Is it Pointless? Modeling and Evaluation of Category Transitions of Spatial Gestures

Adam Stogsdill
astogsdill@mymail.mines.edu
Colorado School of Mines
Golden, Colorado, US

Grace Clark
geclark@mymail.mines.edu
Colorado School of Mines
Golden, Colorado, US

Aly Ranucci
aranucci@mymail.mines.edu
Colorado School of Mines
Golden, Colorado, US

Thao Phung
thaophung@mymail.mines.edu
Colorado School of Mines
Golden, Colorado, US

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

ABSTRACT

To enable robots to select between different types of nonverbal behavior when accompanying spatial language, we must first understand the factors that guide human selection between such behaviors. In this work, we argue that to enable appropriate spatial gesture selection, HRI researchers must answer four questions: (1) What are the factors that determine the form of gesture used to accompany spatial language? (2) What parameters of these factors cause speakers to switch between these categories? (3) How do the parameterizations of these factors inform the performance of gestures within these categories? and (4) How does human generation of gestures differ from human expectations of how robots should generate such gestures? In this work, we consider the first three questions and make two key contributions: (1) a human-human interaction experiment investigating how human gestures transition between deictic and non-deictic under changes in contextual factors, and (2) a model of gesture category transition informed by the results of this experiment.

KEYWORDS

deixis, gesture, spatial reference, human-robot interaction

ACM Reference Format:
Adam Stogsdill, Grace Clark, Aly Ranucci, Thao Phung, and Tom Williams. 2021. Is it Pointless? Modeling and Evaluation of Category Transitions of Spatial Gestures. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21 Companion), March 8–11, 2021, Boulder, CO, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3434074.3447199

1 INTRODUCTION

Suppose that you, wherever you are, were to mention to a colleague that HRI 2021 was expected to be held in Boulder, Colorado. Even if you know quite well the general direction that Boulder is from your current location, you would probably not introduce it by turning to face Boulder, extending your arm precisely, and gazing intently at your office wall. Instead, you would likely simply wave your hand as if to say "elsewhere" – a purely non-deictic gesture which, even if somewhat resembling a deictic gesture in form, is not intended to communicate a vector or cone to be followed for, e.g., establishment of 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 discuss it by gesturing vaguely, but would instead use a deictic gesture such as pointing or presenting, gaze at your referent directly, and perhaps even actively check to ensure your interlocutor was following your gaze and gesture, for the purposes of establishing shared attention.

For humans, we expect that in many cases the selection of gestures to accompany referring expressions will be straightforward, simply due 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 towards a far-off location, we have little to no idea the heading along which target referents lie, and would be unable to precisely gesture towards most referents without seconds or minutes of careful deliberate thought and geometric reasoning.

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. For robots to effectively and naturally use *spatial* gestures, i.e., gestures that accompany language about entities and their locations, we thus argue that it is critical to provide robots with an understanding of the contexts in which *humans* find it natural and appropriate to generate spatial gestures, and to understand how their own gestures will then need to be employed in those contexts.

More formally, to enable appropriate gesture selection, we argue that HRI researchers need answers to four questions:

- **(Q1)** What are the factors that determine the categories of gesture used to accompany spatial language?
- **(Q2)** What parameters of these factors cause speakers to switch between these categories?
- **(Q3)** How do the parameters of these factors inform the performance of gestures within these categories? and

---

HRI '21 Companion, March 8–11, 2021, Boulder, CO, USA

ACM ISBN 978-1-4503-8290-8/21/03…$15.00
https://doi.org/10.1145/3434074.3447199

- **(Q4)** How does human generation of gestures differ from human expectations of how robots should generate such gestures?

To begin to answer some of these questions, we make two key contributions:

- **(C1)** a human-human interaction experiment investigating how human gestures transition between deictic and non-deictic under changes in contextual factors, and
- **(C2)** a model of gesture category transition informed by the results of this experiment.

2 RELATED WORK

2.1 Human Gesture

Gesture is one of the most important ways in which humans communicate with each other, not only serving as a channel for communication, but also playing a major role in human cognition during verbal communication [18]. Gesture has been demonstrated to enable speakers to work through and better articulate concepts, with gestures generated during speech even by blind speakers who are unable to see others use such gestures [25].

Moreover, not only does gesture facilitate the production of human speech, it also allows listeners to better understand the meaning and intentions behind speakers' utterances, both in typical dialogue as well as in contexts in which words cannot be used or in which interlocutors speak different languages [27, 39]. Gesture is especially useful in such contexts due to its visual nature; as Kendon [27] argues, gesture allows speech to convey additional *mental imagery* that persists even once the speaker has finished speaking. Research has demonstrated that gestures become more common when a speaker is referencing spatial information, with speakers gesturing significantly more often when discussing a spatial topic versus when they are discussing a more abstract topic [3]. As argued by McNeill [29], this is in part because gestures are primarily used to communicate visuospatial information to supplement the primarily non-visuospatial communication of speech.

As delineated by McNeill [29] human gestures can be divided into five main categories: deictics (which serve to pick out physical referents), iconics (which resemble physical shapes), metaphorics (which represent more abstract concepts), cohesives (in which abstract gestures are used to metaphorically connect narrative elements), and beats (which do not reflect concepts at all and instead provide emphasis and reflect tempo). While categories such as *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, such as pointing, presenting, and sweeping, are also inherently spatial. Use of deictic gestures is especially pronounced, for example, in contexts where speakers are giving directions [4] or describing room layouts [36]: contexts in which the topic of discussion is explicitly spatial.

But moreover, deictic reference, whether in the form of deictic language, deictic gaze, or deictic gesture, is a critical part of situated human-human communication [28, 30]. Deixis is one of the earliest forms of communication, both anthropologically and developmentally. Beginning around 9-12 months, humans learn to use deictic gesture, especially pointing, during speech [5], with mastery of deictic reference attained by around age 4 [12].

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 [20], that require (or at least benefit from) more communication channels beyond speech [13, 16, 17, 19, 26].

2.2 Robot Gesture

Due to the ubiquity and utility of gestures, deictic or otherwise, in human-human communication, Human-Robot Interaction researchers have also sought to enable this effective and natural communication modality in robots.

Deictics have been of particular interest within human-robot interaction due to the situated, task-oriented nature of much human-robot communication. Research has shown that robots' use of deictic gesture is effective at shifting attention in the same way as humans' use of deictic gesture [9], and that robots' use of deictic gesture improves both subsequent human recall [24] and human-robot rapport [6]. Research has also shown that robots' use of deictic gesture is especially effective when paired with other nonverbal signaling mechanisms [10], such as *deictic gaze*, in which a robot (actually or ostensibly) shifts its gaze towards its intended referent [1, 2, 11], and that this is especially effective when gaze and gesture are appropriately coordinated [33]. These findings have motivated a variety of technical approaches to deictic gesture generation [22, 34, 38], as well as a number of approaches for integrating gesture generation with natural language generation [14] (see also [15, 32, 37]).

In addition, there have also been a number of studies examining other gestures in HRI, including beat [8], iconic [7, 23], and metaphoric gestures [23]. Moreover, many of these approaches (e.g., [23, 24]) look at deictic and other gestures together within the context of a single gesture generation system. However, in these systems it is typically clear exactly when to use deictic gesture (e.g., when picking out an object or representation in a prominent display that is featured within the task environment).

In contrast, little work has been done exploring contexts in which robots must refer to target referents that are not prominently displayed, and must thus choose between meaningful deictic *and* non-deictic gestures – or something in between. This is in part due to the meaningful distinctions between deictic and non-deictic gestures. Deictic gestures serve a very specific, task-relevant, cognitive-pragmatic purpose that connects robots' communication to the world around them and seeks to acutely direct the attention of others. Meanwhile, other gestures such as beat, iconic, and metaphoric gestures instead serve to "grease the wheels" of conversation, improving conversational flow, providing emphasis, smoothing or facilitating the communication of actions and abstract ideas, and aiding speaker cognition. Because these two overarching categories of gesture, deictic and non-deictic, serve fundamentally different purposes, they are largely noninterchangeable, and it would be typically inefficient or inappropriate to try to use a deictic gesture to communicate an action or abstract concept or vice versa.

However, as we will describe, there are in fact cases in which deictic and non-deictic gestures *can* be used to communicate the same thing, and in which a robot cannot simply rely on utterance semantics to select gestures, but must instead rely on key *contextual* factors. In this paper, we consider one such case: *spatial gestures* in which gestures are generated to accompany *spatial references*.

Most examinations of gesture accompanying spatial reference have, within the HRI literature, been focused on deictic gestures. This is mainly a factor of the types of contexts in which HRI research is typically conducted. In typical HRI domains, a limited and finite set of objects is assumed to be under discussion and/or known to interlocutors, and these objects are typically located immediately in front of the robot or are at least visible in the environment (e.g., on a table [1, 2, 14, 21, 35] or screen [23]) [although cp. 31]. In such cases, the most natural gesture to accompany spatial language is a deictic gesture, pointing towards and gazing 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 often much larger. As highlighted in recent work [40, 41] 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). In these cases, such as the example used at the beginning of this paper, humans use significantly different gestural behavior, and it can be similarly expected that robots would need to as well. In this work we seek to understand the design requirements for robot gesture generation in such contexts.

Our investigation is guided by two key research hypotheses:

- **(H1)** Humans use both deictic and non-deictic gestures to accompany spatial references.
- **(H2)** Humans' choice bewteen deictic and non-deictic gestures is dependent on contextual factors such as the visibility, distance, or expected knowledge of their target referents.

3 GESTURE MODELING AND DESIGN

To assess our hypotheses, we plan to use a research paradigm similar to that used in other gesture-related work in HRI [23], in which we first *model* human use of deictic and non-deictic gestures in the course of spatial referring, then *implement* exemplars of those gestures on a humanlike robot, and finally *evaluate* how that robot's generation of those gestures according to that model might impact key measures associated with successful human-robot interactions. In this preliminary work, we describe the modeling phases of this research paradigm. We will implement these gestures and evaluate their impact on human-robot interactions in future research.

3.1 Studying Category Transitions in Human Gestures

To investigate human category transitions between gestural categories accompanying spatial referring expressions, we designed an IRB-approved spatial reference task in which participants needed to sequentially refer to each of a series of objects and locations expected to be familiar to all participants. 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.

We recruited 14 participants from a mid-sized US college campus to engage in this spatial reference task. Demographic information was not collected from the participants. Each participant 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 verbally describe the location of each object or location, with the experimenter themselves only referring to each target by name (i.e., without themselves using any gaze or gestural behaviors or describing the target in any way). If participants asked for clarification on how to describe an object, they were encouraged to tell the location of the object in the way that made the most sense to them. Gestures were not mentioned by the experimenter at any time during the task; the participants were not encouraged, instructed, or required to use gestures when describing the location of the objects.

Participants' gestural behaviors were videotaped using an RGB video camera for coding and analysis. A supplemental recording of each participant was taken using an RGB-D camera (i.e., the XBox One Microsoft Kinect) in order to track joint positions for future work involving more fine-grained analysis and modeling. All recordings from the RGB video camera were coded by a primary rater, who categorized the gesture accompanying participants' spatial referring expressions to the target objects as either deictic, non-deictic, or non-gestural. Here, gestures such as pointing, sweeping, and presenting were categorized as deictic (c.p. [35]), and all other hand motions (including metaphoric, iconic, and beat gestures) categorized as non-deictic. Examples of deictic and non-deictic gestures are displayed in figure 1. 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. In total, 254 recorded descriptions were retained.

3.2 Modeling Category Transitions in Human Gestures for Robots

To identify factors relevant to category transitions in human gestures for deployment on robots, we conducted a Bayesian analysis of our coded interactions. The brms package for R was used to fit and compare a series of General Linear Mixed Models to these coded interactions, with each model using 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. The analysis performed by brms operated by (1) fitting these models, (2) quantifying the evidence for each of these models (including a null model that did not include distance, visibility, or speaker), and (3) computing Bayes Inclusion Factors Across Matched Models (i.e., "Baws Factors") to quantify the relative evidence for inclusion vs. noninclusion of each of these three factors as well as their potential interactions.

The results of our statistical tests suggest that both visibility and distance are important for choosing whether and how to gesture when generating spatial referring expressions. Specifically, our

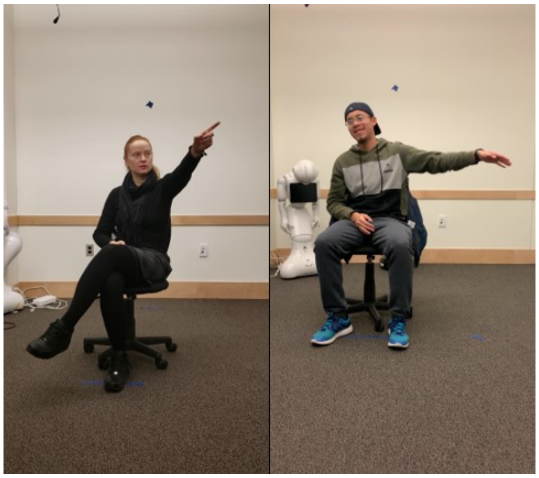

Figure 1: Example Deictic and Non-deictic Gestures

Baws Factor analysis suggests that it is unlikely but uncertain whether distance directly informs spatial 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). However, it is very likely that visibility directly informs spatial 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 suggests 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 deictic gestures when their target was visible than when it was not (with 87.6% of the 97 visible referent descriptions using deictic gestures vs 13.4% of the 157 non-visible referent descriptions using deictic gestures), and that when target referents were not visible, speakers became increasingly less likely to use deictic gestures and more likely to use non-deictic gestures as their targets grew increasingly far away (with, for example, 22.2% of descriptions using deictic gestures and 51.9% of descriptions using non-deictic gestures for referents at a distance of 12 feet, vs 7.1% of descriptions using deictic gestures and 71.4% of descriptions using non-deictic gestures for referents at a distance of 270 feet).

Algorithm 1 is a simple, preliminary method for determining the type of gesture a robot might use in a spatial-reference situation, based on the relevant factors suggested by our statistical analysis. When referring to some object, if that object is visible **or** if that object is within some threshold distance from the robot, the robot should select a deictic gesture. Otherwise (i.e., if the object is **not** visible and lies beyond some threshold distance from the robot), the robot should choose a non-deictic gesture.

Algorithm 1 Gesture Selection Model

```
procedure SELECT-GESTURE($O_V$, $O_D$)
    $O_V$: a binary variable indicating whether the target referent is visible or not
    $O_D$: a continuous variable indicating the distance to the target referent
    $\tau$: a threshold distance beyond which non-deictic gestures should be used
    if $O_V$ == True or $O_D < \tau$ then
        Return Deictic
    else
        Return Non-deictic
    end if
end procedure
```

4 DISCUSSION

Our first hypothesis (**H1**) was that humans use both deictic and non-deictic gestures to accompany spatial gestures. The results of our initial spatial reference task support this hypothesis. Every participant used both deictic and non-deictic gestures to accompany spatial references, with 71.4% of participants using each gesture type more than twice. Similarly, all but two of the objects (87.5%) were referred to using both deictic and non-deictic gestures.

Our second hypothesis (**H2**) was that humans' use of deictic vs. non-deictic gesture is dependent on contextual factors such as the visibility, distance, or expected knowledge of their target referents. The results of our initial spatial reference task support this hypothesis and suggest a model that includes visibility and distance of an object as factors.

Limitations and Future Work— One limitation of our experiment was that there were no objects that the participant observed that were both non-visible and very close or visible and far away. We plan to address this in future work.

In future work, we hope to implement typical gestures from this study on a humanlike robot and evaluate how their generation and use of such gestures according to the model developed above affects various aspects of human-robot interactions. Additionally, we hope to use a Generative Adversarial Network model to produce animations learned from the Kinect data that we collected in the study. This will help provide automated methods for robots to (1) choose appropriate gesture categories to use when referring to given objects and (2) generate those gesture themselves (i.e., generate relevant gestures without puppeteering), based on the known properties of target referents.

5 CONCLUSION

The use of gestures is integral to human communication of spatial information. This work sought to understand what contextual parameters affect the use of different gesture types in a spatial reference task and developed a model for gesture usage based on these parameters.

ACKNOWLEDGMENTS

This work was funded in part by National Science Foundation grant IIS-1909864.

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

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

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

[8] 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.

[9] Andrew G Brooks and Cynthia Breazeal. 2006. Working with robots and objects: Revisiting deictic reference for achieving spatial common ground. In *Proc. Int'l Conf. HRI*.

[10] 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).

[11] 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 *RO-MAN*.

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

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

[14] 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.

[15] 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.

[16] 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.

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

[18] 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.

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

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

[21] 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.

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

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

[24] 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.

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

[26] 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.

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

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

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

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

[31] 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.

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

[33] 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.

[34] Maha Salem, Stefan Kopp, et al. 2010. Towards an integrated model of speech and gesture production for multi-modal robot behavior. In *RO-MAN*.

[35] 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.

[36] 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.

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

[38] David Whitney, Eric Rosen, James MacGlashan, Lawson LS Wong, and Stefanie Tellex. 2017. Reducing errors in object-fetching interactions through social feedback. In *ICRA*.

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

[40] 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.

[41] 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.