Fresh Start: Encouraging Politeness in Wakeword-Driven Human-Robot Interaction

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

Alyssa Hanson
Colorado School of Mines
Golden, CO, USA
abhanson@mines.edu

Zhao Han
Colorado School of Mines
Golden, CO, USA
zhaohan@mines.edu

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

ABSTRACT

Deployed social robots are increasingly relying on *wakeword-based interaction*, where interactions are human-initiated by a wakeword like "Hey Jibo". While wakewords help to increase speech recognition accuracy and ensure privacy, there is concern that wakeword-driven interaction could encourage impolite behavior because wakeword-driven speech is typically phrased as commands. To address these concerns, companies have sought to use wakeword design to encourage interactant politeness, through wakewords like "⟨Name⟩, please". But while this solution is intended to encourage people to use more "polite words", researchers have found that these wakeword designs actually decrease interactant politeness in text-based communication, and that other wakeword designs could better encourage politeness by priming users to use Indirect Speech Acts. Yet there has been no previous research to directly compare these wakewords designs in in-person, voice-based human-robot interaction experiments, and previous in-person HRI studies could not effectively study carryover of wakeword-driven politeness and impoliteness into human-human interactions. In this work, we conceptually reproduced these previous studies (n=69) to assess how the wakewords "Hey ⟨Name⟩", "Excuse me ⟨Name⟩", and "⟨Name⟩, please" impact robot-directed and human-directed politeness. Our results demonstrate the ways that different types of *linguistic priming* interact in nuanced ways to induce different types of robot-directed and human-directed politeness.

CCS CONCEPTS

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

KEYWORDS

human-robot communication, indirect speech act, politeness

HRI '23, March 13–16, 2023, Stockholm, Sweden

ACM ISBN XXX…$15.00
https://doi.org/XXX

ACM Reference Format:
Ruchen Wen, Alyssa Hanson, Zhao Han, and Tom Williams. 2023. Fresh Start: Encouraging Politeness in Wakeword-Driven Human-Robot Interaction. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23), March 13–16, 2023, Stockholm, Sweden.* ACM, New York, NY, USA, 10 pages. https://doi.org/XXX

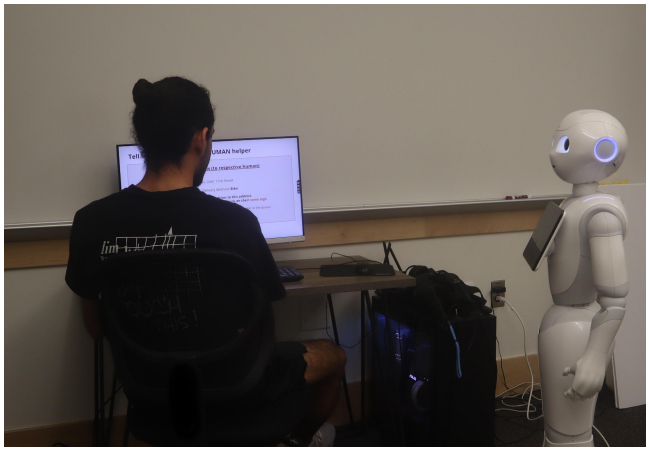


Figure 1: Participants interacted with a Softbank Pepper using to one of three condition-specified Wakewords: "Excuse me Pepper", "Pepper Please", and "Hey Pepper".

1 INTRODUCTION

Social robots are increasingly being deployed into domains where they must interact with everyday users who do not have prior experience with robots. To ensure natural and fluid interaction with these users, robots need to be able to interact in a way that is natural, humanlike, and conforming to interactants expectations and the social conventions of a given context. There has thus been a sustained research effort toward developing robots with natural language capabilities [46, 47, 57].

While most Human-Robot Interaction (HRI) research on Human-Robot Dialogue focuses on enabling natural, fluid, mixed-initiative communication, many voice-interactive robots that have been successful deployed in various real-world contexts instead rely on *wakeword-based* interaction, in which interactions are human-initiated

and begin with a keyphrase (i.e., wakeword), such as "Hey Jibo". This wakeword design is commonly used for other voice interactive technologies, such as virtual intelligent assistants (e.g., Amazon Alexa, Google Assistant and Siri), for a variety of important reasons. Wakeword-driven interaction has been found to be an effective means of interaction not only because it helps to constrain speech recognition and ensure accurate and responsive execution of user commands [35], but also because it aims to promote privacy and security [58], to minimize the risk of these technologies being used as surveillance tools. These concerns are not unique to virtual intelligent assistants; and moreover, they touch on key sociotechnical concerns that are not easily and immediately addressable by technical improvements to speech recognition models. As such, there is good reason to expect that wakeword-driven interaction will be an important interaction design paradigm for both virtual intelligent assistants and social robotics, at least for the near future.

Nevertheless, a number of key concerns have been raised about wakeword-driven interaction. One such concern is the way that many wakeword-based interactions seem to be intrinsically command-based, inducing people to speak in ways that often being perceived as impolite (e.g., "Hey Jibo, take a picture."). Because of this orientation towards command-based interactions, a number of journalists and parents have questioned whether wakeword-driven intelligent agents might entrain users (especially children) into predominantly command-based interaction patterns, in essence teaching these users to be domineering and rude [1, 20, 21, 68]. Moreover, these journalists and parents have questioned whether these effects might carry over into human-human interactions as well [19]. These concerns have been echoed by HRI researchers [69, 70], who have pointed to previous evidence for so-called "ripple effects" in which human influence on robot moral and social norms may carry over into human-human interactions in just this way [43, 65].

To address these concerns when raised for virtual intelligent assistants, companies have been exploring solutions to intentionally encourage politeness from people [1, 66]. For example, Amazon deployed an optional "Alexa please" mode which encourages children to use the "magic words" (e.g., "please" and "thank you") when talking to Alexa [4]. While this solution is intended to encourage people to use more "polite words", some researchers have argued that these wakeword designs may not actually induce more human politeness, and in some cases could in fact decrease interactant politeness. From a politeness theoretic perspective, using "please" at the beginning of a sentence is in fact negatively correlated with perceived politeness [16]. Sentence-frontal please usage is a strategy used to phrase commands more politely; conversely, this means that requiring a request to start with "please" has the effect of inducing the rest of the sentence to be phrased as a direct command, thus resulting in a high level face threat. For the word "please" to positively correlate with politeness, it needs to be used alongside other politeness communication strategies [16].

One common strategy that speakers in many cultures regularly use to reduce face threat (i.e., to adhere with sociocultural politeness norms) is the use of *Indirect Speech Acts* (ISAs), such as "*Could you please bring me some water*", in which a speaker's literal meaning does not match their intended meaning. ISAs are one of the most commonly used politeness strategies, especially in cases where speakers need to issue commands and requests. ISAs have been shown to be common in human-robot interaction, and in some contexts, speakers are noticeably reticent to use anything *but* ISAs [72][1]. As such, encouraging robots' interactants to use ISAs could be a more effective strategy than Please-based wakeword design if our true goal is to encourage interactant politeness.

In fact, previous research has found that, at least in text-based communication, while ostensibly polite wakewords such as "⟨Name⟩, please" do increase the usage the word "please", they actually discourage politeness as measured by ISA use, due to the particular types of lexical and syntactic priming induced by Please-based wakewords [69]. Research has also shown that in live human-robot interactions, wakewords that conform to social conventions, such as "Excuse me ⟨Name⟩", could prime users to phrase their commands as ISAs [70]. However, to the best of our knowledge, there has been no previous research to directly compare those two types of wakewords in in-person, voice-based human-robot interaction experiments. Moreover, previous in-person studies of wakeword-based Human-Robot Interaction were not able to effectively study the potential carryover of wakeword-driven politeness and impoliteness into human-human interactions.

In this work, we thus present a human subject study (n=69), as a conceptual reproduction of previous research from Wen et al. [69] and Williams et al. [70], to assess how the wakewords "Hey ⟨Name⟩", "Excuse me ⟨Name⟩", and "⟨Name⟩, please" impact both robot-directed and human-directed politeness in human-robot interaction. Our experiment builds off on the wakeword-based human-robot interaction paradigm used by Wen et al. [69] in their online experiments, while converting this paradigm for use in in-person HRI studies. Our results demonstrate the ways that different types of *linguistic priming* interact in nuanced ways to induce different types of robot-directed and human-directed politeness.

2 RELATED WORK

2.1 Persuasive Robotics

In the HRI literature, a vast amount of research has demonstrated the persuasive power of embodied robots [44]. While robots' persuasive capabilities can be influenced by general features of their designs [53, 56], they are especially impacted by robots' communication strategies, both nonverbal [22] and verbal [3, 13, 15]. When robots use verbal cues, researchers have shown that different types of verbal strategies, such as indirectness, can be used to increase robots' persuasive capabilities [34, 56].

Due to physical robots being perceived as both social and *moral* agents [26, 27], researchers in HRI have also started studying the moral aspect of robot's persuasive power, and the ways that robots can influence the systems of moral norms that govern human behaviors, intentionally or unintentionally, and for better or for worse. Jackson and Williams [25] showed that robots can unintentionally weaken humans' perceptions of moral norms through common dialogue patterns; an effect they computationally remedy through later architectural work [28]. Briggs and Scheutz [10] found that through displays of verbal protest and distress, robots can intentionally dissuade people from undesired or potentially unethical

[1]These findings mainly hold in western, English-speaking contexts, and do not hold for human-robot interactions conducted in Korean [59].

actions. Sandoval et al. [55] specifically demonstrated robots' ability to encourage interactants to accept bribes.

The potential for this moral influence to be exerted unintentionally creates a significant challenge for robot designers. Even if robot designers do not have an explicit goal to persuade interactants toward specific actions or to maintain a specific norm, they must nevertheless be attentive to and try to head off these potential negative effects. Moreover, this challenge becomes especially salient due to evidence of "ripple effects" in human-robot interaction [65], where the ways that robots shape human behaviors during interactions carry over into those humans' future interactions with other humans [43, 65, 71]. This is a particularly acute concern when robots wield morally persuasive power. If the effects of inadvertent negative moral influence lingers and carries over into future human-human interactants, this risks negative and potentially long-lasting influence on the broader moral ecosystems into which humans are embedded [63].

On the other hand, however, this challenge of avoiding negative, unintentional moral influence also comes with an opportunity for promoting positive moral influence, and opportunities to help cultivate humans' moral ecosystems [12, 14, 73, 74]. One way that robots might seek to cultivate their social and moral ecosystem is by encouraging interactants to adhere to human politeness norms.

2.2 Politeness

Politeness norms are fundamental to governing human interactions. According to Brown and Levinson's Politeness Theory [11], people negotiate the level of threat to one another's *Face* (i.e., the public image that the other person wants to maintain and enhance) through regular communication. Face is composed of two aspects: *Positive Face* (i.e., one's desire for a self-image) and *Negative Face* (i.e., one's desire to have freedom of action) [11]. From the politeness theoretic perspective, politeness is negatively correlated with face threat.

People use a variety of strategies to express politeness and decrease the level of face threat in everyday communication, such as by expressing gratitude and providing compliments [16]. As introduced in the previous section, one of most effective linguistic strategies to reduce face threat is the use of ISAs. In contrast, the word "please", which is the sterotypical denotation of politeness, is actually not always perceived as polite. In order for the word "please" to be considered as polite, it in fact need to be used along with other linguistic constructions (e.g., ISAs) [16].

Research has shown that people apply politeness strategies, including the use of ISAs, when communicating with robots [69, 70, 72]. Yet humans are not automatically and uniformly polite towards robots, and their use of these politeness cues is mediated by a number of different factors. Some of these factors are things that robots and their designers have no control over. For example, Williams et al. [72] showed that interactants use ISAs more frequently with robots in contexts with strong sociocultural norms and conventions (e.g., restaurants) than in more novel and task-oriented contexts.

In contrast, some of these politeness-mediating factors can be steered through robot design in explicit or implicit ways. Robots and other language-capable technologies can explicitly encourage politeness by requesting and requiring human interactants to use polite language when communicating with them [7], although the success of such *moral interventions* is highly dependent on the way they are structured [33] (see also [24, 29, 36–39, 49] in the context of moral interventions beyond politeness).

Recently, HRI researchers have been exploring how to use more implicit dimensions of robots' interaction designs to encourage human politeness. Specifically, Wen et al. [69] and Williams et al. [70] have demonstrated that the choice of different robots *wakewords*, which humans are required to use from a technical perspective (rather than explicitly encouraged or required by the robot according to ostensible robot intent) may also be able to encourage human politeness towards robots, and possibly also towards other humans.

2.3 Robot Wakeword Design

As described above, in the context of smart speakers, companies like Amazon and Google have sought to encourage politeness through wakewords involving the word "please". But inspired by Danescu-Niculescu-Mizil et al. [16]'s finding that sentence-frontal please use is anticorrelated with politeness, Williams et al. [70] hypothesized that this approach could in fact backfire, and sought to explore the use of other possible wakewords for encouraging politeness.

Williams et al. [70] compared politeness indicators for the "Hey, ⟨Name⟩" and "Excuse me, ⟨Name⟩" wakewords in a restaurant setting, requiring the participant to speak to both the robot and human during the experiment. This research found that participants with the impolite wakeword condition "Hey, ⟨Name⟩" were less likely to use Indirect Speech Acts than the participants with the polite wakeword condition "Excuse me, ⟨Name⟩", suggesting that "Excuse me, ⟨Name⟩" could serve as an effective wakeword for encouraging politeness. Moreover, Williams et al. [70] explain their results through an interesting account grounded in linguistic priming, which we will return to later. Yet Williams et al. [70]'s work does not actually provide experimental evidence against the effectiveness of "please".

In contrast, in more recent work, Wen et al. [69] explicitly compare "Hey, ⟨Name⟩" and "Excuse me, ⟨Name⟩" with the "⟨Name⟩, Please" strategy that had been critiqued by Williams et al. [70]. Wen et al. [69]'s research provided extremely strong evidence that different wakewords led to different uses of ISAs. But while Williams et al. [70] had found that "Excuse me ⟨Name⟩" led to more robot-directed politeness than "Hey ⟨Name⟩", Wen et al. [69] found no such effect, finding instead that both wakewords were *equally* as effective at promoting ISA use, and that "⟨Name⟩, Please" fared much worse than either. Unlike previous research, Wen et al. [69] were also able to look at carryover effects into human-human interactions, and found that "Excuse me ⟨Name⟩" and "Hey ⟨Name⟩" *may* have better enabled carry-over of politeness into human-human interaction, in the sense that human interactants were more likely to use ISAs with other humans when these wakewords were used to govern robot-directed speech, while also finding that "⟨Name⟩, please" *was* more effective at (the less nuanced goal of) encouraging people to use "please" more often with one another.

One limitation of Wen et al. [69]'s work, however, was that it was conducted online in a text based chat application, with limited ecological validity for real human-robot interaction experiments. This clearly motivates in-person human-robot interaction experiments to confirm or refute Wen et al. [69]'s findings and arbitrate between the differences found in Williams et al. [70] and Wen et al.

[69]'s work. But moreover, there is a need to interrogate the underlying theoretical assumptions made by Wen et al. [69]. Wen et al. [69] argue that their results support a high-level theory in which different wakewords lead to different downstream behaviors due to different types of *linguistic priming* effects. However, they only make attempt to reason about the priming effects at play for the wakeword "Please ⟨Name⟩". We argue that a priming based explanation of wakeword effects needs to be able to explain the downstream effects of *all* wakewords under consideration. As such, before formulating our experimental hypotheses, we will consider the different types of linguistic priming that might be at play, so that we can ground our hypotheses in those specific mechanisms.

In the following section, we will introduce three types of *linguistic priming*, which motivate our hypotheses and provide a critical lens through which to interpret and explain our results.

2.4 Linguistic Priming

During human-human dialogue, speakers subtly influence each other's linguistic choices at multiple levels of linguistic abstraction, including phonetics, lexical choice, syntax, and semantics [52], in which the psycholinguistics literature effects can directly influence politeness [2]. This influence is exerted by activating mental representations associated with those different levels of linguistic abstraction, through different sorts of *priming*.

At least three main types of priming could determine the effects of different wakewords: (1) syntactic priming [8, 9, 51], where the wakeword most easily facilitates a post-wakeword clause with a particular syntactic construction; (2) semantic priming [18, 48], where the wakeword most easily facilitates a post-wakeword clause with a particular meaning; and (3) lexical priming [6, 23], where the wakeword most easily facilitates a post-wakeword clause in which particular words are used. We expect these different types of linguistic priming to combine in different ways to lead to different effects on robot- and human-directed ISA and "Please" use.

First, we expect both syntactic and semantic priming to play key roles in determining ISA use in post-wakeword clauses of participants' robot-directed utterances, with syntactic priming being more important. We do not expect lexical priming to play a significant role here. Second, we expect robot-directed ISA use to directly carry over into human-directed utterances, and thus expect this influence to ultimately be grounded in these same types of priming. Finally, we expect both semantic and lexical priming to play key roles in determining "please" use in participants' human-directed utterances, with lexical priming being more important. We do not expect syntactic priming to play a significant role here.

If these linguistic expectations are accurate, they should inform how different wakewords influence both robot and human-directed speech. We formulate these expectations through a set of concrete hypotheses that we delineate in the next section.

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". Based on our expectations for how each type of linguistic priming will affect human-directed and robot-directed ISA and please use, we propose the following concrete hypotheses:

H1: We hypothesize that different required wakewords will lead to differences in robot-directed politeness as assessed by ISA use. Specifically, robot-directed ISA use will be:
(H2a) higher under "Excuse me, ⟨Name⟩" than under "⟨Name⟩ Please" (confirming [69]);
(H2b) higher under "Excuse me, ⟨Name⟩" than under "Hey ⟨Name⟩" (confirming Williams et al. [70] and refuting [69]); and
(H2c) higher under "Hey ⟨Name⟩" than under "⟨Name⟩ Please" (confirming [69]).

H2: We hypothesize that these effects will carry over into differences in human-directed politeness as assessed by ISA use. Specifically, human-directed ISA use will be:
(H2a) higher under "Excuse me, ⟨Name⟩" than under "⟨Name⟩ Please" (confirming [69]);
(H2b) higher under "Excuse me, ⟨Name⟩" than under "Hey ⟨Name⟩" (refuting both Williams et al. [70] and [69]); and
(H2c) higher under "Hey ⟨Name⟩" than under "⟨Name⟩ Please" (confirming [69]).

H3: We hypothesize that different required wakewords will lead to differences in human-directed politeness as assessed by use of the word "please". Specifically, human-directed "Please" use will be:
(H3a) higher under "⟨Name⟩ Please" than under "Hey ⟨Name⟩" (confirming [69]);
(H3b) higher under "⟨Name⟩ Please" than under "Excuse me, ⟨Name⟩" (confirming [69]); and
(H3c) higher under "Excuse me, ⟨Name⟩" than under "Hey ⟨Name⟩" (confirming [69]).

3 METHOD

3.1 Experimental Design

To test these hypotheses, we conducted an IRB-approved human-subject study with a between-subjects design, with each participant assigned to one of three conditions (*Excuse Me*, *Please*, *Hey*).

3.2 Task Design

A simulated food delivery task based off that used by Wen et al. [69] was used in this experiment, in which participants provided a series of requests to robot and human teammates. This experiment primarily deviated from the task method used by Wen et al. [69] in two ways: (1) it was conducted in person rather than online; and (2) participants interacted with their human and robot teammates through spoken language rather than through text.

The participant was situated in front of a desktop computer in a room with three teammates: a robot (the Softbank Pepper) and two human confederates. Each Human Teammate was associated with a different delivery method (car or bike). Each Human Teammate wore a nametag bearing their name and their associated delivery method. Each participant was provided with a series of ten food orders, which were presented on a series of slides in a presentation that was advanced by hitting the [Space] key. Each food order was broken into two slides.

On the first slide, participants were given the name of a dish that needed to be prepared (e.g., "Cheeseburger"). The slide then stated "Pepper now needs to tell the kitchen to prepare this order. Get Pepper to do this." This phrasing was chosen to avoid an implication that any particular type of Speech Act should be used. When

speaking to Pepper, participants were required to use a wakeword to trigger the robot's speech recognition. After participants delivered this instruction (repeating the instruction if necessary due to speech recognition failure), they advanced to the second slide.

On the second slide, participants were given a delivery address that the previously indicated dish should be delivered to, and the optimal delivery method to use for that address. The slide then stated "Dispatch the appropriate driver to this address." Again, this phrasing was chosen to avoid an implication that any particular type of Speech Act should be used. After participants delivered this instruction to the appropriate Human Teammate, the teammate left the room for a few moments. Once they returned, the participant advanced to the next slide, i.e., the first slide for the next order.

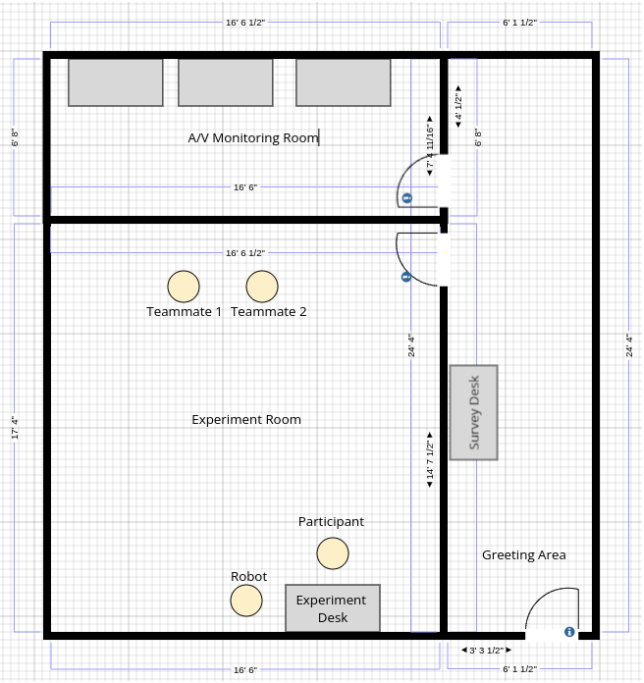


Figure 2: Experiment Environment Layout.

3.3 Measures

The primary data collected in this experiment was participants' utterances directed towards their robot and human teammates. This data was annotated according to two key criteria:

- **Directness** — Utterances were annotated as *Indirect* if a conventionally indirect form such as "Could you ⟨X⟩" or "I need ⟨X⟩" was used; *Direct* if the participant's intent was stated directly, and *Keyword-Only* if the participant used a keyword rather than a complete sentence.
- **Please Use** — Human-directed utterances were annotated based on the appearance of "Please".

Participants also completed a survey requesting age, college major, gender (forced choice ∈ {Female, Male, Nonbinary, Genderfluid, Prefer Not to Say} and familiarity with robots (1="Not familiar at all" to 5="Very familiar").

3.4 Procedure

After providing informed consent, participants were introduced to the experimental task. Participants were informed that they would be testing a new algorithm for dispatching food deliveries, and that their job in the task was to use a desktop application to pass food orders and delivery instructions to a robot and human teammate. Participants were then walked through an example task instance, and were told to make sure to preface all instructions given to the robot teammate with the wakeword needed to trigger the robot's speech recognition. The wakeword each participant was instructed to use was determined by their experimental condition: "Excuse me Pepper", "Pepper Please", or "Hey Pepper". Participants then then began their series of ten food orders. The robot used in this experiment was fully autonomous, and was programmed using the Softbank Choregraphe software to listen for utterances beginning with the condition-specified wakeword and reply with "Got It. I sent your request to the kitchen. The order is now ready to deliver." After all ten food orders had been completed, the participant completed a demographic survey, was debriefed about the purpose of the experiment and paid for their time.

3.5 Participants

Seventy-eight participants were recruited from the campus of a small engineering university. Nine of these participants were discarded due to robot mechanical failure, experimenter miscommunication of instructions, or dramatic participant deviation from instructions. Complete data was thus collected from 69 participants (25 in the "Hey Pepper" condition, 21 in the "Excuse me Pepper" condition, and 23 in the "Pepper Please" condition). Twenty-three self-identified as Female, forty-one as Male, three as Nonbinary, one as Genderfluid, and one did not self-identify. Participant ages ranged from 18 to 65 (M=22.464, SD=8.165). Average participant self-report for familiarity with robots was 2.71 out of 5 (SD=0.925). All participants reported being STEM majors, with the exception of six participants, three who were undecided, and three (non-students) who did not report.

3.6 Analysis

To analyze our data, experiment recordings were transcribed, and divided into robot- and human-directed utterances. These transcriptions were then annotated by three annotators according to each measure described in Sec. 3.3. In cases where the two annotators disagreed, a third annotator vote was used to resolve the conflict. When participants repeated themselves (due to Speech Recognition Errors that resulted in no response from the robot), only the first of the participant's utterances was retained. When a participant spoke with only a keyword, such as "Hey Pepper, <food item>," that utterance was excluded due to the inability to perform utterance-level analysis. One participant who spoke with all keyword-only phrases was not included in the data for this same reason. The mean value for each measure was then calculated for each participant's robot-directed utterances and each participant's human-directed utterances. These means were analyzed using Bayesian Analyses of Variance with Bayes Factor Analysis [54] using the JASP 0.16.4 statistical software [32].

While the Bayesian statistical analysis has been gaining attention within the scientific community, it is still relatively uncommon in HRI research. We chose Bayesian analysis over Frequentist Null Hypothesis Significance Testing (NHST) due to its robustness to small sample-sizes [61], support for incremental and flexible sampling, support for collecting evidence both for *and against* hypotheses [5, 64, 67], and ease of interpretability [31].

Each Bayesian analysis produced a Bayes Factor BF_{10} representing the ratio of evidence in favor versus against an effect of Wakeword. That is, the BF_{10} represents how much more likely the data were to be generated under a model accounting for Wakeword (H_1) than under one that did not (H_0). These Bayes Factors were then interpreted using the classification scheme proposed by Lee and Wagenmakers [41]. Under this scheme, Bayes Factors $BF_{10} \geq 3$ are interpreted as providing at least moderate evidence in favor of H_1 relative to H_0, Bayes Factors $\frac{1}{3} < BF_{10} < 3$ are interpreted as providing anecdotal (and thus inconclusive) evidence, and Bayes Factors $BF_{10} \leq \frac{1}{3}$ are interpreted as providing at least moderate evidence *against* H_1 (and thus *in favor* of H_0), allowing H_1 to be ruled out. When our Bayes Factor analysis of our ANOVA results could not rule out an effect of Wakeword, post-hoc Bayesian t-tests were used to effect pairwise comparisons between each of the three wakewords.

4 RESULTS

4.1 ISA Use in Robot-Directed Utterances

A Bayesian ANOVA provided very strong evidence for an effect of wakeword on robot-directed ISA use ($BF_{10} = 50.251$). We thus conducted post-hoc analyses. Post-hoc analysis provided very strong evidence ($BF_{10} = 65.127$) that participants in the *Excuse Me* condition (M=0.738, SD=0.412) used more ISAs in their robot-directed utterances than did participants in the *Please* condition (M=0.272, SD=0.394), supporting **H1a**. Post-hoc analysis also provided strong evidence ($BF_{10} = 20.150$) that participants in the *Hey* condition (M=0.684,SD=0.456) also used more ISAs in their robot-directed utterances than did participants in the *Please* condition, supporting **H1c**. However, post-hoc analysis also provided moderate evidence against such a difference between the *Hey* and *Excuse Me* condition ($BF_{10} = 0.315$), refuting **H1b**.

4.2 ISA Use in Human-Directed Utterances

A Bayesian ANOVA provided moderate evidence against an effect of wakeword on human-directed ISA use ($BF_{10} = 0.211$), refuting **H2a**, **H2b**, and **H2c**. We thus did not proceed with post-hoc analyses.

4.3 Please Use in Human-Directed Utterances

A Bayesian ANOVA provided anecdotal evidence for an effect of wakeword on human-directed please use($BF_{10} = 2.392$). We thus conducted post-hoc analyses. Post-hoc analysis provided moderate evidence ($BF_{10} = 7.016$) that participants in the *Please* condition (M=0.457,SD=0.445) were more likely to include "Please" in their human-directed utterances than were participants in the *Hey* condition (M=0.144,SD=0.303), supporting **H3a**. However, post-hoc analysis also provided anecdotal evidence ($BF_{10} = 0.606$) against such a difference between the *Please* condition and the *Excuse Me* condition (M=0.286, SD=0.400), and anecdotal evidence ($BF_{10} = 0.620$) against such a difference between the *Excuse Me* condition and the *Hey* condition, potentially refuting **H3b** and **H3c**.

5 DISCUSSION

5.1 ISA Use in Robot-Directed Utterances

Our first hypothesis was that different required wakewords would lead to differences in robot-directed politeness, as assessed by ISA use. Specifically, we hypothesized that *Excuse me* would promote the greatest ISA use, followed by *Hey*, followed by *Please*. This hypothesis was partially supported: *Excuse Me* and *Hey* both led to more robot-directed politeness than *Please*, but no difference was found between *Excuse Me* and *Hey*.

This effect suggests a particular account of the role of linguistic priming in determining participants' robot-directed utterances in wakeword-based interaction. For robot-directed utterances, we are not interested in lexical priming, as there was no reason to expect that the required wakewords would re-occur in the post-wakeword clause. This leaves syntactic and semantic priming.

We expected both syntactic and semantic priming to play key roles in determining ISA use in the post-wakeword clauses of robot-directed utterances, but expected syntactic priming to be more important. That is, we expected syntactic priming to produce large differences between *Excuse Me* / *Hey* and *Please*, as we thought it would be more syntactically natural to follow *Excuse Me* and *Hey* with an indirect request, and more natural to follow *Please* with a command. We then expected semantic priming to increase ISA use in both the *Excuse Me* and *Please* conditions, creating a difference between the polite *Excuse Me* and the impolite *Hey*, and reducing the difference between the impolite *Hey* and the polite *Please*.

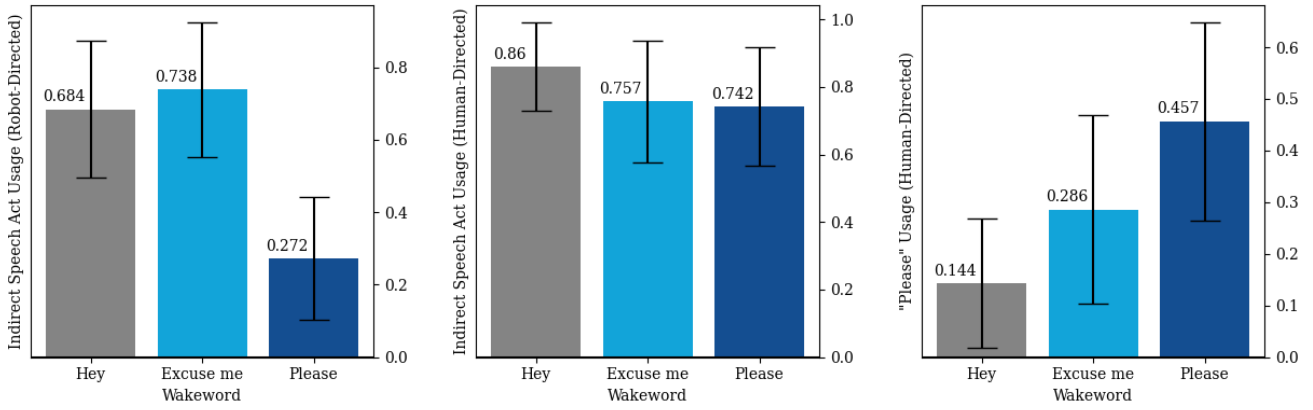
The fact that we observed a difference between *Excuse Me* / *Hey* and *Please* but not between *Excuse Me* and *Hey* suggests that the first part of our prediction was correct, but the second was not. That is, if our general account is correct, ISA use in the post-wakeword clauses of participants' robot-directed utterances was influenced by syntactic priming, but not by semantic priming. Put more plainly, to encourage robot-directed politeness, it was more effective to use a wakeword that made it more syntactically natural to follow-up with an ISA than it was to use an explicitly "polite" wakeword. This finding, which confirms the results of Wen et al. [69] but refutes the results of Williams et al. [70], produces an obvious, concrete design guideline for robot designers:

Design Guideline 1: To encourage robot-directed politeness through wakeword design, robot interaction designers should require the use of wakewords that syntactically prime the use of indirect speech acts by allowing the wakeword clause to be followed by a complete sentence.

These results also suggest the online, text-based experimental paradigm used by Wen et al. [69] had adequate ecological validity to accurately predict robot-directed speech patterns in live interaction.

5.2 ISA Use in Human-Directed Utterances

Our second hypothesis was that different wakewords would lead to differences in human-directed politeness, as assessed by ISA use. Again, we hypothesized that *Excuse me* would promote the greatest ISA use, followed by *Hey*, followed by *Please*. This hypothesis was



(a) Indirect Speech Act Use in Robot-Directed Utterances. **(b) Indirect Speech Act Use in Human-Directed Utterances.** **(c) "Please" Use in Human-Directed Utterances.**

Figure 3: Results from the experiment. Error bars represent 95% Credible Intervals.

partially supported: *Excuse Me* and *Hey* both led to more robot-directed politeness than *Please*, but no difference was found between *Excuse Me* and *Hey*. That is, we expected that robot-directed politeness encouraged by a required wakeword would carry-over into (non-wakeword-based) human-human interaction. We did not observe any such effects. That is, regardless of wakewords' impacts on robot-directed utterances, these effects did not carry over into human-directed utterances. Unlike the results reported by Wen et al. [69], no "ripple effects" were found.

One explanation could be that the influence of syntactic priming ends at the sentence boundary. While syntactic priming constrains the construction of the post-wakeword clauses of a robot-directed utterance, once the utterance is complete, this constraint is removed. However, this would not explain why carry-over effects *were* found by Wen et al. [69]. Similarly, it could be that our human-robot interactions were not long enough to evoke observable effects, but as in other areas of HRI [17, 40, 42, 60], these effects could emerge under longitudinal interactions. But again, this would not explain why carry-over effects *were* found by Wen et al. [69]. We must thus consider the differences between Wen et al. [69]'s testbed and ours, to understand why online, text-based interaction was enabled accurate study of robot-directed utterances but not human-directed utterances. We consider two distinct possibilities.

5.2.1 Online Agent Equivalence Hypothesis. First, it is possible that in Wen et al. [69]'s online, text-based experiment, participants did not buy into the ostensible humanity of their "human" interactants (and may not have even bought into the robotic nature of their "robot" interactants), and treated humans and robots as disembodied intelligent agents, or even just as part of the mechanism of the experimental platform. Under this interpretation, robot and human teammates were treated the same way, and thus the rules that participants learned for how they needed to interact with their robotic teammates carried over into how they interacted with their "human" teammates. This account might be supported if participants in Wen et al. [69]'s experiment continued to use the wakewords required for robot-directed speech in their human-directed speech. This appears to have been the case for the *please* wakeword, but Wen et al. [69] do not report whether this was also true in the other two conditions. Future research could replicate Wen et al. [69]'s online experiment while collecting and reporting this data, and could also use free response questions to try to qualitatively probe participants' rationale. However, these results would say more about online text-based interactions than it would about real-world human-robot interactions.

5.2.2 Social Presence Driven Politeness Hypothesis. Second, it is possible that face-to-face communication imposes more social presence, and thus more awareness of interactants' face needs. Under this explanation, no differences were observed in our in-person study because awareness of interactants' face needs uniformly boosted ISA use. That is, perhaps in in-person interactions, participants were uniformly polite to each other regardless of wakeword use. Comparing general mean use of ISAs in human-directed utterances between our two experiments bears out this hypothesis. While in Wen et al. [69]'s experiment, mean human-directed ISA use varied between 6% and 44% depending on condition, in our experiment, mean human-directed ISA use varied between 74% and 86% depending on condition. Humans are less likely to use ISAs in contexts without highly conventionalized social norms [72], in contexts with increased potential for harm [62], and possibly when speaking to interlocutors in authority positions [45]. Based on these results, future work should replicate our experiment in a context designed to discourage ISA use. The results of such an experiment could either provide evidence for universality of our findings regardless of context (suggesting that human-human politeness may not be achievable through wakeword design alone), or could provide evidence for context-sensitivity of our findings (suggesting that human-human politeness may be achievable through wakeword design only in particular types of task contexts).

5.3 Please Use in Human-Directed Utterances

Our third hypothesis was that different required wakewords would lead to differences in human-directed politeness as assessed by literal use of the word "please". Specifically, we hypothesized that *Please* would promote the greatest "please" use, followed by *Excuse Me*, followed by *Hey*. This hypothesis was partially supported: *Excuse Me* and *Hey* both led to less human-directed "please " use than *Please*, but no difference was found between *Excuse Me* and *Hey*. Similar to robot-directed ISA use, this effect suggests a particular account of the role of linguistic priming in determining participants' human-directed utterances in wakeword-based interaction. Unlike with robot-directed utterances, for human-directed utterances we *are* interested in lexical priming, as interactants could certainly repeat the wakewords required in their robot-directed utterances when speaking with humans; and Wen et al. [69]'s work suggests that at least in the case of "please", this may well be the case.

We expected both semantic and lexical priming to play key roles in determining "please" use in participants' human-directed utterances, but expected lexical priming to be more important. That is, we expected lexical priming to produce large differences between *Excuse Me* / *Hey* and *Please*, based on previous research [69]. We then expected semantic priming to increase "please" use in both the *Excuse Me* and *Please* conditions, creating a difference between the polite *Excuse Me* and the impolite *Hey*, and exacerbating the difference between the impolite *Hey* and the polite *Please*.

The fact that we observed a difference between *Excuse Me* / *Hey* and *Please* but not *necessarily* between *Excuse Me* and *Hey* suggests that, similar to our first hypothesis, the first part of our prediction was correct, but the second *may* not have been correct. That is, if our general account is correct, "please" use in human-directed utterances was influenced by lexical priming, but not *necessarily* by semantic priming. We caveat these claims here because our Bayes Factor analysis produced evidence against differences between *Hey* and *Excuse me*, and between *Excuse me* and *Please*, but this evidence was not strong enough to completely rule out an effect; and visual inspection certainly suggests that our originally hypothesized ordering could still be accurate after all. Overall, however, our results generally suggest that to encourage human-directed "please" use, it was more effective to explicitly require *please* than it was to generally encourage politeness, e.g. through *Excuse Me*, and that more data would need to be collected to make any concrete claims about the existence or nonexistence of semantic priming effects. This presents a clear direction for future work. Even without a conclusive ruling on semantic priming effects, however, our findings in this work confirm the results of Wen et al. [69], and again produces an obvious, concrete design guideline for robot designers:

Design Guideline 2: To encourage the use of "please" in human-human conversation, robot interaction designers should require the use of wakewords that lexically prime this keyword.

6 LIMITATIONS AND FUTURE WORK

Before we conclude, we will describe several key limitations of our experiment, as well as possible directions for future work. First, as we discussed in the previous section, the human-robot interaction designed in this study may not have lasted long enough to cause carry-over effects. Future work should explore longer term interactions, as well as repeated interactions across multiple sessions.

Second, the structure of our experimental task requires participants to pass very specific information to the designated agent. Specifically, even though participants have the freedom to organize their language however they want (except for using the assigned wakeword), participants often tended to say one sentence in each round of robot-directed or human-directed communication, and the entire interaction (ten rounds in total) usually only last around five to seven minutes. Future work should encourage more free-form human-human interactions in order to explore how carry-over effects might manifest outside of single-shot utterance generation.

Finally, as shown in previous HRI research [30], gender plays a critical role in both human-directed and robot-directed politeness. While we did not examine gender effects in this experiment, our human confederates all identified as women, and the Pepper robot is typically perceived as more feminine than masculine [50]. In contrast, the majority of our participants were men. Given the gendered nature of both robot- and human-directed politeness, this may have influenced the politeness dynamics observed in our work. Future work should explore how the gender representation of the robot, human teammate, and participant might interactively influence and inform the politeness dynamics of both human-robot and human-human communication.

7 CONCLUSION

Like many researchers within the HRI community, we hope that the future of verbal human-robot interaction will not be one that necessitates the use of wakewords. But in the immediate future, accurate and privacy preserving speech recognition may nevertheless necessitate this interaction paradigm. As we have discussed in this paper, this presents new challenges that must be addressed (avoiding wakeword-driven encouragement of impoliteness) but also new opportunities that can be seized (pursuing wakeword-driven encouragement of politeness). Our results in this work help us to make sense of the partial and conflicting results obtained by previous attempts to study these challenges and opportunities, and suggest a specific account of how different types of linguistic priming interact to determine the specific types of influence that wakeword choices can have on different types (ISA use and "Please" use) in robot-directed and human-directed language. Moreover, our results provide clear evidence that different wakewords may serve different goals, with "Excuse me, ⟨Name⟩" encouraging robot-directed politeness, and "⟨Name⟩ Please" encouraging human-directed "please" use, which produces a definite effect but risks accidental encouragement of impoliteness. Finally, these results produce clear directions for future work, with new scientific hypotheses that can be tested to develop an even clearer understanding of the origins and extents of the phenomena observed in this work.

ACKNOWLEDGMENTS

This work was funded in part by NSF grant IIS-1849348 and IIS-1909847. We thank Nichole Starr, Anjana Radha, Poulomi Pal, Thao Phung, Terran Mott, Alexa Bejarano, Gabriel Del Castillo, Polina Rygina, Shane Romero, Rena Zhu, Sebastian Negrete-Alamillo, Cloe Emnett, and the students of Mines' Spring 2020 Robot Ethics class for assisting with data collection.

REFERENCES

[1] EC Baig. 2018. Kids were being rude to alexa, so amazon updated it. *USA Today* (2018).

[2] John A Bargh, Mark Chen, and Lara Burrows. 1996. Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of personality and social psychology* 71, 2 (1996), 230.

[3] 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*.

[4] BBC. 2018. Amazon Alexa to reward kids who say: 'Please'. https://www.bbc.com/news/technology-43897516.

[5] James O Berger and Thomas Sellke. 1987. Testing a Point Null Hypothesis: The Irreconcilability of p-values and Evidence. *Journal of the American Statistical Association (ASA)* 82, 397 (1987).

[6] J Kathryn Bock. 1986. Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 12, 4 (1986), 575.

[7] Michael Bonfert, Maximilian Spliethöver, 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.

[8] Holly Branigan. 2007. Syntactic priming. *Language and Linguistics Compass* 1, 1-2 (2007), 1–16.

[9] Holly P Branigan, Martin J Pickering, Simon P Liversedge, Andrew J Stewart, and Thomas P Urbach. 1995. Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research* 24, 6 (1995), 489–506.

[10] 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.

[11] Penelope Brown, Stephen C Levinson, and Stephen C Levinson. 1987. *Politeness: Some universals in language usage*. Vol. 4. Cambridge university press.

[12] Massimiliano L Cappuccio, Anco Peeters, and William McDonald. 2020. Sympathy for Dolores: Moral consideration for robots based on virtue and recognition. *Philosophy & Technology* 33, 1 (2020), 9–31.

[13] 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.

[14] Simon Coghlan, Frank Vetere, Jenny Waycott, and Barbara Barbosa Neves. 2019. Could social robots make us kinder or crueller to humans and animals? *International Journal of Social Robotics* 11, 5 (2019), 741–751.

[15] 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*.

[16] Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A Computational Approach to Politeness with Application to Social Factors. In *51st Annual Meeting of the Association for Computational Linguistics*. ACL, 250–259.

[17] Ewart J De Visser, Marieke MM Peeters, Malte F Jung, Spencer Kohn, Tyler H Shaw, Richard Pak, and Mark A Neerincx. 2020. Towards a theory of longitudinal trust calibration in human–robot teams. *International journal of social robotics* 12, 2 (2020), 459–478.

[18] Donald J Foss. 1982. A discourse on semantic priming. *Cognitive psychology* 14, 4 (1982), 590–607.

[19] Elisa Giaccardi, Pieter Desmet, and Olya Kudina. 2020. The Repertoire of Meaningful Voice Interactions How to Design Good Smart Speakers. *The State of Responsible IoT 2020* (2020).

[20] Ken Gordon. 2018. Alexa and the age of casual rudeness. *The Atlantic* (2018).

[21] 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.

[22] 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.

[23] Michael Hoey. 2012. *Lexical priming: A new theory of words and language*. Routledge.

[24] Ryan Blake Jackson, Ruchen Wen, and Tom Williams. 2019. Tact in noncompliance: The need for pragmatically apt responses to unethical commands. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. 499–505.

[25] Ryan Blake Jackson and Tom Williams. 2019. Language-capable robots may inadvertently weaken human moral norms. In *Companion of the 14th ACM/IEEE International Conference on Human-Robot Interaction (alt.HRI)*. IEEE, 401–410.

[26] Ryan Blake 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*. 401–410.

[27] Ryan Blake Jackson and Tom Williams. 2021. A Theory of Social Agency for Human-Robot Interaction. *Frontiers in Robotics and AI* (2021).

[28] Ryan Blake Jackson and Tom Williams. 2022. Enabling morally sensitive robotic clarification requests. *ACM Transactions on Human-Robot Interaction (THRI)* 11, 2 (2022), 1–18.

[29] Ryan Blake Jackson, Tom Williams, and Nicole Smith. 2020. Exploring the role of gender in perceptions of robotic noncompliance. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*. 559–567.

[30] Ryan Blake Jackson, Tom Williams, and Nicole Smith. 2020. Exploring the Role of Gender in Perceptions of Robotic Noncompliance. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 559–567.

[31] Andrew F. Jarosz and Jennifer Wiley. 2014. What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. *The Journal of Problem Solving* 7 (2014).

[32] JASP Team et al. 2016. Jasp. *Version 0.8. 0.0. software* (2016).

[33] Malte F. Jung, Nikolas Martelaro, and Pamela J. Hinds. 2015. Using Robots to Moderate Team Conflict: The Case of Repairing Violations. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM, 229–236.

[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 (HRI)*. 198–199.

[35] Veton Këpuska. 2011. Wake-up-word speech recognition. *Speech Technologies* (2011), 237–262.

[36] Boyoung Kim and Joanna Korman. 2022. Treading lightly toward behavior change: Moral feedback from a robot on microaggressions. In *RSS Workshop on Social Intelligence in Humans and Robots*.

[37] Boyoung Kim and Elizabeth Phillips. 2021. Robots as Legitimate Moral Regulators: Humans' Assessment of Fairness based on the Proportionality of Punishment. In *IROS Workshop on Building and Evaluating Ethical Robotic Systems*.

[38] Boyoung Kim, Ruchen Wen, Ewart J de Visser, Qin Zhu, Tom Williams, and Elizabeth Phillips. 2021. Investigating robot moral advice to deter cheating behavior. In *TSAR Workshop at ROMAN 2021*.

[39] Taemie Kim and Pamela Hinds. 2006. Who Should I Blame? Effects of Autonomy and Transparency on Attributions in Human-Robot Interaction. In *RO-MAN*.

[40] Kheng Lee Koay, Dag Sverre Syrdal, Michael L Walters, and Kerstin Dautenhahn. 2007. Living with robots: Investigating the habituation effect in participants' preferences during a longitudinal human-robot interaction study. In *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 564–569.

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

[42] Min Kyung Lee, Jodi Forlizzi, Sara Kiesler, Paul Rybski, John Antanitis, and Sarun Savetsila. 2012. Personalization in HRI: A longitudinal field experiment. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 319–326.

[43] Min Kyung Lee, Sara Kiesler, Jodi Forlizzi, and Paul Rybski. 2012. Ripple effects of an embedded social agent: a field study of a social robot in the workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 695–704.

[44] Baisong Liu, Daniel Tetteroo, and Panos Markopoulos. 2022. A Systematic Review of Experimental Work on Persuasive Social Robots. *International Journal of Social Robotics* (2022), 1–40.

[45] 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*.

[46] Cynthia Matuszek. 2018. Grounded Language Learning: Where Robotics and NLP Meet.. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*. 5687–5691.

[47] Nikolaos Mavridis. 2015. A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems* 63 (2015), 22–35.

[48] Timothy P McNamara. 2005. *Semantic priming: Perspectives from memory and word recognition*. Psychology Press.

[49] Aidan Naughton and Tom Williams. 2021. How to Tune Your Draggin': Can Body Language Mitigate Face Threat in Robotic Noncompliance?. In *International Conference on Social Robotics*. Springer, 247–256.

[50] Giulia Perugia, Stefano Guidi, Margherita Bicchi, and Oronzo Parlangeli. 2022. The Shape of Our Bias: Perceived Age and Gender in the Humanoid Robots of the ABOT Database. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction* (Sapporo, Hokkaido, Japan) *(HRI '22)*. IEEE Press, 110–119.

[51] Martin J Pickering and Holly P Branigan. 1999. Syntactic priming in language production. *Trends in cognitive sciences* 3, 4 (1999), 136–141.

[52] Martin J Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences* 27, 2 (2004), 169–190.

[53] 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.

[54] Jeffrey N Rouder, Paul L Speckman, Dongchu Sun, Richard D Morey, and Geoffrey Iverson. 2009. Bayesian t tests for Accepting and Rejecting the Null Hypothesis. *Psychonomic Bulletin & Review* 16, 2 (2009), 225–237.

[55] Eduardo Benítez 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)*.

[56] 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.

[57] Matthias Scheutz, Paul Schermerhorn, James Kramer, and David Anderson. 2007. First steps toward natural human-like HRI. *Autonomous Robots* 22, 4 (2007), 411–423.

[58] Lea Schönherr, Maximilian Golla, Thorsten Eisenhofer, Jan Wiele, Dorothea Kolossa, and Thorsten Holz. 2020. Unacceptable, where is my privacy? exploring accidental triggers of smart speakers. *arXiv preprint arXiv:2008.00508* (2020).

[59] Sukyung Seok, Eunji Hwang, Jongsuk Choi, and Yoonseob Lim. 2022. Cultural Differences in Indirect Speech Act Use and Politeness in Human-Robot Interaction. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*. 470–477.

[60] Sofia Serholt and Wolmet Barendregt. 2016. Robots tutoring children: Longitudinal evaluation of social engagement in child-robot interaction. In *Proceedings of the 9th nordic conference on human-computer interaction*. 1–10.

[61] Joseph P Simmons, Leif D Nelson, and Uri Simonsohn. 2011. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science* 11 (2011).

[62] Cailyn Smith, Charlotte Gorgemans, Ruchen Wen, Saad Elbeleidy, Sayanti Roy, and Tom Williams. 2022. Leveraging Intentional Factors and Task Context to Predict Linguistic Norm Adherence. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 44.

[63] Robert Sparrow. 2016. Kicking a robot dog. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 229–229.

[64] Jonathan AC Sterne and George Davey Smith. 2001. Sifting the Evidence – What's Wrong with Significance Tests? *Physical Therapy* 81, 8 (2001), 1464–1469.

[65] 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.

[66] J Vincent. 2018. Google Assistant adds Pretty Please functionality, updated lists, and other features. *The Verge* (2018).

[67] Eric-Jan Wagenmakers. 2007. A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review* 14, 5 (2007), 779–804.

[68] Hunter Walk. 2016. Amazon Echo is magical. It's also turning my kid into an asshole. *Hunter Walk* (2016).

[69] 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 *Proceedings of the 2022 ACM international conference on multimodal interaction*.

[70] 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.

[71] Tom Williams, Ryan Blake Jackson, and Jane Lockshin. 2018. A Bayesian Analysis of Moral Norm Malleability during Clarification Dialogues. In *Proceedings of the Annual Meeting of the Cognitive Science Society (COGSCI)*. Cognitive Science Society, Madison, WI.

[72] 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 (HRI)*.

[73] Tom Williams, Qin Zhu, Ruchen Wen, and Ewart J de Visser. 2020. The Confucian Matador: Three Defenses Against the Mechanical Bull. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (alt.HRI)*. 25–33.

[74] 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.