

Unpretty Please: Ostensibly Polite Wakewords Discourage Politeness in both Robot-Directed and Human-Directed Communication

Ruchen Wen
rwen@mines.edu
Colorado School of Mines
Golden, Colorado, USA

Sebastian Fauré
sebastianfaure@ufl.edu
University of Florida
Gainesville, Florida, USA

Brandon Barton
brandonbarton@mines.edu
Colorado School of Mines
Golden, Colorado, USA

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

ABSTRACT

For enhanced performance and privacy, companies deploying voice-activated technologies such as virtual assistants and robots are increasingly tending toward designs in which technologies only begin attending to speech once a specified *wakeword* is heard. Due to concerns that interactions with such technologies could lead users, especially children, to develop impolite habits, some companies have begun to develop use modes in which interactants are required to use ostensibly polite wakewords such as “<Name> Please”. In this paper, we argue that these “please-centering” wakewords are likely to backfire and actually *discourage* polite interactions due to the particular types of lexical and syntactic priming induced by those wakewords. We then present the results of a human-subject experiment (n=90) that validates those claims.

KEYWORDS

Wakewords, Politeness, Indirect Speech Acts

ACM Reference Format:

Ruchen Wen, Brandon Barton, Sebastian Fauré, and Tom Williams. 2022. Unpretty Please: Ostensibly Polite Wakewords Discourage Politeness in both Robot-Directed and Human-Directed Communication. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '22)*, November 7–11, 2022, Bengaluru, India. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3536221.3556615>

1 INTRODUCTION

Voice activated and speech capable technologies have become increasingly prevalent, including stationary personal assistants (e.g., Google Home, Amazon Echo, and Apple HomePod), smartphone-based virtual assistants (e.g., Google Now, Amazon Alexa, and Siri), and voice-interactive robots (e.g. Pepper and Jibo). While much of the human-robot interaction (HRI) research literature on spoken language capabilities focuses on long-term a vision of a long-term

future of spoken language interaction grounded in fluid, flexible, mixed-initiative dialogue, successful deployments of these voice-interactive technologies – including voice-interactive social robots – have nearly uniformly used an interaction design centered on wakeword-based interaction, in which voice-based interactions are human-initiated and uniformly begin with a keyphrase (wakeword) such as “Hey Siri”, “Alexa”, or “Okay Google”.

One of the key benefits of wakeword-driven interaction is enhanced privacy and security, as the devices need not be truly “listening in” on conversations until the selected wakeword is heard, instead running reduced speech-recognition models only capable of picking out wakewords, and then switching to full speech-recognition models once the wakeword is detected. Recently, however, many journalists and the parents they have interviewed have raised concerns about the command-based interaction patterns required by these technologies, and thus whether these technologies could be leading users, especially children, to increasingly communicate in ways that are command-based and fundamentally impolite, not only with the devices but also with other people [26]. In response, some companies have implemented novel wakeword designs intended to encourage polite language use by interactants. Amazon, for example, deployed an optional “Alexa Please” mode in which children are encouraged through positive feedback to use “magic words” such as “please” and “thank you” [7].

We could foresee robot companies following this same tactful tack. We argue, however, that these designs may be more effective at temporarily appeasing consumers than actually addressing those consumers’ concerns. Specifically, we argue that the approach taken by companies like Amazon could backfire and result in *less* polite human-technology conversations because it is in conflict with what is actually known about politeness in the research literature. As we will discuss, the politeness literature suggests that merely saying “please” does not automatically make an utterance polite, and in fact the way in which interaction designs like “Alexa Please” encourage users to speak – using “Please” at the beginning of an utterance – is *negatively* correlated with perceived politeness, because sentence-frontal politeness is typically followed up with a face-threatening act in the form of a direct command. Accordingly, nudging users of social robots or other voice-activated technologies towards the use of “please” at the beginning of their sentences may be syntactically priming those users to continue their utterances

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).
ICMI '22, November 7–11, 2022, Bengaluru, India
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9390-4/22/11.
<https://doi.org/10.1145/3536221.3556615>

with direct commands, steering them away from utterance forms that are *positively* correlated with politeness, such as conventionalized Indirect Speech Acts, due to the syntactic awkwardness of using one of those forms after saying “<Name> please...”.

In this work, we experimentally assess whether wakewords like “Name please” actually discourage the use of other, more effective, politeness strategies, and whether alternatives recently proposed in the HRI literature [57], such as “Excuse me <Name>” could be more effective at achieving the ostensible goal of encouraging polite interactions. Our experiment builds off of a wakeword-based HRI paradigm used in previous work, as mediated through a novel chat-based research platform that facilitates exploration of our research questions while adhering to COVID-19 safety constraints [21].

2 BACKGROUND

2.1 The Persuasive Influence of Technologies

As described in the previous section, a number of articles in the popular press have raised concerns about the potential influence that voice-interactive technologies may have on users (especially vulnerable users such as children) and the way they think and act. To be clear, while the articles raising these concerns present the perspectives of journalists and parents, those concerns are well justified by the scientific literature. The systems of social and moral norms that govern human behavior can be readily influenced by agents with appropriate standing as community members, because these norms are dynamic and malleable [25], and because the very way that principles become (and stay) norms is by their definition, communication, and enforcement by community members [53]. While the potential for normative influence is typically considered with respect to human agents, technologies, broadly defined, have been shown to have potential for such influence as well, with technologies changing the way we perceive, think about, and act within the world in morally relevant ways [53]. Moreover, it has been demonstrated that voice-interactive technologies like robots have unique potential for influence [5, 6, 18, 19, 27, 38, 40–42, 49, 50], especially language-capable robots [14, 31, 34, 59], due to their joint perception as both social and moral agents [30, 32].

Many previous concerns regarding the persuasive influence of interactive technologies have centered on their ability to persuade interactants to pursue various courses of action (or, more broadly, influence their beliefs, desires, and intentions). However, the concerns raised by parents regarding voice-interactive technologies’ influence over users’ linguistic behaviors is instead a more subtle linguistic influence, relating to these technologies’ abilities to engage in various forms of lexical, syntactic, and semantic priming. Models of dialogue such as Pickering and Garrod [39]’s *Interactive Alignment* model suggest that dialogue is a highly negotiated process, with interlocutors subtly influencing each other’s (and their own future) linguistic choices at the phonetic, lexical, syntactic, and semantic levels, by activating mental representations maintained at these different levels through different sorts of priming, as demonstrated by phenomena such as *lexical entrainment* [13], where speakers exert mutual influence to converge on shared terminology over the course of a dialogue.

Parents’ concerns over the influence of voice-activated technologies’ ability to influence children’s linguistic behaviors may be best

understood through the lens of these linguistic priming phenomena for two reasons. First, parents’ fears focus on the ways that these technologies may be encouraging users to communicate in the form of commands; a fear that is motivated (1) by the way users *need to speak* to have their requests met by these technologies (based on the syntactic patterns these technologies are able to identify *and* the syntactic patterns that most fluidly and naturally follow the wakewords chosen by technology companies); and (2) by the way that needing to speak in such a way might carry over into future linguistic interactions with humans based on these sorts of priming effects. Second, it is productive to think about these concerns in terms of linguistic priming effects because there is a host of evidence in the literature suggesting that voice-activated technologies are indeed able to trigger these types of effects. In the HRI literature, for example, there has been a variety of research demonstrating robots’ ability to trigger effects such as lexical entrainment [9, 10, 28, 29]. Similar results have been found in the broader HCI literature as well [11, 12].

2.2 Using Persuasion to Exert Positive Moral Influence

Thus far we have described voice-interactive technologies’ potential for persuasion and influence as a risk to be considered and avoided; however, researchers in the HRI and Human-Computer Interaction fields have long recognized that these forms of persuasion and influence represent not only risks to be avoided, but opportunities for positively influencing users and their broader moral ecologies. For example, Zhu et al. [60] have argued that, from a Confucian ethical perspective, interactive technologies such as robots can actively invite their human interactants to cultivate virtues and develop their moral selves. Companies like Amazon have attempted to respond to parents’ concerns about the potential linguistic influence of their technologies through interaction designs that attempt to not only ameliorate those concerns about their technologies, but moreover, to positively influence the moral ecologies into which their technologies are embedded [3, 7]. As discussed in the Introduction, Amazon in particular has done so by deploying an “Alexa Please” mode that actively rewards children for using “Please” in their communications with Alexa. This focus on rewarding positive behaviors would, if successful, have the outcome of exerting positive influence on the moral ecosystem of Alexa’s users, especially if the polite language patterns they are incentivized to use with Alexa carry over into their interactions with other people. Unfortunately, interaction designs like “Alexa Please” narrowly encourage a very particular type of polite language use [8]. That is, they encourage the mere use of the word “please”, and moreover, as the name of the mode suggests, they encourage the use of “please” as part of the wakeword at the beginning of a user’s utterance. This particular strategy is potentially problematic when understood through the lens of politeness theory.

2.3 Politeness

Politeness is often conceptualized with respect to the way that actors negotiate threats to each others’ *Face*, i.e., the public self-concept that social agents seek to preserve and enhance [16]. Face, as defined by Brown et al. [16], has two aspects: Positive Face (i.e.,

self-image and wants) and Negative Face (i.e., freedom of action and from imposition). From a politeness theoretic perspective, the parental concerns described above are primarily oriented around Negative Face; if voice-interactive technologies are priming users to express their desires in terms of commands, they may be priming those users to use linguistic constructions that threaten the Negative Face of others by violating listener’s freedom from imposition. This lens is particularly apt for considering voice-interactive robots, as recent theoretical work in HRI has argued that politeness theory and its notions of face are central to understanding and conceptualizing robot social agency and the unique persuasive power robots hold [32].

Some researchers have attempted to computationally model the ways that different linguistic patterns can have subtle influences on whether a sentence is perceived as polite or not, operating within this politeness theoretic lens [20]. Notably, that work has revealed that “please”, while stereotypically viewed as polite, in fact has drastically varying impacts on perceived politeness depending on how it is used. The precise type of please usage that Amazon encourages with “Alexa Please” is actually, according to the work of Danescu-Niculescu-Mizil et al. [20], *negatively* correlated with politeness. We argue that this is because when “please” appears at or near the front of a sentence, it is usually followed by a direct command. This is potentially unsurprising since the default mode of interaction with technologies like Amazon’s Alexa outside of the “Alexa Please” mode are in fact direct commands.

This default emphasis on direct commands is critical not only because direct commands are inherently face threatening and impolite, but moreover because the type of “please” use that is actually positively correlated with politeness is sentence-medial please-use, typically in the context of “please” being used to intensify the sentiment conveyed by *other* politeness strategies, such as *indirect* speech acts (ISAs) [46] (such as when saying “Could you *please* close the window”). To reiterate, “please” on its own is not inherently polite, but is only polite when it is used to intensify the politeness conveyed by linguistic constructions such as ISA.

It will thus be helpful to briefly describe the nature of ISAs. An ISA is an utterance whose literal meaning mismatches its intended meaning. For example, if someone asks “Can you tell me what time is it?”, people can immediately understand that while this utterance is *literally* a yes-or-no question, interpreting the utterance as such would violate fundamental assumptions underlying cooperative communication (in this case, the *sincerity* condition, as it is unlikely for someone to genuinely want to know, outside of certain precise circumstances, whether someone is merely aware of the time) and that instead, this utterance is likely meant as an *indirect* request for the information that is highlighted to already in fact be known to the listener, i.e., the current time. ISAs are one of the most widely researched topics in the natural language pragmatics literature [2, 46] (especially within the broader context of Speech Act Theory [45, 47]), and are well understood as one of the most effective methods by which speakers can avoid face threat [1, 24], and as such have recently attracted significant attention within the HRI community [15, 23, 43, 44, 48, 52, 55, 56, 58].

It is worth noting that, the usage of ISAs is highly context-sensitive. For example, using ISAs as a politeness strategy when people are facing a great potential for harm or under high time

pressure could be ineffective [37]. Moreover, linguistic politeness norms are also culturally dependent. While in the US, people tend to use ISAs frequently in contexts with strong social conventions (e.g., restaurants) [58]; in Korea, indirect speech is not a commonly used linguistic politeness strategy [48]. However, in this particular study, we set the experimental task in a restaurant context and the participants were all from the United States. Thus, indirectness is a key measure of politeness in our study.

2.4 Designing Wakewords to Encourage Politeness

The considerations described above suggest that if we truly want to encourage users to be polite (i.e., to use robots’ persuasive influence to cultivate user’s moral ecosystems by nudging users in ways that encourages them to treat others with respect and civility), then the current approach adopted by companies like Amazon (i.e., encouraging sentence-frontal please-usage) may in fact backfire and lead to less politely perceived language. Instead, the considerations described above suggest that a more promising approach could be to explicitly try to nudge users to use the types of linguistic forms that are actually positively associated with perceived politeness, such as Indirect Speech Acts.

This approach could also be beneficial in that it may be significantly easier to prime users to use ISAs than arbitrary keywords such as “please.” Recent work in the HRI literature suggests that humans automatically tend toward indirect speech act use with both humans and robots, especially in contexts with highly conventionalized sociocultural politeness norms [58].

Within the HRI literature, there has been some previous work on encouraging people to be polite in exactly this way, demonstrating that in the context of live, in-person interactions, wakewords such as “Excuse me <Name>” (a detailed justification for this particular wakeword can be found in [57]) can be highly effective at encouraging users to use politeness strategies such as ISAs when speaking with robots [57]. However, that past research has not actually demonstrated whether these wakewords are any better than “<Name> please”, and as such, the concerns about the potential politeness-discouraging syntactic priming and sentence-frontal please use in Wakeword designs has not actually been empirically verified. Moreover, it isn’t clear whether lexical, syntactic, or semantic priming effects may actually lead to carry-over between robot-directed language and human-directed language, that is, it isn’t clear whether the wakewords used to address technologies actually have any impact on how people choose to address each other. In this work, we aim to investigate these unanswered questions, by assessing several key research hypotheses.

2.5 Hypotheses

In this paper, we examine the efficacy of “Excuse me <Name>” relative to the baseline “Hey <Name>” and the alternative “<Name> Please”. Specifically, we test the following concrete hypotheses:

- H1** Requiring the use of a polite wakeword (i.e., Excuse me <Name>) will result in increased robot-directed politeness.

- H2** Politeness resulting from the use of a polite wakeword (i.e., Excuse me <Name>) will carry over to the following human-human communication and result in increased human-directed politeness.
- H3** Requiring the use of “Excuse me <Name>” will result in more robot-directed politeness than requiring the use of “<Name> Please”.
- H4** The differences of politeness resulting from the use of “Excuse me <Name>” and “<Name> Please” will carry over to the following human-human communication and result in human-directed politeness.

3 METHODOLOGY

3.1 Experimental Design and Procedure

To assess these experimental hypotheses, we performed an IRB-approved between-subjects experiment in which participants interacted with an autonomous system in the context of a hypothetical food delivery service application. Participants were randomly assigned to three experimental conditions associated with our three wakewords of interest (“Excuse me <Name>”, “<Name> Please”, and “Hey <Name>”), which, as we will describe below, participants were required to use when addressing robot interactants.

Participants were intentionally not given any pre-defined restaurant position such as “cashier” to avoid any pre-existing biases towards or against politeness norms that might come with different workplace roles. Instead, the participant’s role in the restaurant was merely to serve as the communication medium between a robot and a pair of delivery drivers.

Due to the COVID-19 Pandemic and its documented impacts on HRI research praxis [21], we conducted this interactive experiment online. Instead of requiring participants to try to interact with the robot remotely via video/audio communication applications, we chose to develop a text-based online chatting interface to avoid serious speech recognition errors and communication latency.

3.1.1 Demographics and Briefing. After providing informed consent and demographic information, participants received an introductory experimental briefing, and watched a series of videos introducing them to the NAO Robot, represented as the “Foodbot”, and to two human delivery workers (car-driving “James” and bike-riding “Peter”). These video introductions were used to give participants familiarity with the robot and humans, and to reinforce task requirements.

In the first video, the NAO was introduced as the “Foodbot,” who explained that the participants should direct the details of the food orders to them (the Foodbot). The Foodbot notes in the video that its text recognition can be activated by starting the sentence with a certain phrase (i.e., the wakeword associated with the participant’s condition). Participants were required to use the same wakeword throughout the entire task to address the Foodbot.

In the second video, the humans were introduced to the participants. James is introduced as the delivery driver who takes the responsibility of the food orders to be delivered by car, and Peter is introduced as responsible for orders to be delivered by bike. In both introductions, James and Peter indicated not only which orders they were responsible for (i.e., those optimally delivered by Car or by

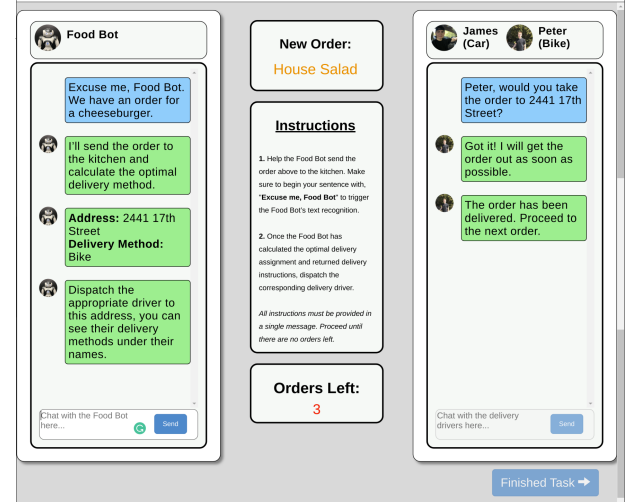


Figure 1: Dual chat box design used in this experiment. The robot chat box is on the left, while the human chat boxes are on the right. Instructions to participants, including incoming orders, are shown between the two chat boxes.

Bike) but also that when messaging with them, participants would need to send them the exact addresses provided by the Foodbot.

After these video-based introductions and instructions, the instructions were reiterated in text form. These written instructions further clarified the details of food order information flow to the “Foodbot” and delivery information flow to delivery drivers based on optimal delivery method.

3.1.2 Experimental Task. Participants were then introduced to and completed a series of tasks assisting the Foodbot, James, and Peter with food orders through automated chat interfaces developed specifically for this experiment. As shown in Figure 1, the chat box on the left side of the screen was used to communicate with the “Foodbot,” while the chat box on the right was used to communicate with delivery team consisting of “James,” or “Peter.” This interface required certain conditions to hold for text messages to be successfully sent. Using this interface, each participant went through the following workflow four times: (1) they received a new order in the center of the experiment panel; (2) they relayed this order to the Foodbot, who responded with the optimal delivery method and information about the delivery address; (3) they relayed this delivery address to the delivery worker associated with the specified delivery message.

3.1.3 Messaging Requirements. To ensure that participants were actually engaging with the experiment and following instructions, participants’ messages were only sent if they fulfilled certain requirements.

When communicating with the Foodbot, participants were unable to send messages if they did not actually include the name of the dish they were instructed to relay to the Foodbot. Communications with Foodbot thus took forms such as “Excuse me Foodbot, we have an order for a cheese pizza”. Any messages sent to the Foodbot that did not contain both the required wakeword and the

instructed Food Item resulted in error messages being returned to the user, as shown in Figure 3. Otherwise, if the condition-specific wakeword and task-specific food order appeared in the message, the user’s chat message would be successfully sent, and the Foodbot would respond appropriately, as shown in Figure 1.

When communicating with the human delivery workers, participants were instead required to include only the name of the driver they intended to contact, and the address the order was to be brought to. Communications with James and Peter thus took forms such as “Peter, would you take the order to 2441 17th Street?”. Any messages sent to the Drivers that did not contain both a driver name and the specified address resulted in error messages being returned to the user, as shown in Figure 3. Otherwise, if a driver’s name and the delivery address appeared in the message, the user’s chat message would be successfully sent, and the driver would respond appropriately, as shown in Figure 1.

3.1.4 Debriefing. After four such rounds of food orders, participants completed a two question follow up survey described in the next section. Participants were then thanked for their time and received a completion code for compensation.



Figure 2: Images from Robot and Human Introduction Videos.

3.2 Measures

Demographics. At the beginning of the experiment, participants were asked to provide their biographical information, including their age, gender, prior experience with robots/AI and prior experience with working in the food industry.

Objective Measures. During the experiment, participants’ chat messages toward both human and robot teammates were recorded,

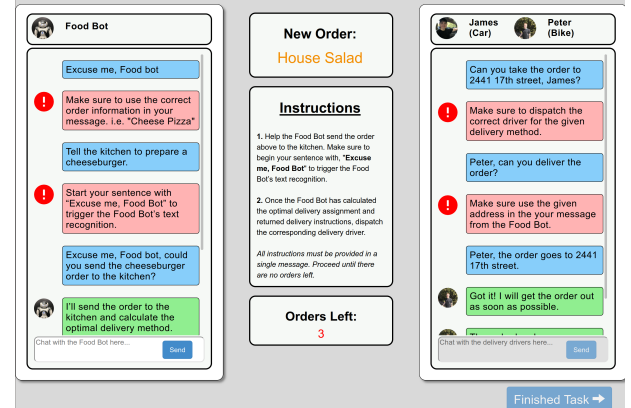


Figure 3: Error messages used in the experiment to ensure participants were genuinely engaging with the experiment and were attending and adhering to experimental instructions.

and were later annotated as follows. Two annotators coded participants’ utterances for common patterns of politeness as described in Danescu-Niculescu-Mizil [20] (see also [16]), including whether or not they were phrased as Indirect Speech Acts – in particular, conventional Indirect Requests [46] (e.g., “Could you <X>”, “Would you mind <X>”, “I need <X>”) – and whether they contained evidence of gratitude (e.g. “Thank you”) or apologies (e.g. “Sorry to bother you, but...”). In cases where the two annotators disagreed, a third annotator vote was used to resolve the conflict. These annotations were used to calculate average use of polite versus non-polite human-directed and robot-directed language for each participant.

Subjective Measures. At the end of the experiment, participants were asked to provide their feedback on the experience of chatting with the robot and with the human delivery team.

3.3 Participants

90 participants (74 female, 15 male, 1 NA) were recruited from Prolific. Participant ages ranged from 18 to 49 years old ($M=23.08$, $SD=5.715$)¹. Most participants (94.44%, 85 participants) reported little to no experience with robots and artificial intelligence, while other 5 participants provided a self-assessment of 4 or 5 out of 7, which indicate that they have some formal training in robotics and/or AI (e.g., university classes). 53 participants reported little to no experience with working in the food industry, 23 participants reported that they have some formal experience with working in the food industry, while the other 14 participants provided a self-assessment of 6 or 7 out of 7, which indicate that they have/had a career in the food industry.

These participants were randomly assigned to the three experimental conditions, resulting in 31 participants in the “Excuse me”

¹The age and gender distribution of our participants reflects a widespread phenomenon across Prolific in late summer 2021, during which Prolific became suddenly and widely popularized on TikTok, with around 30,000 new users (mostly women in their 20s) signing up to participate in experiments [17]. Our experiment was launched the day after the posting of the TikTok video responsible for this wave of new participant signups.

condition, 30 participants in the “Hey” condition, and 29 participants in the “Please” condition. Participants were paid \$3 each for participating in the study.

3.4 Analysis

About half of participants did not type in full sentences, and instead typed messages that only included keywords. When telling the Foodbot what food to prepare, for example, these participants would use messages like “Excuse me Foodbot cheeseburger”; and when dispatching the corresponding delivery drivers to the assigned address, they would use messages like “James 347 Green Avenue”. We will thus report on two versions of each of our analyses. In the first, we analyze the complete collected dataset containing data from all participants ($N=90$). In the second, we analyze a reduced dataset from which those keyword-only participants were excluded ($N=46$). The data and experimental materials are available in our OSF repository, at <https://osf.io/bke78/>.

Both analyses were performed using the JASP software [33], through which Bayesian Analyses of Variance (ANOVAs) with Bayes Factor Analysis were performed to assess the effect of wakeword (IV) on the usage of our politeness markers of interest (DV). Tests with Bayes Factors (BFs) greater than 0.5 (no more than 2:1 evidence against an effect) were subjected to further Bayesian post-hoc pairwise t-test analysis. We follow recommendations from previous researchers in our linguistic interpretations of reported BFs [35, 54].

4 RESULTS

4.1 Usage of ISA in Robot-Directed Messages

Our first set of analyses assessed whether the wakeword participants were required to use when addressing the robot led to differences in the proportion of robot-directed utterances phrased as ISA, similar to analyses performed in previous related work [57].

Analysis I: Full Dataset. Our analysis of the usage of Indirect Speech Act (ISA) in robot-directed messages provided extreme evidence in favor of an effect of wakeword ($BF\ 211.404$)². Post-hoc analysis provided extreme evidence specifically for differences in proportion of robot-directed utterances phrased as ISAs between the “Excuse me” ($M=0.363, SD=0.460$) and “Please” ($M=0, SD=0$) conditions ($BF\ 268.689$), and between the “Hey” ($M=0.408, SD=0.466$) and “Please” ($M=0, SD=0$) conditions ($BF\ 1098.660$). In contrast, post-hoc analysis provided moderate evidence *against* a difference between the “Excuse me” and “Hey” conditions ($BF\ 0.277$), which suggests there was no difference in the usage of ISA in robot-directed messages between those two conditions. Participants in both the “Hey” and “Excuse me” conditions phrased a significant percentage of their robot-directed utterances as ISAs, whereas *no one* used *any* robot-directed ISAs in the “Please” condition.

Analysis II: Reduced Dataset. Re-analysis on the reduced dataset in which keyword-only utterances were excluded reaffirmed and intensified these results (as shown in Fig. 4a), with extreme evidence in favor of an effect of wakeword on robot-directed ISA

use ($BF\ 9725.636$), with post-hoc tests revealing extreme evidence of differences between the “Excuse Me” ($M=0.673, SD=0.472$) and “Please” ($M=0, SD=0$) conditions ($BF\ 5560.962$) and between the “Hey” ($M=0.656, SD=0.437$) and “Please” ($M=0, SD=0$) conditions ($BF\ 16919.496$), and anecdotal evidence against a difference between the “Excuse Me” and “Hey” conditions ($BF\ 0.351$). This re-analysis reveals that for participants who used complete sentences as opposed to keyword-only messages, *the majority* of robot-directed communications in both the “Hey” and “Excuse me” conditions were phrased as ISAs, whereas, again, *no one* used *any* robot-directed ISAs in the “Please” condition.

4.2 Usage of ISA in Human-Directed Messages

Next, to see whether these wakeword-based restrictions on robot-directed utterances carried over into unrestricted *human-directed* utterances, we analyzed the effect of wakeword condition on proportion of ISAs used in human-directed utterances.

Analysis I: Full Dataset. Our analysis of the use of Indirect Speech Acts (ISA) in human-directed messages provided anecdotal evidence of an effect of wakeword ($BF\ 1.065$), suggesting that there was roughly equal evidence for or against an effect, with slightly more evidence for than against. Post-hoc analysis provided stronger yet still anecdotal evidence for differences between the “Excuse me” ($M=0.226, SD=0.400$) and “Please” ($M=0.034, SD=0.186$) conditions ($BF\ 2.531$), and between the “Hey” ($M=0.225, SD=0.396$) and “Please” ($M=0.034, SD=0.186$) conditions ($BF\ 2.549$), but, again, moderate evidence against a difference between the “Excuse me” and “Hey” conditions ($BF\ 0.261$). Participants in both the “Hey” and “Excuse me” conditions phrased a significant percentage of their human-directed utterances as ISAs, whereas very few participants used any human-directed ISAs in the “Please” condition.

Analysis II: Reduced Dataset. Re-analysis on the reduced dataset in which keyword-only utterances were excluded reaffirmed and intensified these results (as shown in Fig. 4b), with anecdotal evidence in favor of an effect of wakeword on human-directed ISA use ($BF\ 2.585$), with post-hoc tests revealing moderate evidence of differences between the “Excuse Me” ($M=0.442, SD=0.502$) and “Please” ($M=0.059, SD=0.243$) conditions ($BF\ 5.123$) and between the “Hey” ($M=0.391, SD=0.474$) and “Please” ($M=0.059, SD=0.243$) conditions ($BF\ 3.588$), and anecdotal evidence against a difference between the “Excuse Me” and “Hey” conditions ($BF\ 0.361$).

4.3 “Please” Usage in Human-Directed Messages

Next, to see whether these wakeword-based restrictions on robot-directed utterances resulted in increased use of “please” in unrestricted *human-directed* utterances in the “Please” condition, we analyzed the effect of wakeword condition on frequency of “please” use in human-directed utterances.

Analysis I: Full Dataset. Our analysis of the usage of the word “please” in human-directed messages provided extreme evidence in favor of an effect of wakeword ($BF\ 10936.264$). The post-hoc analysis provided extreme evidence for differences between the “Excuse me” ($M=0.105, SD=0.301$) and “Please” ($M=0.621, SD=0.471$) conditions ($BF\ 3785.222$), and between the “Hey” ($M=0.175, SD=0.366$) and “Please” ($M=0.621, SD=0.471$) conditions ($BF\ 157.681$), but anecdotal

²Bayes Factors above 100 indicate extreme evidence in favor of a hypothesis [35, 54]. Here, for example, our Bayes Factor of 211.404 suggests that our data were 211 times more likely to be generated under models in which wakeword condition is included than under those in which it is not.

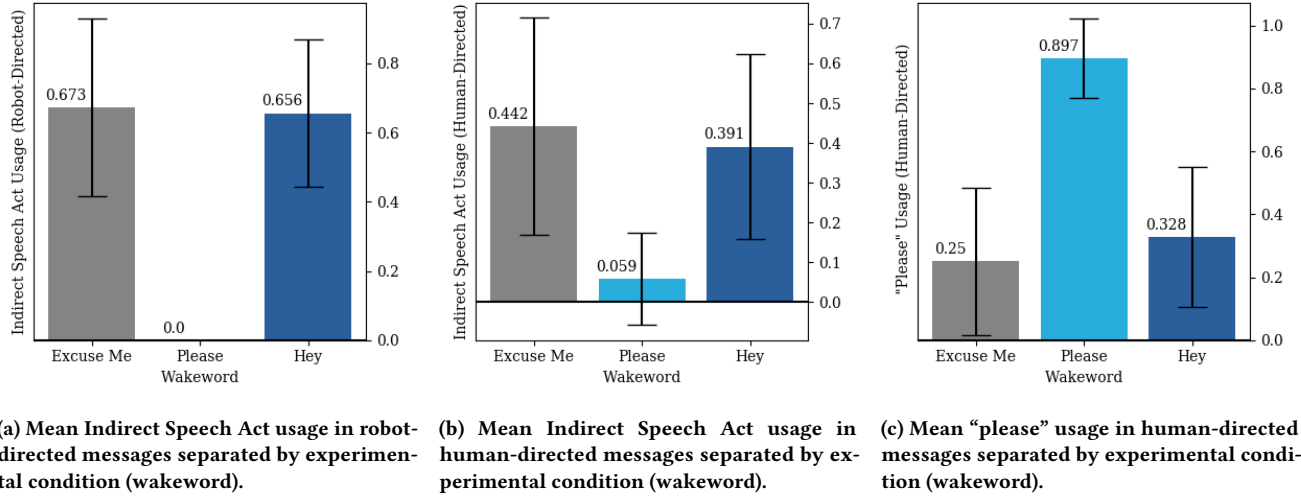


Figure 4: Results from our Reduced Dataset. Error bars represent 95% Credible Intervals.

evidence against differences between the “Excuse me” and “Hey” conditions (BF 0.345). Participants in both the “Hey” and “Excuse me” conditions used “please” in their human-directed utterances much more rarely than those in the “Please” condition.

Analysis II: Reduced Dataset. Re-analysis on the reduced dataset in which keyword-only utterances were excluded reaffirmed and intensified these results (as shown in Fig. 4c), with extreme evidence in favor of an effect of wakeword on human-directed “please” use (BF 645.050), with post-hoc tests revealing moderate evidence of differences between the “Excuse Me” ($M=0.250, SD=0.433$) and “Please” ($M=0.897, SD=0.266$) conditions (BF 711.150) and between the “Hey” ($M=0.328, SD=0.454$) and “Please” ($M=0.897, SD=0.266$) conditions (BF 194.849), and anecdotal evidence against a difference between the “Excuse Me” and “Hey” conditions (BF 0.381).

4.4 Usage of other politeness markers in Robot- and Human-Directed Messages

No use of any of the other politeness markers were identified in any Robot-Directed and Human-Directed messages.

5 DISCUSSION

In this work, we hypothesized that compared to the baseline wakeword “Hey <Name>”, participants would be more polite in robot-directed communication when required to use a polite wakeword “Excuse me <Name>” (H1), and that this politeness would carry over to human-human communication (H2). We also hypothesized that participants would be more polite in robot-directed communication when required to use wakeword “Excuse me <Name>” than when required to use wakeword “<Name> please” (H3), and that this politeness would similarly carry over to human-human communication (H4).

Our analysis demonstrated consistent support for hypotheses H3 and H4 but consistently refuted hypotheses H1 and H2. That is, while requiring participants to use the “Please” wakeword led to lexical priming effects in which participants more frequently

used “please” at the beginning of their utterances with humans despite not being instructed to do so, this wakeword also led to the complete erasure of the use other politeness strategies, such as ISAs, which are *more strongly correlated with perceived politeness than sentence-frontal please-usage* from robot-directed communication, and nearly complete erasure of such strategies from human-directed communication as well.

These results thus support a high level theory of the influence of wakewords on human politeness strategies, in which the “Please” wakeword (1) syntactically primes speakers to frame their robot-directed utterances as direct commands rather than ISAs; and (2) lexically primes speakers to continue to use sentence-frontal “please” in subsequent human-directed utterances, producing another round of syntactic priming effects that result in consistently less polite language human-directed utterances.

In this section, we will further discuss these findings and their implications. We will first discuss our findings on human politeness between wakeword “Excuse me” and “Hey” (hypotheses H1 and H2), and then discuss our findings between wakeword “Excuse me”/ “Hey” and “Please” (hypotheses H3 and H4).

5.1 “Excuse me” versus “Hey”

As described above, there were no differences in the usage of ISAs or other politeness markers between the “Excuse me” and “Hey” conditions. This finding was initially surprising because previous research has shown clear differences between these wakewords in robot-directed communication. Specifically, it has previously been shown that the wakeword “Excuse me” more strongly encourages people to phrase their utterances as ISAs than does “Hey” when speaking with robots [57]. An explanation for this discrepancy may be found in the differing deployment contexts of our experiment versus that of Williams et al. [57].

While in both our experiment and that presented by Williams et al. [57] participants relay instructions to robots in a restaurant

context, the two experiments were deployed in fundamentally different ways. The experiment presented by Williams et al. [57] was deployed as an in-person experiment where participants were able to directly communicate to a physically embodied, autonomous robot over the course of several minutes. In contrast, due to COVID-19 (cf. [21]), our experiment had to be deployed as a crowdsourced online experiment, in which participants could not directly observe the robot after an initial 24-second introduction video. Previous research has clearly demonstrated that people perceive robots differently based on their embodiment [4, 22, 36, 51]. Specifically, a robot with a physical embodiment elicits stronger feelings of social presence [51] and is perceived more positively as an interactive partner [4, 36] than is a robot that is only embodied virtually. Accordingly, participants may not apply the same sociocultural politeness norms when they “interact” with a virtual robot, or may not apply those norms in the same ways or to the same extents. Moreover, because the robot used by [57] was physically embodied, participants’ communication with it was voice-based, with participants communicating directly with the robot through spoken language, whereas in our experiment, participants needed to communicate with the robot through text. These different modalities of verbal communication certainly impact the way in which communicators form and phrase their verbal communications. For instance, verbally uttering “Excuse me” takes less time than does typing out “Excuse me”, and thus, participants in our experiment may have tended towards shorter (even if less appropriate) communications than did the participants in the experiment presented by [57], because it was easier, less tiring, and less irritating. This tendency could have washed out any potential differences between the “Excuse me” and “Hey conditions” of the sort previously observed by [57]. This explanation is partially justified by the high frequency at which participants elected to use brief keyword-only communications.

5.2 “Excuse me” / “Hey” versus “Please”

As described above, significant differences in both robot- and human-directed speech were found between participants in the “Please” condition and participants in the other two conditions, with participants in the “Please” condition being much more likely to use “please” with humans, but much less likely to use more effective politeness strategies such as ISAs with either humans or robots. This finding directly supports our expectations regarding the likely effects of wakeword “Please”. As discussed in the introduction, this suggests that the “please” wakeword in fact backfires and causes people to be *impolite*, as sentence-frontal “Please usage is in fact negatively correlated with perceived politeness [20]. In addition, this further suggests that the “please” wakeword in fact discourages the use of linguistic forms that are actually perceived as polite, such as ISA forms like conventionally indirect requests [20]. These findings suggest that the approaches taken by Amazon, for example, are likely to be effective at ameliorating the potential risks of current approaches to wakeword-driven interaction that have been raised in the popular press. However, our results also suggest that previous proposals to use ISA-encouraging wakewords like “Excuse me” may not be effective either (although as discussed above there

are important caveats to this point due to the online nature of this experiment).

6 CONCLUSION

In this paper, we used an online chat-based research platform to examine how wakewords used in HRI influence the politeness of users toward both robots and other humans. The results of this experiment have two major implications. First, the differences between non-polite wakewords (e.g., “Hey <Name>”) and politeness-forward wakewords (e.g., “Excuse me <Name>”) that were previously observed in laboratory studies with autonomous robots may not extend to interactions with ostensible robots in non-situated domains. Second, both of these wakeword forms are much more effective than please-centering wakewords (e.g., “Please <Name>”), which while intended to promote polite language in fact backfire and result in less polite language beyond the mere use of the word “please”. These findings have clear implications for robot and voice assistant designers (suggesting that such designers should absolutely *not* use please-centering wakewords if they are genuinely interested in encouraging polite behavior on the part of interactants) as well as for robotics researchers (by demonstrating some of the difficulties with COVID-friendly online experimentation). However, our work does have several key limitations that suggest a number of directions for future work.

Most notably, the COVID-19 Pandemic required us to conduct our experiment online, and to do this effectively we needed to develop and deploy a novel chat-based research platform. While this facilitated the safe performance of this experiment during a global pandemic, there are obvious differences between chat-based and spoken communication. For instance, the level of politeness might be different when changing from written language to spoken language. Once it becomes safe to do so, we will replicate this experiment through a testbed similar to that used by Williams et al. [57], to more clearly understand how our results may have been shaped by robot embodiment. By running this experiment in person, we may also be more likely to observe politeness modifiers such as gratitude, which were notably missing from our experimental data but are commonplace in spoken interaction.

On the other hand, we argue that evidence found in this text-based domain should serve as strong motivation for the likelihood of even more pronounced in vocal interactions. Given the online nature of this experiment, we would expect participants to be less willing to type out indirect language or add other politeness indicators when they do not have to. It is thus very encouraging that we are able to find effects in this text-based testbed, and would thus expect that in an actual situated domain where participants can have face-to-face interaction, results should be even stronger. To summarize, while the text-based platform was selected due to safety limitations imposed by COVID-19, the use of this platform does not undercut the value of our work, but rather showcases why an in-person follow-up study would be likely to yield positive results.

ACKNOWLEDGMENTS

This work was funded in part by Grants IIS-1849348 and IIS-1909847 from the National Science Foundation, and in part by a Google exploreCSR grant.

REFERENCES

- [1] Liliana Ardisson, Guido Boella, and Leonardo Lesmo. 1999. Politeness and speech acts. In *Proc. Workshop on Attitude, Personality and Emotions in User-Adapted Interaction*. Citeseer, 41–55.
- [2] Nicholas Asher and Alex Lascarides. 2001. Indirect speech acts. *Synthese* 128, 1 (2001), 183–228.
- [3] Edward Baig. 2018. Kids were being rude to Alexa, so Amazon updated it. <https://www.usatoday.com/story/tech/columnist/baig/2018/04/25/amazon-echo-dot-kids-alexa-thanks-them-saying-please/547911002/>.
- [4] Wilma A Bainbridge, Justin W Hart, Elizabeth S Kim, and Brian Scassellati. 2011. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics* 3, 1 (2011), 41–52.
- [5] Ilaria Baroni, Marco Nalin, Mattia Coti Zelati, Elettra Oleari, and Alberto Sanna. 2014. Designing motivational robot: how robots might motivate children to eat fruits and vegetables. In *Int'l Symp. Robot and Human Interactive Communication*.
- [6] Christoph Bartneck, Timo Bleeker, Jeroen Bun, Pepijn Fens, and Lynnyrd Riet. 2010. The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. *Paladyn, Journal of Behavioral Robotics* 1, 2 (2010), 109–115.
- [7] BBC. 2018. Amazon Alexa to reward kids who say: 'Please'. <https://www.bbc.com/news/technology-43897516>.
- [8] Michael Bonfert, Maximilian Spliethöfer, Roman Arzaroli, Marvin Lange, Martin Hanci, and Robert Porzel. 2018. If you ask nicely: a digital assistant rebuking impolite voice commands. In *proceedings of the 20th ACM international conference on multimodal interaction*. 95–102.
- [9] Jürgen Brandstetter and Christoph Bartneck. 2017. Robots will dominate the use of our language. *Adaptive Behavior* 25, 6 (2017), 275–288.
- [10] Jürgen Brandstetter, Clay Beckner, Eduardo Benitez Sandoval, and Christoph Bartneck. 2017. Persistent lexical entrainment in HRI. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. 63–72.
- [11] Holly P Branigan, Martin J Pickering, Jamie Pearson, and Janet F McLean. 2010. Linguistic alignment between people and computers. *Journal of pragmatics* 42, 9 (2010), 2355–2368.
- [12] Susan E Brennan. 1998. The grounding problem in conversations with and through computers. *Social and cognitive approaches to interpersonal communication* (1998), 201–225.
- [13] Susan E Brennan and Herbert H Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of experimental psychology: Learning, memory, and cognition* 22, 6 (1996), 1482.
- [14] 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, 3 (2014), 343–355.
- [15] Gordon Briggs, Tom Williams, and Matthias Scheutz. 2017. Enabling robots to understand indirect speech acts in task-based interactions. *Journal of Human-Robot Interaction* 6, 1 (2017), 64–94.
- [16] Penelope Brown, Stephen C Levinson, and Stephen C Levinson. 1987. *Politeness: Some universals in language usage*. Vol. 4. Cambridge university press.
- [17] Nick Charalambides. 2021. We recently went viral on TikTok - here's what we learned. <https://blog.prolific.co/we-recently-went-viral-on-tiktok-heres-what-we-learned/>.
- [18] Vijay Chidambaram, Yueh-Hsuan Chiang, and Bilge Mutlu. 2012. Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In *International conference on Human-Robot Interaction (HRI)*. ACM.
- [19] Derek Cormier, Gem Newman, Masayuki Nakane, James E Young, and Stephane Durocher. 2013. Would you do as a robot commands? An obedience study for human-robot interaction. In *International Conference on Human-Agent Interaction*.
- [20] Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. *arXiv preprint arXiv:1306.6078* (2013).
- [21] David Feil-Seifer, Kerstin S Haring, Silvia Rossi, Alan R Wagner, and Tom Williams. 2020. Where to next? The impact of COVID-19 on human-robot interaction research.
- [22] Kerstin Fischer, Katrin Lohan, and Kilian Foth. 2012. Levels of embodiment: Linguistic analyses of factors influencing HRI. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 463–470.
- [23] Daniel Fried, Jacob Andreas, and Dan Klein. 2018. Unified Pragmatic Models for Generating and Following Instructions. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. 1951–1963.
- [24] Raymond W Gibbs Jr. 1986. What makes some indirect speech acts conventional? *Journal of memory and language* 25, 2 (1986), 181–196.
- [25] Francesca Gino. 2015. Understanding ordinary unethical behavior: Why people who value morality act immorally. *Current opinion in behavioral sciences* 3 (2015), 107–111.
- [26] Jennifer Graham. 2019. Are Alexa and Siri making your kids rude? <https://www.deseret.com/2019/6/25/20676377/are-alexa-and-siri-making-your-kids-rude>.
- [27] Jaap Ham, René Bokhorst, Raymond Cuijpers, David van der Pol, and John-John Cabibihan. 2011. Making robots persuasive: the influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power. In *International conference on social robotics*. Springer, 71–83.
- [28] 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 *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 3727–3734.
- [29] 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.
- [30] Ryan B Jackson and Tom Williams. 2019. On Perceived Social and Moral Agency in Natural Language Capable Robots. In *2019 HRI Workshop on The Dark Side of Human-Robot Interaction: Ethical Considerations and Community Guidelines for the Field of HRI*.
- [31] Ryan Blake Jackson and Tom Williams. 2019. Language-Capable Robots may Inadvertently Weaken Human Moral Norms. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 401–410.
- [32] Ryan Blake Jackson and Tom Williams. 2021. A Theory of Social Agency for Human-Robot Interaction. *Frontiers in Robotics and AI* (2021).
- [33] JASP Team et al. 2021. JASP. Version 0.9.1 software (2021).
- [34] James Kennedy, Paul Baxter, and Tony Belpaeme. 2014. Children comply with a robot's indirect requests. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. ACM, 198–199.
- [35] Michael D Lee and Eric-Jan Wagenmakers. 2014. *Bayesian cognitive modeling: A practical course*. Cambridge university press.
- [36] Jamy Li. 2015. The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies* 77 (2015), 23–37.
- [37] Jane Lockshin and Tom Williams. 2020. "We need to start thinking ahead": the impact of social context on linguistic norm adherence. In *Annual Meeting of the Cognitive Science Society*.
- [38] Raul Benites Paradedá, Maria José Ferreira, João Dias, and Ana Paiva. 2017. How Robots Persuasion based on Personality Traits May Affect Human Decisions. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 251–252.
- [39] Martin J Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences* 27, 2 (2004), 169–190.
- [40] Daniel J Rea, Denise Geiskovitch, and James E Young. 2017. Wizard of awwww: Exploring psychological impact on the researchers in social HRI experiments. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. 21–29.
- [41] Paul Robinette, Wenchen Li, Robert Allen, Ayanna M Howard, and Alan R Wagner. 2016. Overtrust of robots in emergency evacuation scenarios. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. 101–108.
- [42] Eduardo Benitez Sandoval, Jürgen Brandstetter, and Christoph Bartneck. 2016. Can a robot bribe a human?: The measurement of the negative side of reciprocity in human robot interaction. In *Int'l Conf. on Human Robot Interaction (HRI)*.
- [43] Vasanthy Sarathy, Alexander Tsuetaki, Antonio Roque, and Matthias Scheutz. 2020. Reasoning Requirements for Indirect Speech Act Interpretation. In *Proceedings of the 28th International Conference on Computational Linguistics*. 4937–4948.
- [44] Shane Saunderson and Goldie Nejat. 2021. Robots asking for favors: The effects of directness and familiarity on persuasive HRI. *IEEE Robotics and Automation Letters* 6, 2 (2021), 1793–1800.
- [45] John R Searle. 1965. What is a speech act. *Perspectives in the philosophy of language: a concise anthology* 2000 (1965), 253–268.
- [46] John R Searle. 1975. Indirect Speech Acts. *Syntax and Semantics* 3 (1975), 59–82.
- [47] John R Searle, Ferenc Kiefer, Manfred Bierwisch, et al. 1980. *Speech act theory and pragmatics*. Vol. 10. Springer.
- [48] Sukyoung Seok, Eunji Hwang, Jongsuk Choi, and Yoonseob Lim. 2022. Cultural Differences on Indirect Speech Act Use and Politeness in Human-Robot interaction. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*.
- [49] Megan Strait, Cody Canning, and Matthias Scheutz. 2014. Let me tell you! investigating the effects of robot communication strategies in advice-giving situations based on robot appearance, interaction modality and distance. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (HRI)*.
- [50] Sarah Strohkorb Sebo, Margaret Traeger, Malte Jung, and Brian Scassellati. 2018. The ripple effects of vulnerability: The effects of a robot's vulnerable behavior on trust in human-robot teams. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 178–186.
- [51] Kazuaki Tanaka, Hideyuki Nakanishi, and Hiroshi Ishiguro. 2014. Comparing video, avatar, and robot mediated communication: pros and cons of embodiment. In *International conference on collaboration technologies*. Springer, 96–110.

- [52] Sean Trott and Benjamin Bergen. 2017. A theoretical model of indirect request comprehension. In *2017 AAAI Fall Symposium Series*.
- [53] Peter-Paul Verbeek. 2011. *Moralizing technology: Understanding and designing the morality of things*. University of Chicago Press.
- [54] Eric-Jan Wagenmakers. 2007. A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review* 14, 5 (2007), 779–804.
- [55] Ruchen Wen, Mohammed Aun Siddiqui, and Tom Williams. 2020. Dempster-Shafer theoretic learning of indirect speech act comprehension norms. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 10410–10417.
- [56] Tom Williams, Gordon Briggs, Bradley Oosterveld, and Matthias Scheutz. 2015. Going beyond literal command-based instructions: Extending robotic natural language interaction capabilities. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- [57] Tom Williams, Daniel Grollman, Mingyuan Han, Ryan Blake Jackson, Jane Lockshin, Ruchen Wen, Zachary Nahman, and Qin Zhu. 2020. “Excuse Me, Robot”: Impact of Polite Robot Wakewords on Human-Robot Politeness. In *International Conference on Social Robotics*. Springer, 404–415.
- [58] 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 the 13th ACM/IEEE International Conference on Human-Robot Interaction*.
- [59] Katie Winkle, Séverin Lemaignan, Praminda Caleb-Solly, Ute Leonards, Ailie Turton, and Paul Bremner. 2019. Effective persuasion strategies for socially assistive robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 277–285.
- [60] Qin Zhu, Tom Williams, Blake Jackson, and Ruchen Wen. 2020. Blame-laden moral rebukes and the morally competent robot: A Confucian ethical perspective. *Science and Engineering Ethics* (2020), 1–16.