capture program saves joint positions at each keyframe. These keyframes are later used to generate arm-motion trajectories. We puppeteered each gesture and manually edited the resulting motion profile as necessary in Choregraphe, a behavior authoring environment for the NAO. The gesture profiles were then saved on the robot to be executed by our experiment software.

As gaze is an integral part of a natural gesture, we implemented gesture-contingent gaze behavior as described in Huang & Mutlu [12] for all of our gestures (see Figure 3). The gaze trajectory followed the robot's hand as it gestured to the block for all gestures.

4.2. Workspace Design

We designed a layout of wooden blocks for six settings that we divided onto two workspaces, which can be seen in Figure 4. Each workspace contained two sets of blocks, with one set on the left half of the workspace, and the second set on the right half. The first workspace displayed the ambiguity and no visibility settings, and the second workspace displayed the neutral, distant from referrer, clustered objects, and noise settings. The following are the descriptions of each setting from the robot's point of view:

- *Neutral*: An assortment of blocks arranged near the referrer. Blocks were spaced 1.5 to 2 in. (3.8 cm to 5.1 cm) from nearby blocks.
- *Distance from Referrer*: An assortment of blocks arranged far from the referent. Blocks were spaced 1.5 to 2 in. (3.8 cm to 5.1 cm) apart, and all blocks were at least 6.5 in. (16.5 cm) from the referrer.
- *Clustered Objects*: An assortment of blocks near the referrer. Blocks were spaced .5 in. (1.3 cm) from nearby blocks.
- *Noise*: Identical to the neutral setting, but with white noise of people talking loudly played from a nearby speaker.
- *No Visibility*: An assortment of blocks placed behind a 3.5 in. (8.9 cm) partition.
- *Ambiguity*: Blocks which were similar in color, length and shape were arranged near the referrer. Blocks were spaced 1.5 to 2 in. (3.8 cm to 5.1 cm) from nearby blocks.

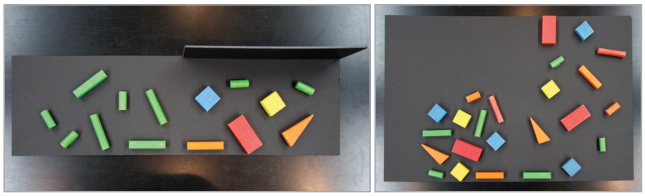

Figure 4: The two workspaces used to represent the six settings we explored. The left workspace displays the ambiguity (left) and the no visibility (right) settings. The right workspace displays the clustered objects setting (left), the distant from referrer setting (top right), and the neutral setting (bottom right), which was used for both the neutral and noise settings.

5. EVALUATION

To explore the effectiveness of gestures in different settings, we used the workspaces from our wooden blocks task to conduct a within-participants study of all gesture-setting combinations. For each condition, participants identified the blocks they believed the robot was referring to and rated the gesture on a number of items. Our results indicate that setting has an impact on gesture.

5.1. Study Design

To better understand the effects of gesture choice and setting on referential communication, we designed a within-participants study to explore every feasible combination of the gesture and setting factors described previously. In addition to the six gestures mentioned, we included two verbal-only baselines in our gesture factor. The first baseline involves the use of only *minimally articulated* verbal references that the robot uses in conjunction with gestures, such as "this block." This baseline follows results from Shah et al. [26], which showed that participants in a collaborative block building task often used generic referential statements in conjunction with gestures to bring attention to a block. The second baseline involves the use of *fully articulated*, descriptive speech to provide a complete description of the block to which the robot is referring, such as "the short green cylinder closest to you." Whereas the first baseline shows the outcome of eliminating gestures, this second baseline demonstrates the consequences of the robot engaging only in verbal communication. Since we designed gaze cues specifically for the gesture they accompanied, we eliminated gaze from our baselines.

In total, the combination of eight forms of communication (six gestures and two verbal baselines) and six settings resulted in 48 conditions. We eliminated two conditions—touching and exhibiting blocks at a distance—due to the physical impossibility of contact with these blocks, resulting in 46 conditions used in the study.

5.2. Task

Participants were asked to observe the robot as it referenced blocks situated on a table between the participant and robot. The participant observed 46 rounds of references made by the robot, where each round was one of the 46 conditions previously outlined. Rounds were broken down into two sets to allow for all of the settings to be displayed. The first set consisted of 30 rounds (neutral, distant from referrer, clustered objects, and noise settings) and the second of 16 rounds (no visibility and ambiguity settings). The order in which participants observed the two sets was balanced across participants, while the rounds within each set were randomly ordered. Additionally, to account for possible participant biases between the left and right arms, workspaces were flipped along the vertical axis for half of the participants. All possible presentations of the workspace were gender balanced. After the robot completed the action for a given round, the participant rated the robot's behavior on both objective and subjective scales on a one page questionnaire.

The experimental setup is shown in Figure 5. Participants were seated 2.5 feet (76.2 cm) away from the robot, with the workspace between them. The participant's questionnaires were placed between the participant and the workspace. A small speaker was placed next to the robot (out of view of the participant) to emit background noise of people talking during any conditions which involved noise.

5.3. Procedure

Following informed consent, participants were seated in the experiment room. The experimenter explained the task to the participant, started the robot, and left the room. The robot initiated the interaction by giving an introduction, followed by starting the first round. During each round, the robot would choose one of the gesture and setting combinations to exhibit according to which set of settings

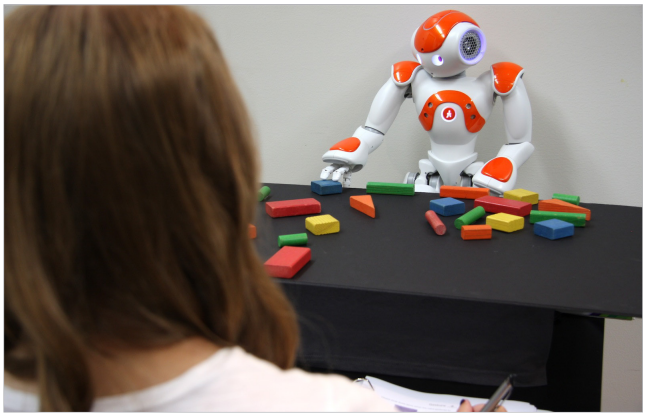

Figure 5: A participant evaluating the robot touching the blue block.

was currently available. The robot would then perform the action associated with the particular combination. Upon completion of the robot's action, the participant would complete the questionnaire.

The top half of the questionnaire showed a picture of the workspace currently being used, where the participant would individually circle all blocks they believed the robot had referred to. On the bottom half of the questionnaire were six seven-point rating items to measure the naturalness of the robot's gesture in that setting. When the participant was satisfied with their answers to the questionnaire, they would say "next" to advance the robot to the next round.

When the first workspace was completed, the experimenter would set up the second workspace and provide new questionnaires that reflected the new workspace layout. At the completion of both workspaces, participants were compensated $5 for their time. Participants took between 15 and 32 minutes to complete the task (M =23 minutes, 40 seconds, SD =3 minutes, 52.8 seconds).

5.4. Participants

We recruited 24 native English speakers (12 males, 12 females) with diverse majors and occupations and ages that ranged 18–46 ($M = 22.7$, $SD = 5.92$) from the University of Wisconsin–Madison.

5.5. Measures & Analysis

For each condition, participants completed a questionnaire in which they identified the blocks they believed the robot was referencing and answered six rating-scale questions on the robot's behavior. Participants identified blocks by circling referenced blocks on a picture of the workspace that was included on the questionnaire. As a measure of *accuracy*, we classified participant's identification of the blocks as either correct or incorrect based on whether the participant's answer exactly matched the blocks that the robot's gesture indicated, considering answers that were a superset or subset of the correct answer to be incorrect. Our subjective measures assessed the perceived qualities of the gesture in the given setting. From the six questions asked, we constructed the following two scales (half of the items were reversed to prevent response sets):

Perceived Effectiveness (Cronbach's $\alpha = .967$)
1. The robot used this gesture effectively.
2. The robot's gesture helped me to identify the object(s).
3. The robot's gesture was appropriate for the context.
4. The robot's gesture was easy to understand.

Naturalness (Cronbach's $\alpha = .790$)
1. The robot's gesture was humanlike.
2. The robot's gesture was fluid.

Data analysis involved a two-way analysis of variance (ANOVA), including gesture and setting as fixed effects. Tukey's honestly significant difference (HSD) test was used for pairwise comparisons.

5.6. Results

We discuss our most significant results below, first discussing gestures across all settings and then highlighting comparisons of gestures within each setting. Due to the large number of pairwise comparisons involved in our analysis, only Omnibus test results are reported in the paragraphs below, and pairwise comparisons are illustrated in Figures 6, 7, and 8.

5.6.1. Comparison of Gestures

A comparison of gestures across settings showed that gesture type had a significant effect on accuracy, $F(7,1073) = 112.06$, $p < .001$ (Figure 6). The fully descriptive baseline was significantly less accurate than exhibiting and pointing, but significantly more accurate than sweeping and grouping. Exhibiting, touching, presenting, and pointing were all significantly more accurate than sweeping and grouping. Consistent with the results on accuracy, gesture type had a significant effect on the perceived effectiveness of the gesture, $F(7,1073) = 134.37$, $p < .001$. Exhibiting and touching were perceived as significantly more effective than the fully articulate baseline and the presenting, pointing, sweeping, and grouping gestures. All gestures were found to be fairly natural, with average ratings between 5.5 and 6.5 out of 7.

5.6.2. Comparison of Gestures by Setting

The following presents results for gestures within each setting. Pairwise comparisons for measures of accuracy and perceived effectiveness are illustrated in Figures 7 and 8, respectively.

Neutral – Gesture type had a significant effect on accuracy in the neutral setting, $F(7,161) = 34.36$, $p < .001$. The fully articulate baseline, as well as the exhibiting, touching, presenting, pointing gestures, were all significantly more accurate in communicating the referent than sweeping and grouping were. Gesture type also had a significant affect on perceived effectiveness, $F(7,161) = 39.48$, $p < .001$. The fully articulate baseline and the exhibiting, touching, and presenting gestures were perceived as significantly more effective than pointing, sweeping, and grouping.

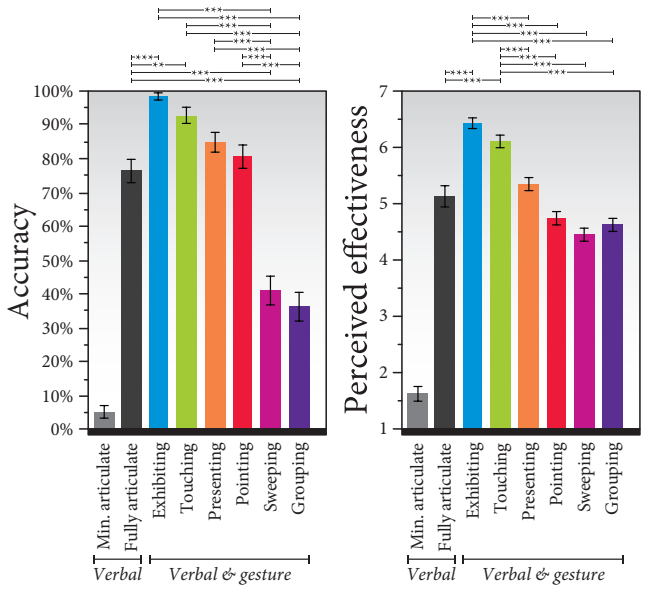

Figure 6: Results for both the accuracy of each gesture and the perceived effectiveness of each gesture across all settings. (***), (**) denotes $p < .001$, $p < .01$, respectively. Exhibiting and touching gestures were more accurate than the two baselines and the sweeping and grouping gestures and were perceived to be more effective than the two baselines and the other gestures.

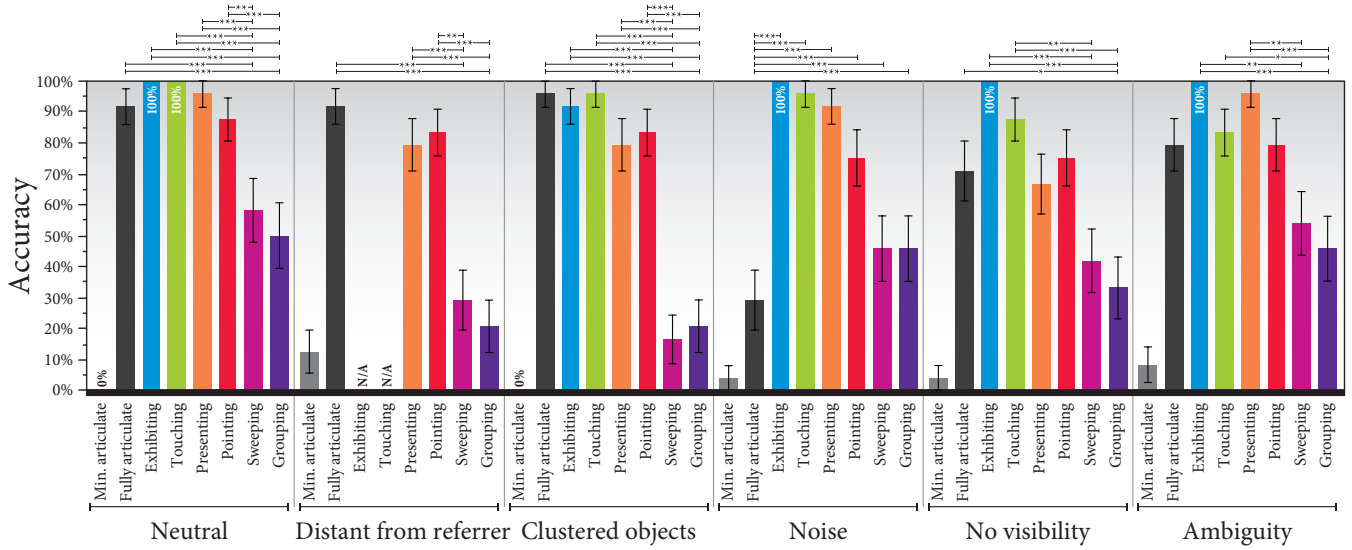

Figure 7: Results for the communicative accuracy of each gesture, displayed by setting. (***), (**), (*) denotes $p < .001$, $p < .01$, and $p < .05$, respectively. Exhibiting and touching were consistently more accurate than sweeping and grouping across the majority of settings.

Distant From Referrer – When referents were distant, gesture type had a significant effect on accuracy, $F(5,115) = 21.73$, $p < .001$. The fully articulate baseline and the presenting and pointing gestures were all significantly more accurate than sweeping and grouping. Additionally, while the effectiveness of presenting and pointing fell by 16% and 5%, respectively, when compared to the neutral setting, sweeping and grouping observed greater losses of effectiveness at 27% and 30%, respectively. Gesture type also had a significant effect on perceived effectiveness in this setting, $F(5,115) = 19.83$, $p < .001$. The fully articulate baseline was perceived as significantly more effective than presenting, sweeping, and grouping.

Clustered Objects – Gesture type had a significant effect on accuracy, $F(7,161) = 43.26$, $p < .001$, when objects were clustered. The fully articulate baseline and the presenting and pointing gestures were all significantly more accurate than sweeping and grouping. Exhibiting, touching, presenting, and pointing were all slightly less accurate than in the neutral setting, losing 5% to 15% accuracy. Sweeping and grouping saw larger drops compared to the neutral setting, losing 48% and 30% accuracy respectively. Perceived effectiveness was significantly affected by gesture type, $F(7,161) = 25.86$, $p < .001$. Exhibiting and touching both were perceived as significantly more effective than presenting, sweeping, and grouping.

Noise – Gesture type also had a significant effect on accuracy in the noise setting, $F(7,161) = 22.49$, $p < .001$. The fully articulate baseline was significantly more accurate than exhibiting, touching, presenting and pointing. Exhibiting, touching, and presenting were all significantly more accurate than pointing, sweeping, and grouping. Gesture type also had a significant effect on perceived effectiveness, $F(7,161) = 38.54$, $p < .001$. The fully articulate baseline was perceived as significantly less effective than every gesture. Additionally, exhibiting and touching were perceived as significantly more effective than pointing, sweeping, and grouping.

No Visibility – Gesture type had a significant effect on accuracy in the no visibility setting, $F(7,161) = 15.86$, $p < .001$. Exhibiting and touching were both significantly more accurate than sweeping and grouping. Additionally, exhibiting was the only level to not experience a drop in accuracy compared to the neutral setting. Gestures type had a significant impact on perceived effectiveness,

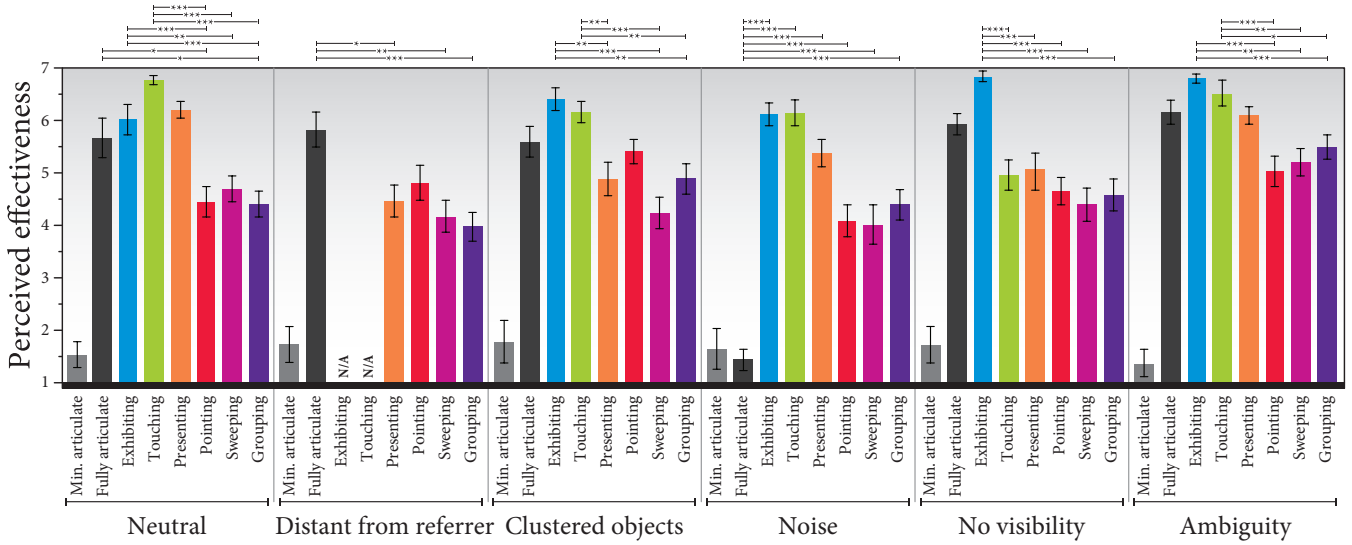

Figure 8: Results for the perceived effectiveness of each gesture, displayed by setting. (***), (**), (*) denote $p < .001$, $p < .01$, and $p < .05$, respectively. Exhibiting and touching were consistently perceived to be more effective than presenting, pointing, sweeping, and grouping across the majority of the settings.

$F(7,161) = 40.18$, $p < .001$. Exhibiting was perceived as significantly more effective than all other gestures. Additionally, the fully descriptive baseline was perceived as significantly more effective than pointing, sweeping, and grouping.

Ambiguity – Gesture type had a significant effect on accuracy under ambiguity, $F(7,161) = 15.52$, $p < .001$. Exhibiting and presenting were significantly more accurate than sweeping and grouping. Gesture type also significantly affected perceived effectiveness, $F(7,161) = 59.40$, $p < .001$. Exhibiting and touching were rated as significantly more effective than pointing, sweeping, and grouping.

6. DISCUSSION

Our results offer implications for designing effective deictic gestures for robots. The paragraphs below summarize the most important results and discusses these implications.

6.1. Properties of Effective Gestures

When the gestures were compared without consideration of context against the fully articulated baseline, our six gestures organized into three groupings: referencing one object with physical contact, referencing one object without physical contact, and referencing multiple objects, with these groupings doing significantly better than, equivalent to, or significantly worse than the fully articulated baseline, respectively. Gestures that involved physical contact with the objects (exhibiting and touching) provided the most effective communication, rarely causing the participant to choose the incorrect block; the effectiveness of physical touch was confirmed in many of our settings as well. This finding confirms prior research on human deictics that report that mothers choose to use physical touch to identify an object for their pre-verbal child due to the unambiguous nature of the gesture [17]. This behavior follows a failed attempt at pointing by the mother, leading her to try a more concrete gesture with a higher likelihood of success. These physical gestures alleviate the cognitive burden placed on those interpreting the gesture by eliminating uncertainty in the referent. Our findings suggest that, in settings or tasks that require precise identification of objects, physical contact with the object provides the best chance of the listener correctly identifying the referent.

6.2. Setting and Gesture Accuracy

Our study highlights instances in which the setting significantly affects the accuracy of gestures. Below, we discuss these results and their implications for designing robot behaviors.

Noise and Verbal Descriptions – Our fully articulated baseline had comparable performance to exhibiting, touching, pointing, and presenting in all levels except noise. Our findings showed that the fully articulated baseline performed much worse in the noise condition than in every other condition, while the accuracy of many of the gestures remained unchanged. This finding supports the use of gestures that come into contact with the object when fully articulated utterances are difficult to form.

Although the most prominent effect of the noise level was seen with the fully articulated baseline, even pointing and presenting were less effective than in the neutral setting, despite their perceived lack of reliance on utterances given the minimal information utterances included when coupled with gestures. While each type of gesture maintained a similar motion profile across conditions, making it easier for participants to learn what the gesture was communicating across repeated viewings, it may be that the simple utterances that the robot used helped the participant to distinguish whether the robot was referring to one or many objects. For example, although sweeping and pointing gestures appear similar when the arm is extended, they follow different trajectories, pointing aiming toward a specific object and sweeping covering an area. In addition, the pointing gesture is accompanied by the phrase "this block," clearly indicating that only one block is being referenced, while the sweeping gesture is accompanied by the phrase "these blocks," suggesting more than one block is being referenced. With noise obscuring these phrases, participants may have doubted how many blocks they should select.

While our study did not look at the interaction between gesture and the complexity of the utterance, the combination of our findings on the fully articulate baseline and the pointing and presenting gestures in the noise setting suggest that the robot attempting to clarify non-physical gestures with verbal descriptions—a common human behavior [8]—would not improve the accuracy of the gesture.

Gestures for Obstructed Objects – In the no visibility level, where some blocks were obscured by a partition, the number of correctly identified objects was significantly lower for many gestures compared to our fully articulated baseline. Only the exhibiting gesture maintained its effectiveness, since the robot grasped the object and exhibited it above the partition for the participant to clearly see. In real-world applications, the robot can take the perspective of its user [29] to determine whether a referent is obstructed and whether it is necessary to exhibit it for the listener.

Diminishing Effectiveness of Multiple Object Gestures – The gestures that referred to multiple objects were consistently less effective than other gestures and the fully articulated baseline. This result is likely a product of the greater ambiguity inherent in these gestures. Such ambiguity occasionally led participants to include objects in the set that were not intended to be referenced. Because our objective measure only counted the answers that were a perfect match to the intended blocks as correct, participants' answers which were a superset or subset of the intended blocks were incorrect. The robot might correct the participant's understanding of which blocks should be included by engaging in repair, such as providing clarifications, which may be considered less costly than precisely identifying the correct blocks the first time. Expecting the listener to process and react to many objects they are expected to identify and manipulate would likely result in high cognitive load, leading to greater frustration with the robot [22]. In the cases where identifying only the correct objects is imperative, the robot might use a gesture intended for one object multiple times.

In the *distant from referrer* and *clustered objects* settings, the low accuracy of gestures is compounded. We found that gestures referring to multiple objects were significantly less effective than gestures for a single object in both of these settings and required more precision than other settings did. Prior work provides support for this observation; when objects are distant, humans use pointing gestures to indicate a spatial region rather than a specific object, instead relying on speech to convey the object [2]. Likewise, when the robot finds itself working with objects that may be difficult for the listener to disambiguate, either due to distance from the robot or from other objects, the robot should rely on a combination of gesture accompanied with speech to help identify the referents.

Consistent Accuracy of Exhibiting and Touching – In five of the settings, exhibiting and touching maintained relatively consistent accuracy, almost always outperforming the remaining gestures and baselines. Only in the distant from referrer level were exhibiting and touching outperformed, and even then, this result was due to the physical impossibility of these gestures in that setting. These results seem to support the use of exhibiting and touching over pointing and presenting. However, while exhibiting and touching more consistently supported accurate identification of the referent, these gestures are more costly to execute, requiring the robot to physically lift and/or to relocate to be within physical touch of the object, which places limits on their use in real-world applications.

6.3. Limitations and Future Work

While we chose gestures and settings for our exploration based on a projection of what might best serve the design of robot behavior, there are other gestures and settings to explore. We did not explore how language, an integral part of deixis, influences the effectiveness of gestures in these settings. Preliminary work has explored some of the issues regarding the influence of language on gestures [19], but a more comprehensive exploration of this area is needed.

The choice of the robot might also have an impact on how gestures are interpreted, as robot platforms vary in their ability to reproduce human gestures. Through iterative design, we sought to design the robot's gestures to mimic human gestures as closely as possible in terms of the intended communication. While we chose to implement the gestures used in this study through puppeteering, a Wizard-of-Oz technique, we hope to develop a robust gesture synthesis system that enables robots to autonomously generate accurate deictic gestures.

In the study, the robot's references included only one gesture. However, communicating complex ideas might require the use of a sequence of gestures. Although we expect our findings to generalize to independent evaluations of each gesture in a sequence, further examination is necessary to conclusively understand how gestures used in a sequence affect referential communication.

7. CONCLUSION

Human collaborations involve the use of deictic gestures, allowing speakers to direct their collaborators' attention to objects in the environment while reducing cognitive load for themselves and their listeners. To function as competent collaborators, robots will need to use deictic gestures to effectively direct the attention of their users toward objects of joint interest. Drawing on literature on human communication, we designed a set of deictic gestures for the NAO robot and specified a set of settings that provided a diverse set of conditions in which humans and robots might engage in deictic communication. We conducted an exploratory human-robot interaction study that examined the communicative accuracy and perceived effectiveness of these gestures against two verbal baselines and how the setting of the communication affected these measures. The results suggest many design implications regarding the use of gestures in each setting, including what properties might make certain gestures more effective, the tradeoffs involved in referring to multiple objects, the effects of noise on verbal deictic references, and how gestures might function when objects are obscured. Our findings offer new research directions into deictic communication and design implications for human-robot collaboration.

8. ACKNOWLEDGMENTS

This research was supported by National Science Foundation award 1149970 and National Institutes of Health award 1R01HD071089-01A1. We thank Brandi Hefty, Jilana Boston, Jason Sauppé, Lindsay Jacobs, and Catherine Steffel for their help with our work.

9. REFERENCES

[1] M. W. Alibali, L. M. Flevares, and S. Goldin-Meadow. Assessing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of Educational Psychology*, 89(1):183, 1997.

[2] A. Bangerter. Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, 15(6):415–419, 2004.

[3] D. Bolinger. Intonation and gesture. *American speech*, 58(2):156–174, 1983.

[4] C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman, and M. Berlin. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *Proc. IROS 2005*, pages 708–713, 2005.

[5] A. G. Brooks and C. Breazeal. Working with robots and objects: Revisiting deictic reference for achieving spatial common ground. In *Proc. HRI 2006*, pages 297–304, 2006.

[6] M. C. Caselli. Communicative gestures and first words. In V. Volterra and C. Erting, editors, *From gesture to language in hearing and deaf children*, pages 56–67. Springer, 1990.

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

[8] C. J. Fillmore. Towards a descriptive framework for spatial deixis. In R. Jarvella and W. Klein, editors, *Speech, Place and Action*, pages 31–59. John Wiley, 1982.

[9] S. Goldin-Meadow. The role of gesture in communication and thinking. *Trends in cognitive sciences*, 3(11):419–429, 1999.

[10] S. Harrison. The creation and implementation of a gesture code for factory communication. In *Proc. GESPIN 2011*, 2011.

[11] Y. Hato, S. Satake, T. Kanda, M. Imai, and N. Hagita. Pointing to space: modeling of deictic interaction referring to regions. In *Proc. HRI 2010*, pages 301–308, 2010.

[12] C.-M. Huang and B. Mutlu. Modeling and evaluating narrative gestures for humanlike robots. In *Proc. RSS 2013*, 2013.

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

[14] A. Kendon. *Gesture: Visible action as utterance*. Cambridge U, 2004.

[15] S. Kita. Pointing: A foundational building block of human communication. *Pointing: Where Language, Culture, and Cognition Meet*, 1:1–8, 2003.

[16] A. Kobsa, J. Allgayer, C. Reddig, N. Reithinger, D. Schmauks, K. Harbusch, and W. Wahlster. Combining deictic gestures and natural language for referent identification. In *Proc. COLING 1986*, pages 356–361, 1986.

[17] J. D. Lempers. Young children's production and comprehension of nonverbal deictic behaviors. *The Journal of Genetic Psychology*, 135(1):93–102, 1979.

[18] P. Liu, D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita. It's not polite to point: generating socially-appropriate deictic behaviors towards people. In *Proc. HRI 2013*, pages 267–274, 2013.

[19] S. C. Lozano and B. Tversky. Communicative gestures facilitate problem solving for both communicators and recipients. *Journal of Memory and Language*, 55(1):47–63, 2006.

[20] D. McNeill. *Hand and mind: What gestures reveal about thought*. U Chicago, 1992.

[21] Á. Miklósi and K. Soproni. A comparative analysis of animals' understanding of the human pointing gesture. *Animal cognition*, 9(2):81–93, 2006.

[22] G. R. Morrison and G. J. Anglin. Research on cognitive load theory: Application to e-learning. *Educational Technology Research and Development*, 53(3):94–104, 2005.

[23] C. M. Murphy. Pointing in the context of a shared activity. *Child Development*, 49(2):371–380, 1978.

[24] Y. Okuno, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita. Providing route directions: design of robot's utterance, gesture, and timing. In *Proc. HRI 2009*, pages 53–60, 2009.

[25] M. Salem, S. Kopp, I. Wachsmuth, K. Rohlfing, and F. Joublin. Generation and evaluation of communicative robot gesture. *International Journal of Social Robotics*, 4(2):201–217, 2012.

[26] J. Shah and C. Breazeal. An empirical analysis of team coordination behaviors and action planning with application to human–robot teaming. *Human Factors*, 52(2):234–245, 2010.

[27] A. St Clair, R. Mead, and M. J. Mataric. Investigating the effects of visual saliency on deictic gesture production by a humanoid robot. In *Proc. RO-MAN 2011*, pages 210–216, 2011.

[28] O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita. Natural deictic communication with humanoid robots. In *Proc. IROS 2007*, pages 1441–1448, 2007.

[29] J. G. Trafton, N. L. Cassimatis, M. D. Bugajska, D. P. Brock, F. E. Mintz, and A. C. Schultz. Enabling effective human-robot interaction using perspective-taking in robots. *Systems, Man and Cybernetics, Part A: Systems and Humans*, 35(4):460–470, 2005.

[30] D. Vogel and R. Balakrishnan. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proc. UIST 2004*, pages 137–146, 2004.