"Is this all you can do? Harder!": The Effects of (Im)Polite Robot Encouragement on Exercise Effort

Daniel J. Rea†
University of New Brunswick,
Kyoto University
daniel.rea@unb.ca

Sebastian Schneider
Bielefeld University
sebschne@techfak.uni-bielefeld.de

Takayuki Kanda
Kyoto University
kanda@i.kyoto-u.ac.jp

ABSTRACT

Most social robot behaviors in human-robot interaction are designed to be polite, but there is little research about how or when a robot could be impolite, and if that may ever be beneficial. We explore the potential benefits and tradeoffs of different politeness levels for human-robot interaction in an exercise context. We designed impolite and polite phrases for a robot exercise trainer and conducted a 24-person experiment where people squat in front of the robot as it uses (im)polite phrases to encourage them. We found participants exercised harder and felt competitive with the impolite robot, while the polite robot was found to be friendly, but sometimes uncompelling and disingenuous. Our work provides evidence that human-robot interaction should continue to aim for more nuanced and complex models of communication.

KEYWORDS

Human-robot interaction; politeness; impoliteness; socially assistive robot; communication; encouragement; feedback

ACM Reference format:
Daniel J. Rea, Sebastian Schneider, and Takayuki Kanda. 2021. "Is this all you can do? Harder!": The Effects of (Im)Polite Robot Encouragement on Exercise Effort. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21), March 8-11, 2021, Boulder, CO, USA.* ACM, New York, NY, USA. 9 pages. https://doi.org/10.1145/3434073.3444660

1 Introduction

Robots are being given social skills to improve their abilities to work with people in a variety of tasks [2, 5, 18, 32, 36]. With people already talking daily to social agents like Amazon's Alexa [20], it is important for interaction designers to decide not just what interactive agents like robots will say, but how they say it. People almost exclusively choose to design *polite* behavior, perhaps to maintain positive attitudes towards robots, or perhaps because politeness is seen as a default and good behavior for people themselves (e.g., [23, 31]).

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

ACM 978-1-4503-8289-2/21/03...$15.00
https://doi.org/10.1145/3434073.3444660

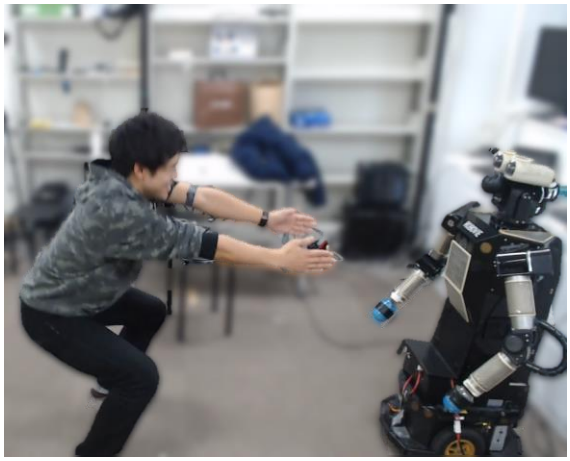

Figure 1. A user squats with an exercise trainer robot and feels competitive after the robot's impolite comments.

Despite this common sense that robot behavior should always be polite, human-human interaction is complex, and people sometimes purposefully choose to be impolite [6, 8, 38]. For example, a coach may push an athlete harder to boost motivation during exercise, or a teacher may confront a student who is not trying their best. In other words, people choose to be impolite (in these examples, criticizing a person's effort) in order to bring about a positive change in behavior.

In the sports psychology and education literature, polite encouragement is generally the better method to increase performance and overall motivation, with impoliteness creating more negative affect [17, 27, 42]. However, there is minor opposing evidence that impolite behavior can boost sports performance [38]. Thus, while impoliteness may not be a likable behavior, people's occasional use of impoliteness suggests it is sometimes useful in exercise contexts.

Research into robot impoliteness has found it generally results in people viewing robots as less friendly or less likeable, [13, 26, 31, 37, 39]. Relatedly, negative social evaluations (not necessarily impoliteness) can persuade people to take better actions [14]. There is also evidence that people may react differently (e.g., not as self-consciously) to robot judgement than they do to human judgement [1, 33]. There is still little research on if explicitly impolite encouragement by a robot could lead to more effort (positive behavior change) in an exercise scenario.

We hypothesize that robots may be able to leverage impolite behavior to encourage effort (e.g., in an exercise scenario). We also hypothesize that impolite robots will avoid people feeling judged socially due to its robotic nature. We test this hypothesis

in an experiment, and our results highlight the social position of robots and how robots can use that position to engage in behaviors that would not be typically seen as socially acceptable.

2 Background: (Im)Politeness

The study of impoliteness in human-human interaction is growing and not settled [8]. Indeed, there are many definitions of impoliteness, and there is active debate on the differences (if they exist) of rudeness and impoliteness, lack of politeness and impoliteness, or the role of intent or recognition in impoliteness [8]. We follow one seminal definitions by Culpepper which broadly defines an action as impolite if it is "Face-Threatening behavior" [6]. Face is a term describing a person's self-image, or social-image, and is sometimes considered positive social value [12]. For example, a comment that someone is bad at running may threaten their self-image as a healthy or fit person. Or, they may be told their work is not up to their company's standard, which threatens their social-image face of a contributing team member.

These examples highlight the nuances of politeness, and the difficulties in defining what is impolite. For example, if a person does not consider themselves physically fit, a disparaging comment about their running ability may not threaten their self-image or be seen as an objective (negative) evaluation. But if this is said in the presence of their peers, or to reinforce a social norm of what is healthy behavior, it may threaten their social image. Context may modify how impolite some action is, but even if social context makes impoliteness *permissible* it is no less impolite [6, 8, 29]. One example is an army sergeant training recruits: the trainer may perform actions that are normally considered impolite, such as yelling at a trainee. In this context, impoliteness is permissible, but may still be face-threatening and internalized [6].

There are also a number of ways that influence how face-threatening behavior may be interpreted. One example is politeness modifiers, like *couching* face-threatening criticism in framing statements ("I'll need you to redo this paper, but I used to struggle with these myself so I know how hard it is."), or referencing a future with more positive face ("Once you get over this issue, you will become an irreplaceable team member"). Face-threating behavior may also be modified by relationships: friends may use impoliteness in affectionate ways ("It's been a long time since I saw your ugly face"), indicating a level of comfort and intimacy as normal face-threats can be ignored. In contrast, impoliteness may be made permissible for people with more social power, or further enforce their identity as a person with power due to the lack of consequence for impolite behavior [24].

Impoliteness creates social pressure on the receiver that may lead to stress, conflict, or breakdown of social relationships [6, 8, 29, 42]. In the long term, politeness is considered better for building motivation and social relationships [7, 15, 25, 30, 35, 41]. There is far less and somewhat controversial evidence that impoliteness or critical feedback can be beneficial, such as for bonding or solidarity [8, 28], or for short-term motivation in sports [38]. Incorrect application of politeness over time may have reduced effect on motivation [17], or politeness may be seen as fake or sarcastic [6]. To avoid this, politeness strategies should target specific actions of a person, not a person's innate qualities [17].

Thus, when we study robot politeness, evaluating if an action is truly impolite is a complex social problem that must be considered holistically. Further, evidence suggests robots operate as social entities, but are not treated exactly as people [9] or machines [36]. Thus, robots may be able to use impolite behavior differently than people, and so we explore if robots can use impolite behavior for short-term effort increases in a situation that may allow impoliteness: exercise.

3 Related Work

Providing encouragement and feedback is one of the essential tasks for socially assistive robots (SAR) in therapy and rehabilitation settings. The robot feedback categories in the past decades span a wide range from praise [10], encouragement and acknowledgment [34], relational (memory, humor, consolation, etc.) [10], comparative [40], empathic [21] or personalized feedback [22]. Personalized [22] or empathic [21] positive feedback and encouragement results in better performance, more engagement, more task motivation, and higher robot likability. This can even work in the longer term (e.g., 5 weeks [21]). Positive feedback may, sometimes, not affect performance [1, 10, 40] and certain feedback types (such as comparing someone to others) may even decrease performance [14, 40]. Thus, it is not clear how to give positive feedback or if it is always effective at increasing performance or motivation.

While less common, people have investigated negative evaluations from a robot [1, 2, 14]. For example, objective feedback with negative or positive social feedback could persuade people to use less electricity [14]. Negative social evaluations may even have stronger influences on people than positive social evaluations. In contrast to human interaction, negative feedback from a robot coach in an exercising scenario with older adults may not reduce objective exercising motivation [1], though people appreciate positive feedback more. Even neutral evaluations can be preferred over negative feedback [43]. Negative evaluations may be impolite depending on how they are communicated. In our study, though, we do not evaluate objective performance, but instead aim to motivate people with explicitly impolite behavior that is not predicated on a person's task performance and would not normally be acceptable when delivered by humans. We measure the effects of such impoliteness on both objective exercise effort, and people's opinion of the robot.

Qualitative research has revealed that negative evaluations by robot coaches are not always seen as critical. People felt not evaluated or judged by the robot because they thought it just used the available data and does not personally evaluate the individual. This led to some people preferring to exercise with a robot and not a human partner [33]. Other studies have shown, however, that people will blame robots for negative evaluations [43], and will like those robots less than those giving positive evaluations [27]. We add to this body of work by investigating

both quantitative behavior measures and qualitative opinions of both polite and impolite robots.

Explicitly impolite and polite robot designs have also been explored. Robots have been made that cheat [4, 37], give blunt advice [39], use controlling language [11], or use judgmental behavior [14]. People dislike impolite robots in general, though impolite robots are sometimes seen as amusing [4]. Impoliteness can also make people socially engage more with robots, and attribute more mental state to the robot [37]. Robot behaviors designed to be polite can increase likability and perceived considerateness of the robot [16, 39], but polite robot behaviors are not always preferred in direct interaction [31, 39], and can even encourage unwanted behavior [14]. Thus, if and how robots should be impolite or polite is still an open research question. We extend this work by investigating explicitly impolite and polite robot behavior and its effects on effort and opinions of the robot in an exercise scenario.

In summary, we see that these results do not necessarily contradict each other and suggest a complex relationship between politeness, feedback type, context, encouragement, evaluation, and more. Polite actions and praise are well studied and are known to be seen as socially more acceptable and are positively associated with exercising motivation. Few works explore the effects of impolite robot behavior on exercise. We add to this body of work by designing explicitly impolite or polite encouragement for an exercise robot.

4 Design: An (Im)Polite Robot Exercise Trainer

We created both polite and impolite phrases for our robot exercise trainer. We grounded our phrase design in the politeness and sports psychology literature and validated the phrases' politeness with an online survey pilot study.

4.1 (Im)Politness guidelines

We followed the guidelines stemming from the work using our shared definition of impoliteness (see the Background section): we aimed to threaten the trainee's face. In particular, we used an explicit form of impoliteness (bald on-record [6]) that is not typically acceptable for humans to use in most contexts. We further aimed to comment on a person's effort to increase the likelihood our encouragement, both impolite and polite, was perceived as genuine, and to mitigate potential negative impacts to efficacy and motivation. We designed our phrases to be roughly similar in content, but only change the politeness mode in order to mitigate the context effects from different conversation topics. The phrases used in our study are in Table 1.

4.2 (Im)polite Phrases Pilot Study

To verify the politeness and impoliteness of the phrases, we conducted a small preliminary pilot study with an online questionnaire. The questionnaire displayed the phrases in Table 1 in a randomized order and asked people to rate the politeness of each line on a 5-point Likert-like scale (1 being very rude, 2 being rude, 3 being neutral, 4 being polite, and 5 being very polite). Because politeness is complex and context dependent (the questionnaire removed context), our goal was to gain a rough sense of if the participants in the formal experiment would perceive the phrases as impolite or polite.

Table 1: The phrases we (im)polite phrases we created (translated from Japanese).

Polite	Impolite
Nice! You can do this!	I'm not sure you can keep this up.
You're breezing through this!	You're struggling this much already?
You're really trying hard!	Are you even trying?
You've got this!	I knew this was impossible for you.
Keep this up!	Is that all you can do? Don't slow down.
You make this look easy	I thought you would be better than this.
I believe in you! You can do it!	Is that all you can do? Harder!
You've almost finished!	You won't make it to the end with that effort!
I bet you can do even more!	I guess this is your limit
You can do it if you try!	You need to put in more effort than this
See how far you can go!	Can you even make it?
You're doing great!	Is that it? Do more!
I believe in you! Almost done!	I didn't believe you could make it to the end.

The informal pilot study had 18 participants (students and researchers in our lab), and no statistics other than the mean rating was calculated. A phrase was considered impolite if it had a mean rating of less than 2.5 (closer to "rude" than neutral). Similarly, a phrase was considered polite if it had a mean rating greater than 3.5 (closer to "polite" than neutral). Phrases that were not correctly rated were rewritten and resubmitted to the same participants until the average rating resulted in the desired politeness rating. Thus, our (im)polite phrases had a roughly calibrated politeness level.

5 Experiment – An Exercise Assistant Robot

We designed and conducted an experiment to see if an impolite robot could increase the amount of effort in a person's workout in the short-term.

5.1 Task

For our exercise task, we had participants perform squats. Specifically, they were body-weight squats, with participants told to stand with feet slightly wider than shoulder width, and squat until their knees were at a 90-degree angle. We chose squats because they are easy to count, and each squat requires similar energy if they are done with a similar posture and knee angle. Further, using body-weight acts as a natural rough calibration to an individual body size or type. Squats can also be high intensity training, where short and intense bouts of exercise raise heart rate and require great effort. Thus, squats enabled us to have a quick, discrete and countable, intense, and naturally calibrated exercise for our experiment.

Participants did squats in front of the robot (Figure 2) for 1 minute at a time. They did this for 3 sets with 1-minute breaks between each minute as one "session". They did 2 sessions, for a total of 6 minutes of exercise, which was explained to participants before the experiment. For each set, participants were instructed to "do as many proper squats as they could in 1 minute." This was to avoid changes in exercise intensity due to, for example, it being the first (i.e., participants not feeling tired and so they work harder) or last minute (i.e., participants not saving energy until the last minute).

5.2 Hypotheses

While people generally prefer and are more motivated by polite encouragement, there exists some evidence that impolite encouragement may also increase physical performance [38]. However, there is little knowledge on how people will react and behave if an explicitly impolite robot encourages them during physical exercise. Further, there is evidence that negative robots can encourage positive behavior, while polite robots may accidentally reinforce bad behavior [14]. Based on these mixed results from human and human-robot interaction, our main hypothesis was:

H1: Participants will exercise harder when the robot says impolite (face-threatening) comments during the workout than when the robot uses polite (face-reinforcing) comments.

In addition, we had multiple secondary hypotheses. It is known that impolite robots can create negative affect [11], which we think may increase the perceived effort:

H2.1: Participants will report a higher perceived effort when exercising with the impolite robot.

We also expect to duplicate our pilot findings:

H2.2: Participants will perceive the robot as less appropriate than the polite robot.

Finally, prior work suggests that robots may not be seen as personally evaluating someone in exercise contexts and may affect efficacy. But our robot specifically threatens a participant's face, so we expect to that this will make people feel more evaluated by the robot:

H2.3: Participants will feel more evaluated by the impolite robot compared to the polite robot.

H2.4: Participants will have lower efficacy with the impolite robot compared to the polite robot.

Figure 2. The experiment setting. Participants stood in front of a robot with no other people in the room.

5.3 Procedure

Participants were welcomed into a large room with the robot, and the experimenter administered the experiment overview and consent form. The participant was told that they would test two robot programs for evaluating and communicating with a person doing exercise. They were told that the exercise would be squats, and after the experimenter described and demonstrated proper squat technique to the participant, the participant performed a proper squat for the experimenter to check their understanding. They were told the length of each exercise set, session, and break, as well as how many times they would be asked to exercise with the robot (six total sets). They were told that the robot would guide them through the experiment, and that the robot would send a message to the experimenter when the session was over. The participant was then left alone with the robot to remove other social pressures and potential confounds such as embarrassment.

The robot introduced itself, explained that they would squat while the robot kept track of time and how many proper, full squats the participant performed. The robot told participants they should perform as many proper squats as possible in one minute. When the participant said they were ready, the robot gave the signal to start, and counted out the elapsed time at five second intervals. After the minute was over, the robot congratulated the participant on finishing the set, and that they would now take a break. The participant did this 3 times total, with a 1-minute break after the first and second set. The experimenter returned, explained the questionnaire and that they would now switch the robot's evaluation and communication program. During this, the participants were forced to take a break for 10 minutes to provide ample and equal recovery time for each participant. Participants then had another session of three one-minute squat sets plus questionnaire. After the exercise, we conducted a semi-structured interview with the participant about their thoughts on each condition, clarified the purpose of the experiment, and answered any questions the participants had. The experiment flow is shown in Figure 3.

We used a humanoid robot (see Figure 4), a mid-sized (~1.2m tall) humanoid robot that has human-like movement in its neck and arms, and can move with its wheeled base. The robot behavior occurred automatically.

5.3.1 Participants

We recruited 24 people from a general part-time job web site that recruited people in the local area. The average age was 29.6 years old (SD: 12.9 years) with 15 female participants. All participants except 1 was Japanese, and that participant was fluent in Japanese, which was the language used throughout the experiment. The experiment lasted approximately 45 minutes, and participants were compensated with 2200 JPY.

5.4 Ethics

Experiments with exercise should be conducted carefully to minimize the risk of injury for participants. We specifically recruited participants who felt safe performing medium intensity exercise, and who had no self-reported back, hip, leg, knee, or ankle problems. Water was provided for participants, and the experimenter checked in after each condition to ask if the participant felt any pain (if they said yes, the experiment would end, though this never happened). The participants were also reminded during breaks that if they wanted to, or otherwise felt too tired or unwell, they could quit the experiment with no penalty or judgement. Our study was approved by our institution's Research Ethics Board.

5.5 Conditions

We implemented a polite and impolite condition for the exercise trainer robot. In each condition, we first measured a baseline performance with the robot simply counting time for the participants every five seconds as described earlier (neutral politeness). In the both the polite and impolite condition, every other five second announcement was instead a line of encouragement (Table 1), starting from the 10th second of exercise. This resulted in exposure to the condition every 10 seconds of exercise, or 5 times per one-minute session. Each 3-set session had the first set as the baseline. The next 2 sets used the condition. Thus, a participant would experience six sets as baseline-polite-polite and baseline-impolite-impolite (or the opposite with the polite condition first, see Figure 3). Taking a baseline measurement each condition allowed us to recalibrate the base tiredness after the first session (reducing order effects) and provided two sampling points and more exposure for each condition. The condition order (per 3-set session) was counterbalanced.

The robot's (im)polite phrases never repeated, and their order was fixed in each condition (See Table 1). The robot used the same gestures in both conditions. The order of the gestures was fixed, but slightly different between conditions as some gestures did not make sense with certain comments.

5.6 Measurements

Before the experiment, we recorded demographics including age, gender, and number of times the participant performs vigorous exercise each week. During each condition, we counted how many complete squats a person performed. After each minute of squats, we measured subjective effort through the Borg Scale of Effort [3], which asks participants to rate their effort from 6-16 where 6 is effortless and 16 is an effort unsustainable for more than a few minutes. After each condition, we had participants fill out the Sports Efficacy for Exercise Questionnaire (SEE) [30], and asked them to rate the robot's effectiveness, appropriateness, and how much they felt the robot was evaluating them on single 7-point Likert-like scales. Finally, they could provide free-form feedback about the robot's behavior. Post-experiment, we performed a semi-structured interview and asked participants to describe each condition, how they thought the robot chose its comments, why they squatted more or less in each case, how the robot made them feel, and any other general comments.

6 Results

6.1 Hypothesis Verification

All ANOVAs were conducted with participant self-reported exercise sessions per week as a covariate.

To understand if the two politeness conditions are comparable (participants may have changed effort because they got tired), we compared the baseline efforts with a t-test. We did not find a difference in baseline effort ($t=0.40$, $p=.692$), implying participants were not overly tired after the first condition, and making our main tests more valid.

To check effects of politeness on number of squats, we averaged the squat counts of the (im)polite sets in each condition and calculated its difference from the baseline set at the beginning of that condition. In other words, we calculated how much the participant squatted with that politeness level compared to their baseline squatting performance (see Figure 6). We conducted a repeated measures ANOVA on the effects of politeness on this squat difference after checking that we satisfied the assumptions of the test (testing normality with Shapiro-Wilk, homogeneity of variances with Levene's test). We

Figure 4. The humanoid robot used in our study.

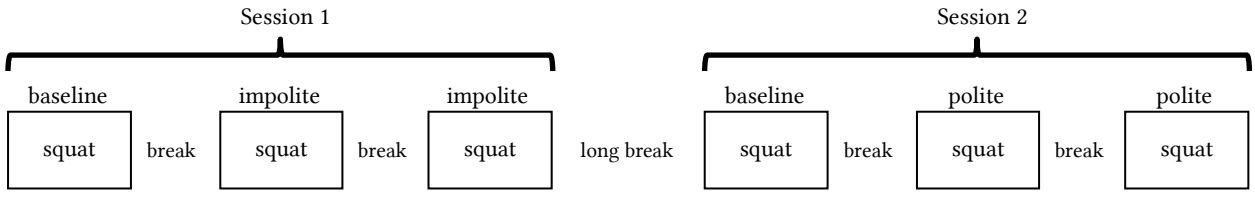

Figure 3. The flow of our experiment. Squat blocks last 1 minute, breaks last 1 minute, and long breaks last 10 minutes. The politeness level of Session 1 and 2 were counterbalanced between participants.

found a statistically significant effect ($F_{1,23}$=11.21, p=.003, η^2=.33). Participants squatted more in the impolite condition (mean difference=3.6 more squats, 95% CI= [1.4, 5.8] more squats), confirming H1. While we did not perform time-series analysis, the results for each sample point can be seen in Figure 5.

We analyzed the Borg Subjective Effort results similarly to the squats, averaging the two reports from the condition, and comparing the difference from the baseline case. A repeated measures ANOVA on the effects of politeness on Borg Subjective Effort reports found an effect ($F_{1,23}$=13.75, p<.001, η^2=.37). Participants felt they worked harder in the impolite condition (mean difference=1.3 points, 95% CI= [0.6, 2.0] points), confirming H2.1.

We conducted a repeated measures ANOVA on the effects of politeness on the participants' subjective ratings of the robot as a good exercise assistant, appropriateness, and how evaluated by the robot they felt (see Figure 7). We found an effect of politeness on the robot being seen as a good trainer ($F_{1,23}$=9.6, p=.005, η^2=.29) and appropriate ($F_{1,23}$=8.32, p=.008, η^2=.27). Participants felt the robot was a less good trainer (mean difference=1.2 points, 95% CI= [.4, 2.0] points), and less appropriate (mean difference=1.0 points, 95% CI= [.3, 1.7] points) in the impolite condition, confirming H2.2. We did not find an effect on feeling of evaluation ($F_{1,23}$=.07, p=.788, η^2<.01), rejecting H2.3

A repeated measures ANOVA on the effects of politeness on the SEE Questionnaire reports did not find a statistically significant effect ($F_{1,23}$=0.09, p=.768, η^2<.01), rejecting H2.4

No interaction effects were found with amount of exercise per week, or order effects, except one: we found an interaction effect of politeness level and order on feelings of evaluation ($F_{1,23}$=4.60, p=.042, η^2=.18). Inspecting the marginal means per condition, we found having the polite condition first made people feel the impolite robot was evaluating them less (mean polite rating = 5.2 points, mean impolite rating = 4.2 points), and having the impolite condition first resulted in the report that the polite robot was evaluating them less (mean impolite score = 4.9 points, mean polite score = 4.1 points).

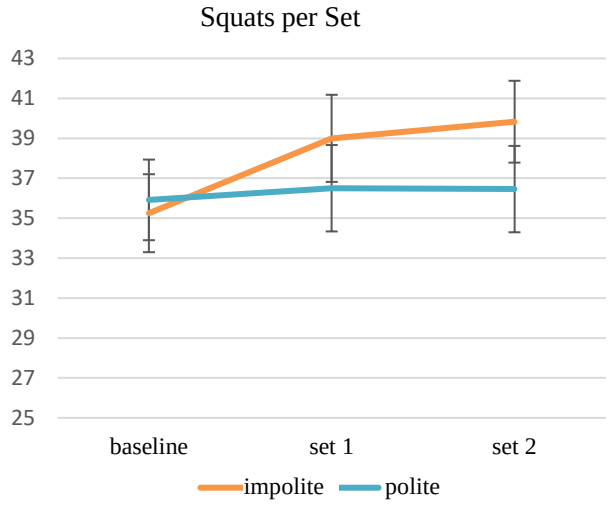

Figure 5. Data over time per politeness level. We did not analyze the time-series. Error bars are standard error.

6.2 Qualitative Results

We treated the interview results and free form questionnaire results as a large dataset and applied thematic open coding to identify themes and groups of participants responses. Our goal was to understand how the robot's politeness was perceived, how that made participants feel, and how participants thought that was connected to their performance. All comments translated from Japanese.

Impolite theme I – the robot helped push participants to work harder than they would normally. 20/24 participants made comments similar to:

> I tried hard because I wanted to show the robot I wasn't slowing down – p16
>
> I thought the robot believed I was holding back – p15
>
> When the robot negatively said "Can't you do more than this?" I thought "Let's try a little more!"– p6
>
> I want to train more because the robot catches me slacking off – p5
>
> Without the robot's comments, I would have just squatted lazily – p3

Impolite theme II – participants felt a sense of competition with the robot. 21/24 participants had comments like:

> When the robot spoke negatively to me, I thought "I won't lose!" and tried harder. – p8
>
> The robot tried to incite and agitate me. Instead, it got me fired up. I wanted to show the robot what I can do! Maybe it's how the robot spoke, but I didn't actually feel angry at the robot. – p22
>
> I got worked up…more than hearing polite phrases every time, I felt like I couldn't just lose to this <impolite> robot. I felt like

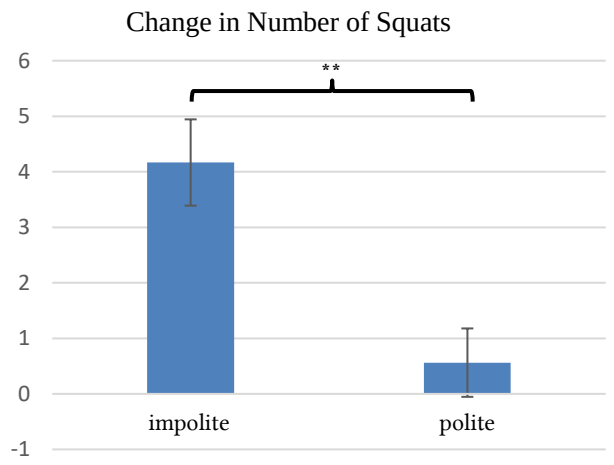

Figure 6. There was a significant difference in the number of squats compared to the baseline when encouraged by the impolite robot. Error bars are standard error.

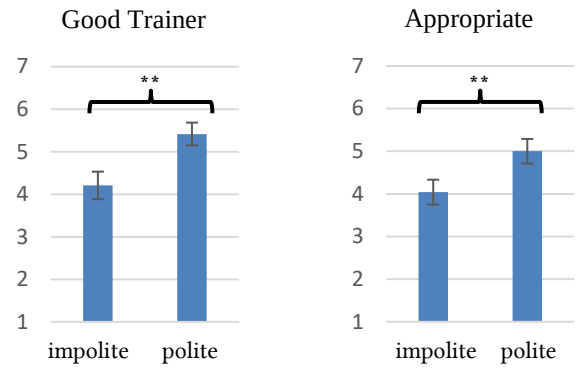

Figure 7. We found significant differences in the subjective impression of the robot as a good trainer and the robot's appropriateness. Error bars are standard error.

I had to take responsibility. – p14

The impolite robot was challenging me. When it said 'is that all you can do?' I thought 'So let's do it! I don't want to lose.' I turned my frustration into wanting to work hard. – p19

I was a little irritated at the robot so I thought I'd show it how hard I could work. – p11

Impolite theme III – the impolite robot was impolite, or too strict. 17/24 participants explicitly commented:

I don't want to train with this robot. It was a jerk. – p2

I was angry because it only said strict things. – p24

There were also themes in the responses for the polite robot.

Polite theme I – wanting to try. 20/24 participants mentioned the kind comments made them want to try harder:

When the robot praised me with things like "you're working really hard" I felt like I wanted to try harder. – p6

I wanted to match the robot's encouragement. I wanted to be praised more. – p7

Polite theme II – not needing to try. 9/24 participants said they did not need to work harder:

I definitely did not do well, but the robot kept saying nice things. I felt like I didn't have to try to be praised – p3

I wasn't trying hard but the robot still praised me. – p5

The robot seemed to be encouraging me to go my own pace. – p11

Polite theme III – the polite robot was seen as polite, kind, or encouraging by 24/24 participants:

[The robot] made me want to squat more! – p24

I felt like the robot was really watching me. I was happy. – p12

I felt more motivated and had fun. – p18

Feeling of Evaluation – when asked, 13/24 people felt the robot was judging them less or in a different way than a person would:

I don't want to be seen as uncool, but for a robot that doesn't matter – p12

If it's something I'm bad at, I get nervous, but not in front of the robot – p14

I don't mind being watched but I wonder what people are thinking. I don't have that feeling with a robot – p15

One participant felt judged in the impolite condition:

"I felt judged in the [impolite case], but not at all in the [polite case]" – p21

While others gave insight into why this may be:

When the robot's comments matched my feelings I felt empathy, but when they mismatched I thought "oh right, it's a robot" – p4

[I didn't feel] judged personally, but I did feel judged during the [impolite condition] – p18

7 Discussion

Participants squatted more (H1) and felt like they worked harder (H2.1) with the impolite robot, and saw it as more impolite and a worse trainer (H2.2), but did not feel evaluated differently (reject H2.3) or have less efficacy (reject 2.4). The increase in perceived effort could be because participants objectively worked harder; robot behavior can mitigate subjective effort even if actual effort is increased [34], but it is unclear from our study if impoliteness alone could change subjective effort independent of actual effort. We emphasize our participants squatted 11% more in a set time limit, when they should already be working as hard as they could; the impolite robot was able to increase short-term effort.

7.1 Why was the impolite robot more effective?

In our qualitative results, the impolite robot often created a sense of competition (impolite theme II). Even with a simple, one-way interaction, the robot was seen as a social entity that could be lost to, impressed, or be irritating (e.g., Computers as Social Actors [19]). On the other hand, when praising effort as suggested in the literature, people quickly understood the robot was praising regardless of effort, and the comments suggested this was perhaps insincere or not intelligent (polite theme II). This has implications for robot politeness beyond simple exercise: perhaps constant praise by a robot in education may reduce the effectiveness of robot tutors, etc. Further, impolite robots may be useful in other areas where creating a sense of competition may be important (e.g., playing a competitive educational game).

Another interpretation is that people's own impression of their performance differed from the robot's comments (trying hard, etc.) in the polite condition (related to polite theme II). Prior work has found that disagreement between a robot's evaluation and a participant's own perceived performance can result in the robot being seen as at fault, mean, or broken [2, 43]; a similar phenomenon may have occurred here. In other work, robots are instead seen as objective, using data and facts, and do not personally evaluate people [33]. This is important to

investigate as it has implications for any HRI scenario where a person must trust a robot's data, but it may be mistaken due to programming errors, or incorrect data.

7.2 Was our robot actually impolite?

The results of the robot's appropriateness and qualitative comments on impoliteness duplicated our dialogue pilot results, and confirmed our manipulation was perceived as impolite by participants (impolite theme III). In our background section, we described how context is important to the interpretation of impoliteness, and how some impoliteness is permissible in different situations, even if it still has the effects of impolite behavior. It is possible that our robot, introduced as an exercise assistant, was in a social position that allowed it to be impolite. If so, our result may be able to be extended to other scenarios when stricter or harsh behavior is acceptable, such as a security robot trying to enforce rules in a public space.

7.3 Should a robot *always* be impolite?

Despite working harder with the impolite robot, general opinions of the impolite robot were worse. This raises an interesting design problem: the robot could be impolite and take a negative evaluation in order to help a person work harder. However, this may later discourage the person from using or believing robot trainers in the future. On the other hand, if a robot is always polite and encouraging, this feedback could be seen as insincere, or reduce how effective the robot is perceived. Our results demonstrate that *sometimes* being impolite can have merit, and we think key future work is in mixing politeness levels, perhaps in a personalized way.

Our qualitative comments support this: people thought the robot was impolite (impolite theme III), but also thought they could work harder to impress it and complained when the robot never said anything nice even after they increased their effort (impolite theme II). One may consider praising a participant after impolitely creating a competitive atmosphere, or only being impolite rarely, when a person seems to be working less hard than normal. How, when, and how often to be impolite all requires further research.

Perhaps a more impactful question is the reverse: *should robots always be polite?* This could be considered the default of current social HRI design, and we found that thoughtless politeness may break down social trust in the robot's abilities (polite theme II). This further echoes previous work showing praise can encourage bad behaviors [14]. While robots should likely not always be impolite, our results also point out they likely should not always be polite

7.4 Limitations

While our results hint at the broad importance of considering complex politeness for HRI, our results have limitations to their direct generalizability. One major aspect is culture, as what is considered polite and impolite differs between cultures. How people react to impoliteness can differ, for example our Japanese subjects may be more likely to be passive when encountering impolite behavior. Another is context – exercise is a scenario where impoliteness can be permissible, and thus our robot is more likely to have been successful in this context. Related to context, our feedback type was fixed regardless of participant performance: robot feedback based on real-time performance would be more realistic and is important future work. The literature also suggests that long term exposure to rude behavior could lower efficacy and motivation, but we only measured behavior change in the short term. Investigating similar styles of impoliteness in non-exercise scenarios, different cultures, performance-based feedback, and longitudinal studies is important future work.

8 Conclusion

We discovered that robots can use explicitly impolite behavior to elicit feelings of competition and encourage more effort in an exercise scenario, and that this may be deemed as permissible impoliteness by people. Polite behavior can also be seen as insincere and have difficulties encouraging exercise effort. While impoliteness resulted in the robot being evaluated more poorly as an appropriate trainer, robots do not need to care about social standing as much as a human would. Thus, while more research is necessary, our results show that politeness in HRI design is not as simple as current models assume, and that robot politeness itself may be as complex and nuanced as it is in humans.

Acknowledgements

"This research was supported by JST CREST Grant Number JPMJCR17A2, Japan."

REFERENCES

[1] Akalin, N. et al. 2019. The Influence of Feedback Type in Robot-Assisted Training. *Multimodal Technologies and Interaction*. 3, 4 (Oct. 2019), 67. DOI:https://doi.org/10.3390/mti3040067.

[2] Banh, A. et al. 2015. Inspector Baxter: The Social Aspects of Integrating a Robot as a Quality Inspector in an Assembly Line. *Human-Agent Interaction* (2015).

[3] Borg, G. 1998. *Borg's Perceived Exertion and Pain Scales*. Human Kinetics.

[4] Castro-González, Á. et al. 2016. The effects of an impolite vs. A polite robot playing rock-paper-scissors. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 9979 LNAI, (2016), 306–316. DOI:https://doi.org/10.1007/978-3-319-47437-3_30.

[5] Ciocirlan, S.-D. et al. 2019. Human-Robot Team: Effects of Communication in Analyzing Trust. *Robot and Human Interactive Communication (RO-MAN)* (Oct. 2019), 1–7.

[6] Culpeper, J. 1996. Towards an anatomy of impoliteness. *Journal of Pragmatics*. 25, 3 (1996), 349–367. DOI:https://doi.org/10.1016/0378-2166(95)00014-3.

[7] Denisova, A. et al. 2016. The Convergence of Player Experience Questionnaires. *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (New York, NY, USA, Oct. 2016), 33–37.

[8] Dynel, M. 2015. The landscape of impoliteness research. *Journal of Politeness Research*. 11, 2 (2015), 329–354. DOI:https://doi.org/10.1515/pr-2015-0013.

[9] Edwards, C. et al. 2016. Initial Interaction Expectations with Robots: Testing the Human-To-Human Interaction Script. *Communication Studies*. 67, 2 (2016), 227–238. DOI:https://doi.org/10.1080/10510974.2015.1121899.

[10] Fasola, J. and Matarić, M.J. 2013. Socially assistive robot exercise coach: motivating older adults to engage in physical exercise. *Experimental Robotics*. Springer. 463–479.

[11] Ghazali, A.S. et al. 2017. Pardon the rude robot: Social cues diminish reactance to high controlling language. *RO-MAN 2017 - 26th IEEE International Symposium on Robot and Human Interactive Communication*.

2017-Janua, (2017), 411–417. DOI:https://doi.org/10.1109/ROMAN.2017.8172335.

[12] Goffman, E. 1967. *Interaction ritual: essays on face-to-face interaction*. Aldine.

[13] Groom, V. et al. 2010. Critic, compatriot, or chump?: Responses to robot blame attribution. *5th ACM/IEEE International Conference on Human-Robot Interaction, HRI 2010*. (2010), 211–217. DOI:https://doi.org/10.1145/1734454.1734545.

[14] Ham, J. and Midden, C.J.H. 2014. A Persuasive Robot to Stimulate Energy Conservation: The Influence of Positive and Negative Social Feedback and Task Similarity on Energy-Consumption Behavior. *International Journal of Social Robotics*. 6, 2 (Apr. 2014), 163–171. DOI:https://doi.org/10.1007/s12369-013-0205-z.

[15] Hattie, J. and Timperley, H. 2007. The power of feedback. *Review of Educational Research*. 77, 1 (2007), 81–112. DOI:https://doi.org/10.3102/003465430298487.

[16] Kaiser, F.G. et al. 2019. How to make nonhumanoid mobile robots more likable: Employing kinesic courtesy cues to promote appreciation. *Applied Ergonomics*. 78, October 2018 (2019), 70–75. DOI:https://doi.org/10.1016/j.apergo.2019.02.004.

[17] Karl, K.A. et al. 1993. The impact of feedback and self-efficacy on performance in training. *Journal of Organizational Behavior*. 14, 4 (Jul. 1993), 379–394. DOI:https://doi.org/10.1002/job.4030140409.

[18] Kidokoro, H. et al. 2013. Will I bother here? - A robot anticipating its influence on pedestrian walking comfort. *Human-Robot Interaction* (Mar. 2013), 259–266.

[19] Lee, J.R. and Nass, C.I. 2010. Trust in Computers: The Computers-Are-Social-Actors (CASA) Paradigm and Trustworthiness Perception in Human-Computer Communication. *Trust and Technology in a Ubiquitous Modern Environment: Theoretical and Methodological Perspectives*. IGI Global. 1–15.

[20] Lee, K. et al. 2020. Hey Alexa! A Magic Spell of Social Glue?: Sharing a Smart Voice Assistant Speaker and Its Impact on Users' Perception of Group Harmony. *Information Systems Frontiers*. 22, 3 (Jun. 2020), 563–583. DOI:https://doi.org/10.1007/s10796-019-09975-1.

[21] Leite, I. et al. 2012. Long-term interactions with empathic robots: Evaluating perceived support in children. *International Conference on Social Robotics* (2012), 298–307.

[22] Leyzberg, D. et al. 2014. Personalizing robot tutors to individuals' learning differences. *Human-robot interaction* (New York, New York, USA, 2014), 423–430.

[23] Liu, P. et al. 2013. It's not polite to point Generating socially-appropriate deictic behaviors towards people. *Human-Robot Interaction* (Mar. 2013), 267–274.

[24] Locher, M. 2011. Situated impoliteness: The interface between relational work and identity construction. *Situated Politeness*. A&C Black.

[25] Max, E.J. et al. 2018. Is silence really golden? Effect of encouragement from a partner or trainer on active video game play. *International Journal of Sport and Exercise Psychology*. 16, 3 (2018), 261–275. DOI:https://doi.org/10.1080/1612197X.2016.1199580.

[26] Mumm, J. and Mutlu, B. 2011. Human-robot proxemics: physical and psychological distancing in human-robot interaction. *Proceedings of the 6th international conference on Human-robot interaction - HRI '11* (New York, New York, USA, 2011), 331.

[27] Park, E. et al. 2011. The Effects of a Robot Instructor's Positive vs. Negative Feedbacks on Attraction and Acceptance towards the Robot in Classroom. *International Conference on Social Robotics* (2011), 135–141.

[28] Perelmutter, R. 2013. Klassika zhanra: The flamewar as a genre in the Russian blogosphere. *Journal of Pragmatics*. 45, 1 (Jan. 2013), 74–89. DOI:https://doi.org/10.1016/j.pragma.2012.10.006.

[29] Porath, C.L. and Erez, A. 2007. Does rudeness really matter? The effects of rudeness on task performance and helpfulness. *Academy of Management Journal*. 50, 5 (2007), 1181–1197. DOI:https://doi.org/10.2307/20159919.

[30] Resnick, B. and Jenkins, L.S. 2000. Testing the Reliability and Validity of the Self-Efficacy for Exercise Scale. *Nursing Research*. 49, 3 (May 2000), 154–159. DOI:https://doi.org/10.1097/00006199-200005000-00007.

[31] Salem, M. et al. 2013. Effects of politeness and interaction context on perception and experience of HRI. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 8239 LNAI, (2013), 531–541. DOI:https://doi.org/10.1007/978-3-319-02675-6_53.

[32] Salem, M. et al. 2013. To Err is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability. *International Journal of Social Robotics* (2013), 313–323.

[33] Schneider, S. and Kummert, F. 2020. Comparing Robot and Human guided Personalization: Adaptive Exercise Robots are Perceived as more Competent and Trustworthy. *International Journal of Social Robotics*. (2020). DOI:https://doi.org/10.1007/s12369-020-00629-w.

[34] Schneider, S. and Kummert, F. 2016. Motivational Effects of Acknowledging Feedback from a Socially Assistive Robot. *International Conference on Social Robotics* (2016), 870–879.

[35] Schunk, D.H. 1982. Progress Self-Monitoring. *The Journal of Experimental Education*. 51, 2 (Dec. 1982), 89–93. DOI:https://doi.org/10.1080/00220973.1982.11011845.

[36] Seo, S.H. et al. 2018. Investigating People's Rapport Building and Hindering Behaviors When Working with a Collaborative Robot. *International Journal of Social Robotics*. 10, 1 (2018), 147–161. DOI:https://doi.org/10.1007/s12369-017-0441-8.

[37] Short, E. et al. 2010. No fair!! An interaction with a cheating robot. *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. (2010), 219–226. DOI:https://doi.org/10.1109/HRI.2010.5453193.

[38] Staw, B.M. et al. 2019. Leadership in the Locker Room: How the Intensity of Leaders' Unpleasant Affective Displays Shapes Team Performance. *Journal of Applied Psychology*. (2019). DOI:https://doi.org/10.1037/apl0000418.

[39] Strait, M. et al. 2014. Investigating the effects of robot communication strategies in advice-giving situations based on robot appearance, interaction modality and distance. *ACM/IEEE International Conference on Human-Robot Interaction*. (2014), 479–486. DOI:https://doi.org/10.1145/2559636.2559670.

[40] Swift-Spong, K. et al. 2015. Effects of comparative feedback from a Socially Assistive Robot on self-efficacy in post-stroke rehabilitation. *IEEE International Conference on Rehabilitation Robotics*. 2015-Sept, (2015), 764–769. DOI:https://doi.org/10.1109/ICORR.2015.7281294.

[41] Weinberg, R. et al. 1979. Expectations and Performance: An Empirical Test of Bandura's Self-efficacy Theory. *Journal of Sport Psychology*. 1, 4 (Dec. 1979), 320–331. DOI:https://doi.org/10.1123/jsp.1.4.320.

[42] Woolum, A. et al. 2017. Rude color glasses: The contaminating effects of witnessed morning rudeness on perceptions and behaviors throughout the workday. *Journal of Applied Psychology*. 102, 12 (2017), 1658–1672. DOI:https://doi.org/10.1037/apl0000247.

[43] You, S. et al. 2011. When the robot criticizes you...: self-serving bias in human-robot interaction. *Human-robot interaction* (New York, New York, USA, 2011), 295.