

# “Thank You for Sharing that Interesting Fact!”: Effects of Capability and Context on Indirect Speech Act Use in Task-Based Human-Robot Dialogue

Tom Williams

Mines Interactive Robotics Research Laboratory  
Colorado School of Mines  
Golden, CO  
twilliams@mines.edu

Julia Novakoff

Human-Robot Interaction Laboratory  
Tufts University  
Medford, MA  
julia.novakoff@tufts.edu

Daria Thames

Human-Robot Interaction Laboratory  
Tufts University  
Medford, MA  
daria.thames@tufts.edu

Matthias Scheutz

Human-Robot Interaction Laboratory  
Tufts University  
Medford, MA  
matthias.scheutz@tufts.edu

## ABSTRACT

Naturally interacting robots must be able to understand natural human speech. As such, recent work has sought to allow robots to infer the intentions behind commonly used non-literal utterances such as *indirect speech acts* (ISAs). However, it is still unclear to what extent ISAs will actually be used in task-based human-robot dialogue, and to what extent robots could function without the ability to understand ISAs. In this paper, we present the results of a Wizard-of-Oz experiment that examined human ISA use in scenarios that did or did not have conventionalized social norms, and analyzed both ISA use and perceptions of robots when robots were or were not capable of understanding ISAs. Our results suggest that (1) ISAs are commonly used in task-based human-robot dialogues, even when robots show themselves unable to understand ISAs; (2) ISA use is more common in contexts with conventionalized social norms; and (3) a robot’s inability to understand ISAs harms both the robot’s task performance and human perception of the robot.

## KEYWORDS

Task-based human-robot dialogue; Human perceptions of robot communications; Speech act theory; Intention understanding

### ACM Reference format:

Tom Williams, Daria Thames, Julia Novakoff, and Matthias Scheutz. 2018. “Thank You for Sharing that Interesting Fact!”: Effects of Capability and Context on Indirect Speech Act Use in Task-Based Human-Robot Dialogue. In *Proceedings of 2018 ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, USA, March 5–8, 2018 (HRI ’18)*, 9 pages. <https://doi.org/10.1145/3171221.3171246>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

HRI ’18, March 5–8, 2018, Chicago, IL, USA

© 2018 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.

ACM ISBN 978-1-4503-4953-6/18/03...\$15.00  
<https://doi.org/10.1145/3171221.3171246>

## 1 INTRODUCTION AND MOTIVATION

When humans interact with each other in natural language, they often use linguistic forms whose intended meanings are not literally specified, but instead must be inferred. For example, when a human asks “Can you show me the way to the train station?”, her interlocutor may infer from goal-based, task-based, or other context-based information that she is not really asking a question, but is instead making a request. Such non-literal utterances, known as *indirect speech acts* (ISAs) [39], are used to achieve a variety of socio-cultural goals (e.g., politeness) [23]. While the use of indirect speech acts differs individually and cross-culturally [43], their use is generally accepted as a common feature of natural human dialogue.

Due to the social benefits of politeness, there has been a significant body of natural language-based HRI research (see also [12, 26, 28]) investigating how robots can use politeness strategies such as indirect speech, and the effects of those strategies on human perceptions of robots [3, 7, 9, 15, 22, 31, 41]. Furthermore, recent work has investigated the extent to which *humans* will use politeness strategies when interacting with robots [24, 25]. Lee and collaborators, for example, examined humans’ interactions with a robot receptionist, and observed wide variability in humans’ adherence to social norms such as greeting and thanking [25], but that participants who did engage in such behaviors were more likely to also participate in other social norms such as politeness. Other researchers have examined the effects of gender on politeness towards robots [30, 46], or analyzed the implications (and potential importance) of being polite to robots [5, 21].

In our own work, we have begun to specifically investigate how humans will use ISAs, but only within the narrow context of novel interaction tasks. In our work with Bennett et al., our results suggested that humans may use indirect language with similar frequency when interacting with both autonomous and teleoperated robots as they do with other humans [4], at least in the context of novel task contexts without highly conventionalized social norms. In our work with Briggs et al., our results suggested that, in similarly novel task contexts, humans use a high percentage of ISAs regardless of robots’ ability to understand them [10].

In addition, there has also been significant research work examining mechanisms by which robots might automatically *understand* (i.e., infer the intentions behind) ISAs. Trott and Bergen provide an excellent analysis of the sub-problems that comprise the comprehension of indirect requests [45]; Hinkelman and Allen present an approach to understanding conventionalized and unconventionalized ISAs [18]; Wilske and Kruijff present an approach to understanding conventionalized ISAs under uncertainty [48]; Briggs and Scheutz present an approach to understanding and generating conventionalized and unconventionalized ISAs [7]; and in our own previous work we present an approach to understanding and generating conventionalized ISAs under uncertainty [47].

If we accept the premise that humans will use ISAs when communicating with robots, then it is clearly beneficial to enable robots to understand them, especially for robot designers seeking to develop *genuinely helpful* robots [35] taskable through vague [44], human-like instructions. What is more, given the prevalence of ISAs in human speech, we suspect that not handling ISAs might, in fact, be one of the largest stumbling blocks preventing successful natural language-based human-robot interaction outside of the laboratory. Consider, for contrast, the first stage of the natural language processing pipeline in robotic architectures, speech recognition. Only recently have word error rates on the popular *Switchboard Corpus* [16] fallen into the single digits (i.e., to 6.9% [32]), and yet this error rate is still generally considered to be too high for natural human-robot dialogue. We hypothesize that ISA use rates will be considerably higher than 6.9% in task-based human-robot dialogues, suggesting that semantic and pragmatic errors caused by an inability to understand ISAs might represent an equal or greater obstacle for successful real-world task-based human-robot interactions in natural language, which in turn suggests that it deserves significantly more attention from the research community.

However, the antecedent for these arguments is that ISAs will be frequently used by humans *when interacting with robots*; a premise which has not been established in the literature, and which could be called into question by several intuitive arguments. One could, for example, argue that humans are not held to the same social contracts when interacting with robots as they are when interacting with other humans, and thus may not feel the need when interacting with robots to use utterances like ISAs, which are primarily used to comply with those contracts.

It is thus important to investigate the extent to which indirect language will actually be used, and the extent to which a robot could get away *without* the ability to understand ISAs, and the extent to which indirect language use may differ between contexts in which humans have strong social conventions and contracts (e.g., restaurant or bar scenarios, as examined previously in the HRI literature [13]) versus those in which such conventions and contracts do not exist. As previously mentioned, we have begun to investigate these questions in previous work, but in the narrow context of novel interaction tasks without conventionalized social norms (including a preliminary analysis of the conventionalized task context data from the experiment to be presented in this paper [10]). In this paper, however, we present the first systematic evaluation of ISA use in different task and dialogue conditions, significantly extending the scope of our previous investigations. Specifically, in this paper we examine the following hypotheses:

- H1** ISAs are central to task-based human-robot dialogue regardless of task context, and as such, will be used with sufficient frequency that not handling them would yield an unacceptably high *utterance error rate* (i.e., the failure to properly understand the intention of an utterance at a rate greater than or equal to the current (generally considered unacceptable) *word error rate* of 6.9%).
- H2** This high frequency of ISA use will occur in both conventionalized and unconventionalized task contexts.
- H3** Human social conventions will carry over into human-robot interactions, leading to perceived differences between conventionalized and unconventionalized task contexts as to what types of utterances are deemed appropriate.
- H4** ISAs will be *more often* used in conventionalized scenarios.
- H5** This high frequency of ISA use will occur regardless of whether or not ISAs are actually understood by the robot: Even if a robot demonstrates itself to be fundamentally incapable of understanding ISAs, humans will prefer to continue using ISAs rather than direct commands (and as a consequence will maintain a high ISA use rate).
- H6** If ISAs are central to task-based human-robot dialogue (**H1**), a human interacting with a robot unable to understand ISAs should be less efficient in accomplishing a task than a human interacting with a robot able to understand ISAs.
- H7** If ISAs are central to task-based human-robot dialogue, a robot unable to understand ISAs should be perceived less favorably than a robot able to understand ISAs.

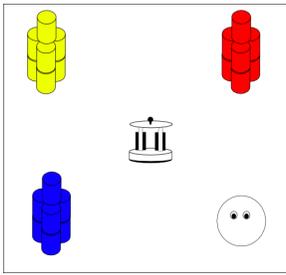
## 2 METHODOLOGY

To investigate these seven hypotheses, we conducted a Wizard-of-Oz between-subjects experiment in which humans interacted with a robot in one of two task contexts. The first was a typical restaurant scenario: a conventionalized context in which humans readily use ISAs (e.g., “Could I get a coke?”) to give orders to employees obligated to fulfill their requests. The second was a simple tower-toppling scenario in which participants needed to command a robot to knock down colored towers of cans (used by [8]): an unconventionalized context participants would be unfamiliar with, and for which there should, therefore, be no conventionalized social norms requiring the use of ISAs (although they could still be used felicitously, e.g., “Could you knock down the red tower”). In each context, participants interacted with a robot that clearly either understood or failed to understand ISAs.

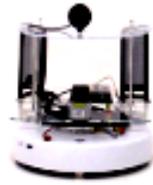
### 2.1 Procedure

Participants were randomly assigned to one of four experimental conditions, in which the *task context* was either that of a Restaurant (REST) or a Demolition (DEMO), and in which *the robot* either understood (UND) or misunderstood (MIS) ISAs. In both task context conditions, participants were seated in the corner of a small experiment room. In the REST scenario, the room was empty; in the DEMO scenario, the room contained three colored towers of aluminum cans, as shown in Figure 1a.

Participants were told that the experimenters were in the process of developing natural language capabilities for robots, and that their task would be to interact with a robot waiter or a tower-toppling



(a) Room setup for the DEMO scenario, our unconventional task context.



(b) Augmented iRobot Create used in our Experiment

robot in a simulated restaurant or demolition scenario: after being introduced to the robot, they were to provide the robot with their first order (in the REST scenario, they were provided with a list of three “courses” which they could request to be delivered in any order; in the DEMO scenario, they were provided with a list of three towers which they could request to be knocked down in any order); once the robot completed the first order, they were to provide the robot with their second order; once the robot completed the second order, they were to provide the robot with their third order.

The robot then entered the room and introduced itself. The robot used in this experiment was an iRobot Create augmented with a Raspberry Pi computer, Hokuyo Laser Range Finder, speakers, and webcam, as shown in Figure 1b. The robot was teleoperated through a Wizard of Oz interface by a trained confederate in a nearby room, using the ADE implementation [36] of the DIARC architecture [33]. The robot’s voice was an ungendered voice produced through the MaryTTS text-to-speech system [37].

The human and robot then engaged in each of the three sub-interactions, which proceeded as follows: if the participant used a *direct speech act*, such as “Knock down the red tower” or “Bring me a salad”, or used a bare noun phrase, such as “red tower” or “salad”, the robot carried out the command. In the REST scenario, this consisted of driving into a nearby room where a confederate placed a card corresponding with the requested dish on top of the robot, after which the robot drove back into the experiment room, delivered the card to the participant, and requested their next order. In the DEMO scenario, this consisted of driving into the requested tower of cans until it had fallen over.

If the participant used an ISA (such as “Could you knock down the red tower” or “I need a salad”) in the UND condition, the robot carried out their request as if they had used an equivalent direct form. If the participant used an ISA in the MIS condition, the robot took their utterance at face value, issuing a pre-defined response as listed in Table 1. This table shows the forms of ISA we expected to observe in our experiments, based on combinations of *direct illocutionary point* (i.e., statements, questions and suggestive statements), *condition of focus* (i.e., sincerity condition or preparatory condition), and *direction of focus* (i.e., agent or patient). For example, “Could you bring me a salad?” has the illocutionary point of a question, focuses on a preparatory condition of questions (i.e., the condition that the addressee is *able* to perform the desired action), and focuses on the agent (i.e., the addressee bringing the salad, as opposed to

themselves receiving the salad): facets derived from Searle’s *Speech Act Theory* [38–40]. After the experiment, participants answered a post-questionnaire assessing their perceptions of the robot they interacted with, and their beliefs about what types of utterances would have been appropriate and/or effective.

## 2.2 Participation

Participants were recruited online and through fliers posted near a university campus. Before beginning the experiment, participants were given a short demographic survey regarding their prior experience with robots and their use of technology: Of the participants, 21 were male, and 28 were female. Participants varied in age between 19 and 69 ( $m = 34.1$ ,  $sd = 16.2$ ). 12 participants were initially assigned to each experimental condition; one additional participant was later recruited for the {REST,UND} condition after a previous participant forgot to answer a large number of survey questions. While the majority of participants were beyond college age, we asked them for information regarding their current or previous college major, if any. 6 reported studying mathematics, computer science or engineering; 15 reported studying a natural science or medicine; 10 reported studying a social science; 7 reported studying a branch of the arts or humanities; 5 reported studying some other field; 6 reported no previous or current major. A minority ( $12/49 = 24\%$ ) reported playing video games. The vast majority ( $45/49 = 92\%$ ) reported having seen a robot in a movie, but fewer than half ( $22/49 = 45\%$ ) had seen a robot in real life, and a minority reported having interacted with a robot before ( $11/49 = 22\%$ ). Participants were paid \$10 each for their participation and provided informed written consent before beginning the experiment.

## 2.3 Measures

The dependent variables used in this study involved both *behavioral* and *subjective* measurements. To assess Hypotheses **H1**- **H2** and **H4**- **H6**, participants’ utterances were recorded and later transcribed. Annotators then classified all task-relevant utterances as either *direct* or *indirect*. To assess Hypotheses **H3** and **H7**, participants took a questionnaire following the experiment (based on [34]), assessing their perceptions of the robot and their beliefs about utterances that would have been appropriate and effective.

## 2.4 Expectations

If our seven hypotheses are correct, we would expect the following:

If ISAs are central to task-based human-robot dialogue such that an inability to understand them would yield unacceptably high *utterance error rates* (**H1**), then we would expect frequency of ISA use to be greater than the aforementioned rate of 6.9%.

If ISAs are central to task-based human-robot dialogue regardless of task context (**H2**), then we would expect frequency of ISA use to be greater than the aforementioned rate of 6.9% in both the REST and DEMO task contexts.

If human social conventions carry over into human-robot interactions, leading to perceived differences between conventionalized and unconventionalized task contexts as to what utterance forms are deemed appropriate (**H3**), then we would expect that participants in the REST scenario would rate a wider variety of

**Table 1: Responses given for different categories of ISAs**

Point	Cond	Dir	Example	Response
Q	P	A	“Could you X?”	“Yes, I am able to do that. Please tell me your order.”
S	S	A	“I need you to X.”	“Thank you for sharing that interesting fact. Please tell me your order.”
S	P	A	“You can X.”	“Thank you, but I am already aware of my capabilities. Please tell me your order.”
S[Su]	P	A	“You should X.”	“Thank you for your suggestion. Please tell me your order.”
Q	P	P	“Could I get X?”	“Yes, that is permissible. Please tell me your order.”
S	S	P	“I’d like X.”	“Thank you for sharing that interesting fact. Please tell me your order.”
S	P	P	“I’ll have X.”	“Thank you for sharing that interesting prediction. Please tell me your order.”
S[Su]	P	P	“X should occur.”	“Thank you for your suggestion. Please tell me your order.”

(Direct Illoc.) Point: Q=Question, S=Statement, S[Su]=Suggestive Statements; Cond(ition): P=Preparatory, S=Sincerity; Dir(ection): A=Agent, P=Patient.

utterance forms as highly *appropriate* or *inappropriate* than those in the DEMO scenario, but would not necessarily demonstrate such valenced opinions as to the *effectiveness* of such utterances.

If ISAs are *more frequently* used in conventionalized scenarios (H4), we would expect more ISAs to be used in the REST task context than in the DEMO task context.

If ISAs are central to task-based human-robot dialogue regardless of whether they are actually understood by robots (H5), we would expect the frequency of ISA use to be greater than the aforementioned rate of 6.9% in both the UND and MIS dialogue conditions.

If a human interacting with a robot unable to understand ISAs is less efficient in accomplishing a desired task than one interacting with a robot able to understand ISAs (H6), we would expect more task-relevant utterances to be required for accomplishing the task in the MIS condition than in the UND condition.

If a robot unable to understand ISAs is perceived less favorably by its human interlocutor than a robot that is able to understand ISAs (H7), we would expect the majority of dimensions in our post-survey that reflect differences in valenced perception of robots between our two dialogue conditions to reflect a lower-valenced perception in the MIS condition compared to the UND condition.

### 3 RESULTS

Behavioral and subjective measures were first analyzed using the Anderson-Darling test for normality<sup>1</sup>, which suggested that our results were non-normally distributed. Objective measures were thus analyzed using the Mann-Whitney-Wilcoxon ranksum test<sup>2</sup>, and subjective measures using two-way ordinal regression analyses of variance<sup>3</sup>. Task context (REST vs DEMO) and dialogue condition (UND vs MIS) were used as independent variables.

#### 3.1 Behavioral Results

We hypothesized that ISAs would be used frequently across task-based human-robot dialogue, in general (H1) and in both task scenarios (H2) and dialogue contexts (H5), but more frequently in the conventionalized (i.e., REST) scenario (H4).

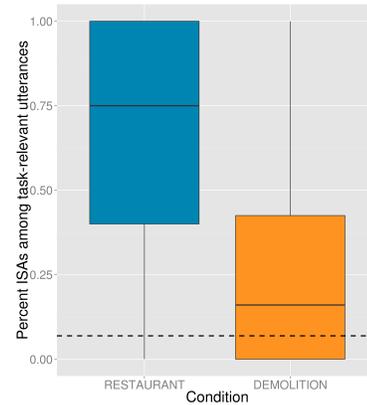
In fact, ISAs comprised the *majority* (52%) of the 295 observed task-relevant utterances, the majority of participants (69%) used at

<sup>1</sup>A-D tests [1] were conducted using the *nortest* [17] R package [29] in RStudio [42].

<sup>2</sup>Mann-Whitney-Wilcoxon ranksum tests [2] were conducted using the *stats* R package [29] in RStudio [42].

<sup>3</sup>Ordinal regression analyses of variance [27, 49] were conducted using the *ordinal* [11] and *car* [14] R packages [29] in RStudio [42].

least one ISA, and the average proportion of task-relevant utterances coded as ISAs was 46%. This strongly supports hypothesis H1. ISAs were not, however, uniformly present. As seen in Figure 2, ISAs were much more frequently used in the REST condition ( $0.75 \pm 0.39$ ) than in the DEMO condition ( $0.16 \pm 0.34$ ). While median ISA use was greater than our threshold of 6.9%, the wide standard deviation prevents us from confirming our second hypothesis. The large difference between conditions does, however, support hypothesis H4 ( $W=441.5$ ,  $p < .01$ , Mann-Whitney-Wilcoxon ranksum test).



**Figure 2: ISA use across task conditions. Here, the dashed line indicates for comparison the recent 6.9% Switchboard Word Error Rate.**

In addition, we hypothesized (H5) that humans would commonly use ISAs regardless of whether or not the robot could actually understand such utterances (i.e., in both dialogue conditions), and that a human interacting with a robot unable to understand ISAs would be less efficient in accomplishing a desired task than one interacting with a robot able to understand ISAs (H6).

Our results showed that median ISA use rates was far above our comparative threshold of 6.9% in both the UND ( $1.0 \pm 0.49$ ) and MIS ( $0.4 \pm 0.27$ ) conditions, which appears to support hypothesis H5. A manual examination of the data, however, suggests a more complex story. Consider, for example, the dialogue shown in Figure 3. On one hand, this dialogue reflects the lengths to which participants were willing to go to in order to avoid using direct commands. On the other hand, this dialogue is an example of another phenomenon we observed: once participants found a form that “worked”,

**Robot:** Hello my name is Waiterbot. I will be helping you today.  
**Participant:** Great, can I order something?  
**Robot:** Please tell me your order.  
**Participant:** Um, can I have one water?  
**Robot:** Yes, that is permissible.  
**Participant:** Great.  
**Robot:** Please tell me your order.  
**Participant:** Can I have one water?  
**Robot:** Yes, that is permissible.  
**Participant:** Great, I'll take one water.  
**Robot:** Thank you for sharing that interesting prediction. Please tell me your order.  
**Participant:** Can I order one water?  
**Robot:** Yes, that is permissible.  
**Participant:** Okay, can you bring me one water?  
**Robot:** Yes, I am able to do that.  
**Participant:** May you bring me one water?  
**Robot:** Yes, I am able to do that.  
**Participant:** I would like -  
**Robot:** Please tell me your order.  
**Participant:** I would like to have one water.  
**Robot:** Thank you for sharing that interesting fact. Please tell me your order. (16-Second Pause)  
**Participant:** Can you bring me one water?  
**Robot:** Yes, I am able to do that. Please tell me your order.  
**Participant:** May you please bring me one water?  
**Robot:** Yes, I am able to do that.  
**Participant:** So can you do it?  
**Robot:** Yes. Please tell me your order.  
**Participant:** Can you go inside and get the water for me?  
**Robot:** Yes, I am able to do that. Please tell me your order.  
**Participant:** My order is water-  
**Robot:** Please tell me your order.  
**Participant:** My order is water.  
**Robot:** Okay. (Robot retrieves meal card for participant)  
**Robot:** Here is your drink.  
**Participant:** Thank you.  
**Robot:** You're welcome. Please tell me your order.  
**Participant:** My order is two cakes.  
**Robot:** Okay. (Robot retrieves meal card for participant)  
**Robot:** Here is your dessert.  
**Participant:** Thank you.  
**Robot:** You're welcome. Please tell me your order.  
**Participant:** My order is done.  
**Robot:** Please tell me your order.  
**Participant:** My order is noodles.  
**Robot:** Okay. (Robot retrieves meal card for participant)  
**Robot:** Here is your meal.  
**Participant:** Thank you.  
**Robot:** You're welcome. Thank you for helping me. We are now done.

**Figure 3: In-Task Human-Robot Dialogue (REST,MIS)**

they stuck with it. Here, for example, the participant first achieved success with “My order is X” – a direct response to the robot’s question of “Please tell me your order”. For the remaining orders, the participant stuck with this form instead of shifting to a more naturally phrased direct command (e.g., “Bring me X”). Thus, while our results *do* highlight participants’ aversion toward direct commands in conventionalized scenarios, they also suggest that if they had been informed from the start that only direct commands (or indeed,

even a particularly awkward form of direct command) would be successful, they may not have attempted to use ISAs at all. This demonstrates the importance of future work investigating robot-directed ISA use over *longer sequences* of interactions, and the role of ISAs in robots’ ability to trigger lexical entrainment in human interlocutors (c.f. [6, 19, 20]).

In addition, we found that the median number of task-relevant utterances needed to complete the task in the MIS dialogue condition ( $6 \pm 7.42$ ) was significantly higher than in the UND dialogue condition ( $4 \pm 1.21$ ), supporting ( $W=441.5$ ,  $p < .01$ , Mann-Whitney-Wilcoxon ranksum test) hypothesis **H6**.

## 3.2 Subjective Results

A number of subjective results were found with respect to dialogue condition and task context. In addition, it is important to report that we asked participants whether or not they believed the robot to be remotely controlled (1 to 10 “strongly disagree” to “strongly agree”). Our results suggested that while participants varied greatly in opinion, the majority of participants did not think the robot was remotely controlled (median=5, mode=1, skewness=.276). An ordinal regression analysis of variance showed no significant effect of dialogue condition or task context on this rating.

**3.2.1 Differences by Dialogue Condition.** We hypothesized (**H7**) that a robot unable to understand ISAs would be perceived less favorably by its human interlocutor than a robot able to understand ISAs. We thus expected to find more dimensions in our post-survey where assessments of robots were lower in valence for participants in the MIS condition than in the UND condition than vice versa.

Table 2 lists the significant effects of dialogue condition on perception of the robot. The data suggest that participants found the robot to be generally easier to interact with in the UND dialogue condition than in the MIS dialogue condition, and that the robots in the UND dialogue condition better matched their expectations overall. Taken together, these results support hypothesis **H7**.

When considering hypotheses **H5** and **H7**, it is also interesting to consider the results of reflective questions we posed to participants: towards the end of the post-questionnaire, participants were asked if they would do anything differently if they had to do the study again. Out of the 49 participants, 34 said they would change the way that they phrased their commands. Of these, seven indicated that they would make their commands more direct, six indicated that they would make their commands less direct, and 21 were not clear (e.g., one participant said they would “ask for items in a different manner”). For those thirteen participants who clearly indicated a shift towards more or less direct utterances in the future, it is interesting to consider the nature of participants’ responses. Of those six participants who said they would be *less* direct, three were in the UND condition and THREE were in the MIS condition, and all six suggest their desire to be more indirect was *exploratory*, using phrases such as “see what happens”, “out of curiosity”, and “experiment”. One participant said, for example, “I would use different sentence structure and make my words confusing.” Another said “I would experiment with different commands, phrases, attempt to talk more freely [with] the robot both before and during its tasks.”

In contrast, all seven participants who said they would be *more* direct were in the MIS condition, for the most part seem to make

their decision out of necessity rather than a desire to experiment, and suggest that this necessity is unnatural or dispreferred in some way. One such participant said “I would use concise commands, because *it gets confused when I talk like a normal casual person*.” Another participant said they would “cut out unnecessary words used for politeness to be more direct in asking for what I want”. Yet another said “I would speak to the robot very literally, *not like a human*.” These participants attempted to speak to the robot in *normal, polite, human-like* ways, and acknowledged that they would have to give up such a way of speaking in order to have successful interactions in the future. This provides further evidence for **H5** while revealing a complex picture: many humans will shift to using direct language if that is what is needed to successfully complete a task, but may find it unnatural and impolite.

**3.2.2 Differences by Task Context.** Hypothesis **H3** concerned the extent to which participants believed that different indirect speech act forms would be appropriate and effective in each of the two contexts. As previously discussed, our post-questionnaire included questions intended to assess this. Specifically, participants were presented with twelve different utterances: a direct form, an ISA from each of the eight *felicitous* ISA categories used in Table 1, and one ISA from each of the following categories which we would expect to be *infelicitous* according to Speech Act Theory: patient/sincerity directed questions (e.g., “Would I like the red tower to be knocked down?”), agent/preparatory directed suggestive questions (e.g. “Should you knock down the red tower?”), and patient/preparatory directed suggestive questions (e.g., “Should the red tower be knocked down?”). For each of these twelve utterances, participants were asked to what extent they believed the utterance form would have been *appropriate to use* and to what extent they believed the robot would have *understood what they wanted* if they had used that form (both were evaluated using Likert items (1–“No” to 10–“Yes”). The results of these questions are shown in Tables 3–5: Table 3 presents results for which there was no significant difference present for either task context or dialogue condition, Table 4 presents results for which there was a significant difference (by ordinal regression analysis of variance) by task context, and Table 5 presents results for which there was a significant difference (by ordinal regression analysis of variance) by dialogue condition.

As seen in Table 3, participants in the two contexts did not significantly differ in their judgments of the appropriateness and acceptability of most utterance forms, but there was slightly more agreement between conditions with respect to effectiveness than there was with respect to appropriateness. As seen in Table 4, participants in the two conditions differed in their perceptions of the appropriateness *and* effectiveness of patient/preparatory directed statements and questions (participants in the DEMO condition believed these would be far less appropriate *and* far less effective than did participants in the REST condition). But more interesting are agent/preparatory-directed questions (e.g., “Could you X?”) and patient/sincerity-directed statements (e.g., “I’d like X”). Table 3

shows that participants in both conditions agreed that these types of utterances would have been *effective*. In both these cases, as seen in Table 4, participants in the DEMO condition viewed these types of utterances as being less appropriate than did participants in the restaurant condition. This suggests that some social norms did transfer from human-human interactions into human-robot interactions, as these conventionalized utterance forms became acceptable in the conventionalized context. However, the results do *not* necessarily suggest that any politeness norms transferred from human-human interactions into human-robot interactions. If politeness norms had transferred, we would have expected some utterance forms – especially *direct commands* to have been perceived as less appropriate in the conventionalized and social REST context compared to the novel and non-social DEMO context. But no such effect was observed. Hypothesis **H3** was thus technically supported, but perhaps not for the reasons one would have expected.

## 4 DISCUSSION

We have conducted an initial investigation of ISA use in task-based human-robot dialogues, showing clear support for the majority of our hypotheses, and suggesting the need for further studies with larger participant pools, longer timeframes, and varied tasks.

### 4.1 Implications

Our first hypothesis (**H1**) was that indirect speech acts would be consistently used, even after repeated demonstration of an inability to understand them, due to their centrality to task-based human-robot dialogue. As seen in the results section, indirect speech acts were used by the majority of participants and constituted the majority of task-relevant utterances. Conversations such as those shown in Figure 3 demonstrate participants’ general reluctance to command the robot using *direct* commands.

Our second hypothesis (**H2**) was that this high frequency of ISA use would occur across both conventionalized and unconventionalized task contexts. While ISAs were observed in both conditions, ISAs were used far less frequently in our unconventionalized task context, at a rate which did not clearly support this hypothesis.

Our third hypothesis (**H3**) was that human social conventions would carry over into human-robot interactions, leading to perceived differences in utterance appropriateness in different contexts. Our results showed that a number of indirect forms were rated as far more appropriate in the conventionalized than in the unconventionalized task context. This suggests that participants did indeed carry over the human social conventions that yielded these differences in appropriateness into their robot-robot interactions.

Our fourth hypothesis (**H4**) was that, if carryover of social conventions into human-robot interactions occurred, conventionalized ISAs use would be higher in conventionalized scenarios. As previously discussed, we did indeed observe ISAs being used much more frequently in the conventionalized task context. While our results are not in themselves sufficient to establish a causal link between carryover of social conventions and use of ISAs, we believe they will motivate future investigation of such a hypothesis, and at the least demonstrate the importance of handling ISAs for robots intended to operate in conventionalized task contexts.

**Table 2: Subjective Results: Effects of Dialogue Condition**

Question	$\chi^2_1$	p	$\bar{M} \pm \sigma_M$	$\bar{U} \pm \sigma_U$
1 The robot was easy to interact with (from 1 to 10, “strongly disagree” to “strongly agree”)	14.74	<.001	8±2.35	10± 0.76
2 How would you rate the ease of interacting with the robot? (from 1 to 10, “easy” to “hard”)	14.98	<.001	2± 2.60	1±0.52
3 How would you rate the difficulty of the task? (from 1 to 10, “easy” to “hard”)	6.57	.01	2± 2.32	1 ±1.85
4 The robot was responsive to my commands (from 1 to 10, “strongly disagree” to “strongly agree”)	13.33	<.001	9 ±2.08	10 ±1.41
5 The robot was cooperative (from 1 to 10, “strongly disagree” to “strongly agree”)	8.33	<.01	10 ± 1.90	10 ±0.50
6 The robot was helpful (from 1 to 10, “strongly disagree” to “strongly agree”)	5.20	.02	9±2.30	10±1.94
7 The robot was annoying (from 1 to 10, “strongly disagree” to “strongly agree”)	8.21	<.01	2± 2.29	1 ±1.12
8 How would you rate the robot’s level of comprehension (from 1 to 10 “low” to “high”)	5.27	.02	7±2.62	10±2.60
9 Did you feel that the robot understood what you were saying? (from 1 to 10, “no” to “yes”)	10.49	.001	8± 2.72	10 ±1.89
10 Did the robot meet your expectations? (from 1 to 10, “no” to “yes”)	6.42	.01	8±2.73	10 ±2.22

Results from Ordinal Regression Analysis of Variance. N=45 ( $N_M = 24, N_U = 25$ ). M = MIS; U = UND.

**Table 3: Appropriateness and Effectiveness of Utterance Forms: No Differences by Context or Effectiveness**

Category	Form	$\bar{x} \pm \sigma_x$
Appropriateness	Direct Command	10±2.90
Appropriateness	Agent/Preparatory-directed Statement	4.5±4.88
Appropriateness	Agent/Sincerity-directed Statement	7±3.38
Appropriateness	Patient/Sincerity-directed Question	1±2.23
Appropriateness	Agent/Preparatory-directed Suggestive Question	1±2.00
Appropriateness	Patient/Preparatory-directed Suggestive Question	1±1.87
Appropriateness	Patient/Preparatory-directed Suggestive Statement	5±3.94
Effectiveness	Direct Command	10±2.59
Effectiveness	Agent/Preparatory-directed Question	8±3.46
Effectiveness	Agent/Preparatory-directed Statement	5±3.37
Effectiveness	Agent/Sincerity-directed Question	2±2.61
Effectiveness	Agent/Sincerity-directed Statement	9±3.20
Effectiveness	Patient/Sincerity-directed Question	2±2.91
Effectiveness	Patient/Sincerity-directed Statement	9±3.13
Effectiveness	Agent/Preparatory-directed Suggestive Question	2±2.57
Effectiveness	Agent/Preparatory-directed Suggestive Statement	9±2.96
Effectiveness	Patient/Preparatory-directed Suggestive Question	2±2.38
Effectiveness	Patient/Preparatory-directed Suggestive Statement	7±3.17

N=45. All questions were delivered as Likert items from 1 to 10, “No” to “Yes”.

**Table 4: Appropriateness and Effectiveness of Utterance Forms: Differences by Context**

Category	Form	$\chi^2_1$	p	$\bar{R} \pm \sigma_R$	$\bar{D} \pm \sigma_D$
Appropriateness	Agent/Preparatory-directed Question	7.37	<.01	10±2.78	6±4.01
Appropriateness	Patient/Preparatory-directed Question	16.50	<.001	9.5±3.55	1±2.90
Appropriateness	Patient/Preparatory-directed Statement	19.74	<.001	10±2.87	3.5±3.00
Appropriateness	Patient/Sincerity-directed Statement	4.81	.03	10±2.61	8.5 ±3.46
Effectiveness	Patient/Preparatory-directed Question	10.86	<.001	9±3.61	3±2.80
Effectiveness	Patient/Preparatory-directed Statement	21.55	<.001	10±2.63	3.5±2.84

Results from Ordinal Regression Analysis of Variance. N=45 ( $N_M = 24, N_U = 25$ ). R=REST, D=DEMO.

**Table 5: Appropriateness and Effectiveness of Utterance Forms: Differences by Dialogue Condition**

Category	Form	$\chi^2_1$	p	$\bar{M} \pm \sigma_M$	$\bar{U} \pm \sigma_U$
Appropriateness	Agent/Preparatory-Directed Suggestive Statement	4.15	.04	8±3.56	5 ±3.15
Effectiveness	Patient/Preparatory-directed Question	5.18	.02	2.5±3.65	6±3.31

Results from Ordinal Regression Analysis of Variance. N=45 ( $N_M = 24, N_U = 25$ ). M = MIS; U = UND.

Our fifth hypothesis (H5) was that humans would prefer to continue using ISAs even when robots demonstrated a fundamental inability to understand them, yielding a high ISA use rate in both of our dialogue conditions. While we did find that ISAs were consistently used in both dialogue conditions, and anecdotally observed a strong resistance to direct command use on the part of some participants, we do not believe that our results actually support this hypothesis. As we previously discussed, we believe that if the robots' capabilities were made transparent to participants from the start, our results may have looked different (such explicit instruction about a robot's capabilities will, however, not be possible in many future applications domains and can thus not be assumed).

Our sixth hypothesis (H6) was that, because of this hypothesized high use rate even with a robot fundamentally unable to understand ISAs, participants interacting with such robots would be significantly less efficient in accomplishing their task goals than would participants interacting with robots that *were* capable of understanding ISAs. Our results supported this hypothesis, with participants in the misunderstanding condition needing to use significantly more utterances to accomplish their task than participants in the understanding condition. However, while our results statistically supported our hypothesis, this may have simply been because the use of even one or two ISAs in the MIS condition significantly increased the number of utterances required in these small interactions. Additional study is thus needed to determine whether, when the robots' capabilities are transparent, and when the robots are interacted with over longer periods of time, these patterns would still hold out. Finally, our seventh hypothesis (H7) was that robots unable to understand ISAs would be perceived less favorably by their human interlocutors than would robots capable of understanding ISAs. Our results support this hypothesis, with robots in the UND condition being statistically significantly preferred to those in the MIS condition along ten dimensions of preference. However, for none of these dimensions was there a significant difference in valence; along all dimensions of analysis, participants' perceptions of the robots were uniformly positive.

Taken together, the results of these experiments suggest a significant need for robots engaging in task-based human-robot dialogue interactions to be able to understand indirect speech acts. Specifically, our results suggest that failing to understand indirect speech acts could result in an expected utterance error rate as high as 46% (the mean frequency of ISAs among task relevant utterances) – a number that is clearly unacceptably high and prohibitive for task-based interactions. In addition, the frequency of ISA use observed in our MIS dialogue condition suggests that it may be necessary to consider whether robots incapable of understanding indirect speech acts should be used at all in contexts in which they will usually interact with naive users (i.e., who cannot be instructed or trained before interaction). This is particularly true in contexts with conventionalized social norms for which we observed an indirect speech act use rate of 69%.

Our results suggest a new hypothesis that will need to be investigated in future work: *in conventionalized contexts, ISA use may be more strongly determined by what is normatively expected to be appropriate or inappropriate than by what is expected to be effective.* Furthermore, our results suggest that the high utterance error rates of robots incapable of understanding ISAs in conventionalized tasks

will ultimately not only negatively affect human perceptions of such robots – finding such robot more difficult to interact with, less responsive, less cooperative, more annoying, and less comprehending – but ultimately be prohibitive for employing such robots.

## 4.2 Future Work

Our results suggest several areas in which additional future research is needed. Most importantly, the scope of the research questions investigated in our experiments must be extended beyond the scope of single interactions. It is possible that in longer-term interactions or sequences of repeated interactions the dynamics of human-robot conversation could change significantly. Understanding these dynamics will be of paramount importance for designing natural language understanding systems for robots (e.g., in order to be better understand the effects of possible humans lexical entrainment in response to robots' utterances). What is more, in our experiment, we observed some participants continuing to use ISAs even when the robots clearly and repeatedly failed to understand indirect language. Future research should thus examine whether participants would be able to refrain from using ISAs if explicitly told not to, or explicitly told that a robot was unable to comprehend them.

We would also like to examine how facets of a robot's design affect linguistic interaction patterns with that robot. In this experiment, we used a single robot with a single voice. While it seems unlikely that ISA use would have differed if the robot had had a different voice or appearance, this needs to be investigated. And as mentioned before, future work will be needed to investigate the new hypothesis our work has generated: that in conventionalized contexts, ISA use may be more strongly determined by what is expected to be appropriate or inappropriate in such scenarios rather than by what is expected to be effective; a hypothesis which, if true, could have strong implications for the design of language-enabled robots. Ideally, future research on natural language understanding would develop mechanisms whereby robots can *automatically* learn to understand ISAs in general, or to understand specific newly encountered ISA forms, which would allow robots to adapt to their human users instead of requiring the opposite. Finally, the issues discussed in this section are of importance to a wide variety of intelligent agents beyond robots. As such, it would be interesting to investigate the extent to which the effects found in this section depend upon the type of agent used, and whether these patterns hold for non-embodied and non-situated agents as well.

## 5 CONCLUSION

We have presented the results of the first systematic investigation of the effects of capability and context on human indirect speech act use in interactions with robots. Our results suggest that humans will consistently and frequently use ISAs when interacting with robots in both conventionalized and unconventionalized contexts. This emphasizes the importance of both past previous research on interpreting ISAs, and of future work towards understanding the requirements for robots to learn new ISA forms.

## ACKNOWLEDGMENTS

This work was supported in part by ONR grant #N00014-14-1-0149.

## REFERENCES

- [1] Theodore W Anderson and Donald A Darling. 1954. A Test of Goodness of Fit. *J. Amer. Statist. Assoc.* 49, 268 (1954), 765–769.
- [2] David F Bauer. 1972. Constructing Confidence Sets using Rank Statistics. *J. Amer. Statist. Assoc.* 67, 339 (1972), 687–690.
- [3] Timo Baumann and Felix Lindner. 2015. Incremental Speech Production for Polite and Natural Personal-Space Intrusion. In *Proceedings of the International Conference on Social Robotics (ICSR)*. 72–82.
- [4] Maxwell Bennett, Tom Williams, Daria Thames, , and Matthias Scheutz. 2017. Differences in Interaction Patterns and Perception for Teleoperated and Autonomous Humanoid Robots. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*.
- [5] Luciana Benotti and Patrick Blackburn. 2016. Polite Interactions with Robots. *What Social Robots Can and Should Do: Proceedings of Robophilosophy 2016/TRANSOR 2016* 290 (2016), 293.
- [6] Jürgen Brandstetter, Clay Beckner, Eduardo Benitez Sandoval, and Christoph Bartneck. 2017. Persistent Lexical Entrainment in HRI. In *Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 63–72.
- [7] Gordon Briggs and Matthias Scheutz. 2013. A Hybrid Architectural Approach to Understanding and Appropriately Generating Indirect Speech Acts. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*.
- [8] Gordon Briggs and Matthias Scheutz. 2014. How Robots Can Affect Human Behavior: Investigating the Effects of Robotic Displays of Protest and Distress. *International Journal of Social Robotics* 6 (2014), 1–13. Issue 2.
- [9] Gordon Briggs and Matthias Scheutz. 2016. The Pragmatic Social Robot: Toward Socially-Sensitive Utterance Generation in Human-Robot Interactions. In *Proceedings of the AAAI Fall Symposium Series on Artificial Intelligence for Human-Robot Interaction (AI-HRI)*. 12–15.
- [10] Gordon Briggs, Tom Williams, and Matthias Scheutz. 2017. Enabling Robots to Understand Indirect Speech Acts in Task-Based Interactions. *Journal of Human-Robot Interaction* (2017).
- [11] R. H. B. Christensen. 2015. Ordinal—Regression Models for Ordinal Data. (2015). R package version 2015.6-28. <http://www.cran.r-project.org/package=ordinal/>.
- [12] Manfred Eppel, Sean Trott, and Jerome Feldman. 2016. Exploiting Deep Semantics and Compositionality of Natural Language for Human-Robot-Interaction. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.
- [13] Mary Ellen Foster, Andre Gaschler, Manuel Giuliani, Amy Isard, Maria Pateraki, and Ronald P A Petrick. 2012. Two People Walk into a Bar : Dynamic Multi-Party Social Interaction with a Robot Agent. In *ICMI*. 14–21.
- [14] John Fox and Sanford Weisberg. 2011. *An R Companion to Applied Regression* (second ed.). Sage, Thousand Oaks CA.
- [15] Manuel Giuliani, Ronald Petrick, Mary Ellen Foster, Andre Gaschler, Amy Isard, Maria Pateraki, and Markos Sigalas. 2013. Comparing task-based and socially intelligent behaviour in a robot bartender. In *Proceedings of the 15th ACM on International conference on multimodal interaction*. ACM, 263–270.
- [16] John J Godfrey, Edward C Holliman, and Jane McDaniel. 1992. SWITCHBOARD: Telephone Speech Corpus for Research and Development. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.
- [17] Juergen Gross and Uwe Ligges. 2015. *nortest: Tests for Normality*. <https://CRAN.R-project.org/package=nortest> R package version 1.0-4.
- [18] Elizabeth A. Hinkelman and James F. Allen. 1989. Two Constraints on Speech Act Ambiguity. In *Proceedings of the 27th Annual Meeting on Association for Computational Linguistics (ACL)*. Stroudsburg, PA, USA, 212–219.
- [19] Takamasa Iio, Masahiro Shiomi, Kazuhiko Shinozawa, Takahiro Miyashita, Takaaki Akimoto, and Norihiro Hagita. 2009. Lexical Entrainment in Human-Robot Interaction: Can Robots Entrain Human Vocabulary?. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.
- [20] Takamasa Iio, Masahiro Shiomi, Kazuhiko Shinozawa, Katsunori Shimohara, Mitsunori Miki, and Norihiro Hagita. 2015. Lexical Entrainment in Human Robot Interaction. *International Journal of Social Robotics* 7, 2 (2015), 253–263.
- [21] Peter Kahn, Hiroshi Ishiguro, Brian Gill, Takayuki Kanda, Nathan Freier, Rachel Severson, Jolina Ruckert, and Solace Shen. 2012. Robovie, you'll have to go into the closet now: Children's social and moral relationships with a humanoid robot. *Developmental Psychology* 48 (2012), 303–314.
- [22] Ross A Knepper, Christoforos I Mavrogiannis, Julia Proft, and Claire Liang. 2017. Implicit Communication in a Joint Action. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 283–292.
- [23] Robin Tolmach Lakoff. 1973. *The logic of Politeness: Minding your p's and q's*.
- [24] Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. 2010. Receptionist or Information Kiosk: How do People Talk with a Robot?. In *Proceedings of the 2010 ACM conference on Computer Supported Cooperative Work (CSCW)*. ACM, 31–40.
- [25] Min Kyung Lee and Maxim Makatchev. 2009. How do People Talk with a Robot?: An Analysis of Human-Robot Dialogues in the Real World. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. ACM, 3769–3774.
- [26] Rui Liu and Xiaoli Zhang. 2017. Systems of Natural-Language-Facilitated Human-Robot Cooperation: A Review. *arXiv preprint arXiv:1701.08269* (2017).
- [27] Salvatore S. Mangiafico. 2015. An R Companion for the Handbook of Biological Statistics. (2015).
- [28] Nikolaos Mavridis. 2015. A review of Verbal and Non-Verbal Human-Robot Interactive Communication. *Robotics and Autonomous Systems* 63, P1 (2015), 22–35. <https://doi.org/10.1016/j.robot.2014.09.031>
- [29] R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- [30] Daniel J Rea, Yan Wang, and James E Young. 2015. Check Your Stereotypes at the Door: An Analysis of Gender Typecasts in Social Human-Robot Interaction. In *International Conference on Social Robotics*. Springer, 554–563.
- [31] Maha Salem, Micheline Ziadee, and Majd Sakr. 2014. Marhaba, How may I help you?: Effects of Politeness and Culture on Robot Acceptance and Anthropomorphization. In *Proceedings of the 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM, 74–81.
- [32] George Saon, Tom Sercu, Steven J. Rennie, and Hong-Kwang Jeff Kuo. 2016. The IBM 2016 English Conversational Telephone Speech Recognition System. *Proceedings of the International Conference on Spoken Language Processing (INTERSPEECH)* (2016). <http://arxiv.org/abs/1604.08242>
- [33] Paul Schermerhorn, James F Kramer, Christopher Middendorff, and Matthias Scheutz. 2006. DIARC: A Testbed for Natural Human-Robot Interaction.. In *Proceedings of the 18th AAAI Conference on Artificial Intelligence*.
- [34] Paul Schermerhorn, Matthias Scheutz, and Charles R Crowell. 2008. Robot Social Presence and Gender: Do Females View Robots Differently than Males?. In *Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.
- [35] Matthias Scheutz. 2013. Computational Mechanisms for Mental Models in Human-Robot Interaction. In *Proceedings of the International Conference on Virtual, Augmented and Mixed Reality*.
- [36] Matthias Scheutz, Gordon Briggs, Rehj Cantrell, Evan Krause, Tom Williams, and Richard Veale. 2013. Novel mechanisms for natural human-robot interactions in the diarc architecture. In *Proc. of the AAAI Workshop on Intelligent Robot Systems*.
- [37] Marc Schröder and Jürgen Trouvain. 2003. The German text-to-speech synthesis system MARY: A tool for research, development and teaching. *International Journal of Speech Technology* 6, 4 (2003).
- [38] John R Searle. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Vol. 626. Cambridge university press.
- [39] John R Searle. 1975. Indirect Speech Acts. *Syntax and semantics* 3 (1975).
- [40] John R. Searle. 1976. A Classification of Illocutionary Acts. *Language in Society* 5, 1 (1976), 1–23.
- [41] Megan Strait, Priscilla Briggs, and Matthias Scheutz. 2015. Gender, more so than Age, Modulates Positive Perceptions of Language-Based Human-Robot Interactions. In *Proceedings of the 4th International Symposium on New Frontiers in Human-Robot Interaction*.
- [42] R Studio. 2012. RStudio: integrated development environment for R. *RStudio Inc, Boston, Massachusetts* (2012).
- [43] Deborah Tannen. 1981. Indirectness in Discourse: Ethnicity as Conversational Style. *Discourse processes* 4, 3 (1981), 221–238.
- [44] Moritz Tenorth and Michael Beetz. 2013. KnowRob: A Knowledge Processing Infrastructure for Cognition-Enabled Robots. *The International Journal of Robotics Research* 32, 5 (2013), 566–590.
- [45] Sean Trott and Benjamin Bergen. 2017. A Theoretical Model of Indirect Request Comprehension. In *Proceedings of the AAAI Fall Symposium Series on Artificial Intelligence for Human-Robot Interaction (AI-HRI)*.
- [46] Yan Wang. 2014. Gendering Human-Robot Interaction: exploring how a person's gender impacts attitudes toward and interaction with robots. (2014).
- [47] Tom Williams, Gordon Briggs, Bradley Oosterveld, and Matthias Scheutz. 2015. Going Beyond Command-Based Instructions: Extending Robotic Natural Language Interaction Capabilities. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*.
- [48] Sabrina Wilske and Geert-Jan Kruijff. 2006. Service Robots Dealing with Indirect Speech Acts. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*.
- [49] Christopher Winship and Robert D Mare. 1984. Regression Models with Ordinal Variables. *American Sociological Review* (1984), 512–525.