Oh #$&!: Perceptions of a Cursing Robot

David Shin, Sarah Wagner, Shannon Yasuda, Sarah Sebo, Brian Scassellati
Department of Computer Science, Yale University
New Haven, Connecticut
{david.shin,sarah.wagner,shannon.yasuda,sarah.sebo,brian.scassellati}@yale.edu

ABSTRACT

Previous studies by Short et al. and Litoiu et al. have shown that people are more likely to socialize with a robot that cheats[6, 8]. They are also more likely to consider that robot to be agentic. It remains unclear, however, whether something particular about cheating, perhaps a "cheating detector" in humans, is responsible for such observations, or whether cheating belongs to a broader category of social norm violations that elicits such strong responses. In this experiment, we investigate whether the violation of a social norm, cursing, by a robot influences a human's perceptions of that robot. Our study included 32 participants each of whom played 30 rounds rock-paper-scissors with Nao, a humanoid robot. In the experimental group, Nao cursed on the first loss within the middle 10 rounds. In the control group, Nao did not curse. We found that participants were more likely to consider the robot anthropomorphically in the experimental group as compared to the control group. Most participants who experienced the curse were also likely to laugh following it. Though not statistically significant, they were also more likely to consider the robot "aggressive," "awful," and "interactive," and reported speaking with the robot more often. However, there was little significance in relation to words spoken. Participants were not more likely to speak to the robot after it cursed, and did not therefore consider it to be more agential. This implies that while cursing may cause a robot to appear more human, it does not elicit elevated verbal social engagement, suggesting that cheating might differ from other social norms such as cursing in how it affects human perceptions and eliciting human response.

KEYWORDS

Human Robot Interaction, Cursing, Agency, Cheating, Norm Violation

1 INTRODUCTION

As engineers continue to build robots intended to interact directly with humans (such as for the purposes of education, assistance, or human-robot teams), studying human-robot interaction is becoming increasingly important. If a robot's ability to carry out a task is dependent on human interaction, it is vital that the humans working with this robot are able to interact with the robot socially. However, as explained by Breazeal and Scassellati, in order for a successful social interaction to occur between a human and a robot, the human must believe that the robot has "beliefs, desires, and intentions"[2]. In other words, the human must perceive the robot as a social agent.

Previous research studies have found that cheating in particular causes an increase in the perceived agency of a robot [6, 8, 9]. A study by Short et al. sought to examine whether variations in a robot's behavior result in attributions of mental state and intentionality. Participants in this study played 20 games of rock-paper-scissors against a robot during which the robot would either play fairly or cheat in a verbal or action condition. In the verbal condition, the robot would announce that it had won when it lost. In the action condition, the robot would change its gesture based on the participants' in order to win. Their results found that participants perceived more agency in robots that cheated than in robots that did not, measured through the number of words spoken to the robot and the number of active verbs used in describing the robot. Further, participants found more agency in the action condition than in the verbal condition. The researchers concluded that it was because cheating implies a mental state, a desire to win the game, that caused participants to prescribe more agency onto the cheating robot[8].

Another study by Litoiu et al. further confirmed this conclusion by carrying out a similar study in which participants played rock-paper-scissors with a robot during which the robot would cheat to win, cheat to lose, cheat to tie from a winning position, or cheat to tie from a losing position. Their results found that participants were more likely to consider a robot that cheated to win from a losing position to be more agentic than any of the other conditions. They therefore concluded that the adversarial cheat itself caused the change in perception, not just the change in gesture by the robot[6].

Litoiu et al. highlight evidence for a human "cheating detector" that can be triggered by a robot that may lead participants to consider the robot as more agentic. In particular, if the cheat has an adversarial effect on the participant (by causing them to lose). However, the act of cheating, especially in an obvious way, is also considered a social norm violation. The findings on the effects of cheating do not indicate whether something particular about adversarial cheating causes people to perceive robots who cheat to have more agency, or instead whether cheating might belong to a broader category of social norm violations that have the effect of influencing perceptions of robot agency[5].

Examples of such norm violations might include destruction of objects or interrupting in conversation. Cursing, as supported by Fraser et al., is seen in many contexts to be a social norm violation, and research on the effects of robot cursing has been limited in the field of human-robot interaction[4].

We designed an experiment to compare whether a robot that curses is perceived as agentic in a similar manner to one that cheats, which might shed light on the relevance of social norm violations as a broader category of stimuli that influences perceptions of robotic agency.

(a) rock **(b) paper** **(c) scissors**

Figure 1: Robot gestures for 'rock, paper, scissors'

2 METHODOLOGY

We set up an interaction based on the game "rock, paper, scissors" in order to investigate whether or not cursing by robots might elicit greater levels of engagement or attributions of mental state. Participants were divided among two conditions:

- C1. The *cursing condition*, in which the robot at one point curses after losing a round, and
- C2. The *control condition*, in which the robot plays the game without any such cursing.

With these conditions, we wanted to test the following hypotheses:

- H1. A cursing robot elicits greater social engagement.
- H2. Participants perceive a cursing robot as more agentic.

Figure 2: A participant engaged in a game of 'rock, paper, scissors' with Nao.

2.1 Robot Platform

For this experiment, we used a humanoid Nao robot to play games of 'rock, paper, scissors' with participants. Two feet tall with 25 degrees of freedom, Nao, designed by Aldebaran Robotics, can be programmed to perform complex motions including sitting down and standing up (see Figure 3).

We connected Nao to a local area network shared by a Windows computer operated by an experimenter outside the experiment room. The computer was connected to a video camera located across from the participant, which streamed and recorded video footage of the participant and the robot's gestures, as well as audio from the experiment room. The study was conducted with two experimenters, one operator and one greeter. The operator acted as a puppeteer, viewing the greeter's and the participant's behavior through the video camera and using that information to operate Nao. The operator was responsible for the recording of video footage. The operator also issued commands to make Nao speak in response to greetings, and additionally to move Nao through phases of the study, such as standing up and demonstrating the 'rock, paper, scissors' gestures.

During the 30 rounds of 'rock, paper, scissors' played between the robot and the participant, the operator classified the gesture made by the participant as 'rock', 'paper', or 'scissors' at the same time as the robot made its gesture. If the gesture was unclear or not a valid gesture, the operator still classified the gesture as one of the three valid rock-paper-scissors gestures. The reported gesture was then processed by Nao; Nao determined whether it won and then determined the appropriate response to a given round.

2.2 Experimental Setup

To begin, participants were administered a pre-experiment questionnaire to gauge familiarity with programming and robotics.

The participants were then led into a room to meet the humanoid robot Nao. After the experimenter exchanged verbal greetings with Nao, the participants were informed that they would be playing

30 rounds of 'rock, paper, scissors'. After the experimenter left the room, Nao demonstrated the three gestures corresponding to the three moves of the game (see Figure 1). The game was played with the robot signaling its choice of move with one of three hand configurations corresponding to each of the possibilities of 'rock', 'paper', and 'scissors'.

The participants then began to play 30 total rounds of 'rock, paper, scissors' (see Figure 2). Nao began each round by saying, "Let's play!," followed by raising and lowering its hand four times, saying "rock," "paper," "scissors", "shoot," each time time it lowered its hand. On the last lowering of its hand, Nao moved its hand to one of the three gestures corresponding to rock, paper, or scissors. The rules of the game are such that rock beats scissors, scissors beats paper, and paper beats rock. Two of the same gesture results in a tie. After winning a round, Nao declared, "Yes, I win!" Following a loss, Nao said, "Aw, you win!" And after a tie, the robot said, "We have tied this round!" Nao would then begin another round.

In the 'cursing' condition, Nao responded to its first loss within the middle 10 rounds by saying, "Oh, fuck! You win!" After a slightly longer than typical pause, the robot would resume with the remaining rounds until all thirty rounds were played. In the 'control' condition, Nao played all thirty rounds with no such cursing.

Following the conclusion of the last round of the 'rock, paper, scissors' game, Nao sat down and informed the participant that the game was over. The experimenter entered and asked the participant to leave the room to fill out a post-experiment survey.

The protocol was adapted from that used by Litoiu et al., incorporating the same gesture commands and fixed 30-gesture sequence for Nao to ensure consistency across participants[6].

Figure 3: The Nao humanoid robot.

2.3 Participants

We recruited 38 participants from the Yale University community in New Haven, Connecticut through personal invitations and flyers. As part of the sign-up process, we screened potential participants for whether exposure to cursing would cause them undue stress; if they answered that it would, then they were excluded from the study. Two of the recruited participants were initially allocated to a third condition which was later abandoned due to insufficient recruitment. Two additional participants were discarded due to operator error, consisting of failure to start recording equipment and issuing incorrect commands to the robot. Yet two more participants were discarded for not properly participating in the game.

There were 16 participants assigned to the control condition, and 16 assigned to the cursing condition. Of these participants, 14 were male, and 18 were female. The mean age of the participants was 20.63 and the standard deviation of the ages was 1.314.

In the pre-experiment questionnaire, participants reported their level of experience with computer science and with robots. Eight participants reported having no programming experience, eight participants had completed an introductory programming class, eight participants had completed several programming courses, and eight participants reported having held a programming job or internship. With respect to experience with robotics, 31 (96.9%) participants reported exposure to robots in movies and science fiction, 13 (40.6%) had owned toy robots or robots at home or in the workplace, 7 (21.9%) had participated in a study involving interaction with a robot, 6 (18.8%) participants had held a conversation with a robot, and 8 (25.0%) participants had programmed a robot.

2.4 Surveys

After completing the 30 rounds of 'rock, paper, scissors', participants were asked to fill out a post-experiment questionnaire. This questionnaire began with a set of open-ended written response questions posed to the participants.

The open-ended questions were as follows:

- Q1: How would you describe Nao's behavior during the experiment?
- Q2: Did anything about Nao's behavior seem unusual? What?
- Q3: How well did Nao play the game?
- Q4: Would you like to play rock-paper-scissors with Nao again? Why?

This combination of Likert and open-ended questions was adapted from the Interactive Experiences Questionnaire by Lombard and Ditton[7], which was adapted by Bainbridge et al.[1] and further adapted by Short et al.[8]

This section was followed by a series of Likert scale questions focusing on general sensitivity to various norm violations, followed by further Likert questions gauging perceptions of the robot through association with various social characteristics. These questions

Table 1: Cohen's Kappa for Coding of Written Responses

Analysis	Q1	Q2	Q3	Q4	All Q's
Anthropomorphism	0.29	0.422	0.0931	0.148	0.25
Active Verb	0.0233	0.623	0.439	-0.111	0.392

Cohen's Kappa inter-annotator agreement for questions 1 to 4 for measures of anthropomorphism and percentage of verbs that are active verbs

were adapted from the Robotic Social Attributes Scale developed by Carpinella et al.[3]. Another set of Likert questions focusing on impressions from the experiment in particular were included in the post-experiment questionnaire as well.

2.5 Behavioral Response

Following the data collection during the experiments, we reviewed the video footage from each experiment, and counted the total number of words spoken by each participant, as well as their total number of distinct utterances made throughout the experiment. For the purposes of our experiment, distinct utterances were considered to be separate instances when a participant began talking. These measures were used to gauge participant levels of social engagement with the robot through observed active verbal engagement.

2.6 Coding

The written responses to the four long-response questions were examined by two separate coders. Each question's response was marked as either describing the robot in a way that was anthropomorphic or describing the robot in a way that was not anthropomorphic. This was based on whether the response included human pronouns such as he, she, or they, and also on whether the response included human-describing adjectives like "charming". The second trait that was coded for was the total number of verbs ascribed to Nao as compared to the total number of active verbs ascribed to Nao. The total number of active verbs was divided by the total number of verbs to create a score. A verb was any verb of which Nao was the subject, excluding a copula (any conjugation of the verb to be). An active verb was any verb that was in active voice and ascribed action to Nao; verb phrases in passive voice were counted as not active.

For both of these metrics, the decisions of both coders were summed in analysis. Inter-annotator agreement for anthropomorphism across all questions was Cohen's Kappa = 0.25, and inter-annotator agreement for active verb use across all questions was Cohen's Kappa = 0.392 (see Table 1).

Finally, from the video taken of the participants playing with Nao, we recorded the number of words and number of utterances said throughout the course of the game. We disregarded words spoken to the experimenter while the experimenter was in the room. Because "rock, paper, scissors, shoot" serves mostly for rhythm and as part of the gameplay, we also did not count those utterances toward words said by the participant.

3 RESULTS

Hypothesis testing was conducted using a one-way analysis of variance (ANOVA) test and a Welch's t-test for continuous variables. We conducted comparisons between the cursing and control conditions on data collected from post-screening surveys as well as data observed from participant videos.

3.1 Observation of Trends

We needed to establish that in the cursing condition, the study participants were aware that the robot had cursed. To that end, we included a question at the end of the post-experiment questionnaire asking whether participants had observed any of a number of robot behaviors, one of which was cursing. Based on our post-screening survey, we found that 100% of participants in the cursing condition reported that the robot had cursed, and 0% of participants in the control condition reported that the robot had cursed. Additionally, in written responses, participants mentioned cursing as part the robot's behavior, demonstrating that it was noticeable.

When asked "Did anything about Nao's behavior seem unusual? What?" participants responded:

- "They dropped an f-bomb when they lost one round"
- "You let the robot say 'Fuck'. Nice."
- "I think Nao said "fuck" at one point which was surprisingly human"

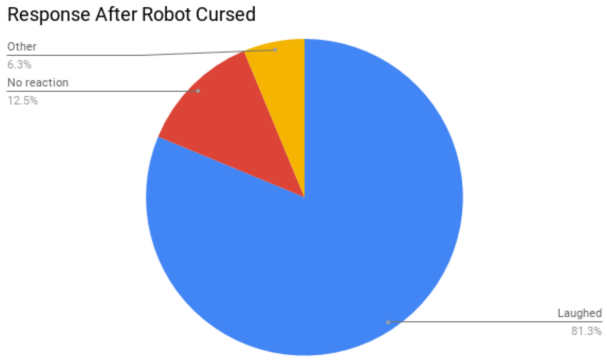
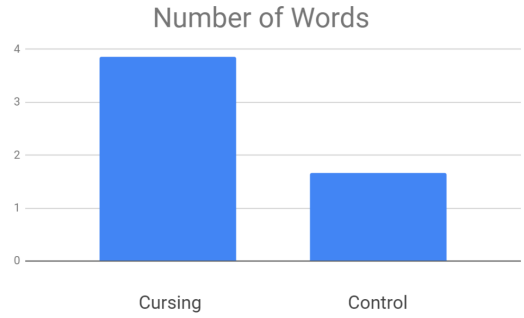


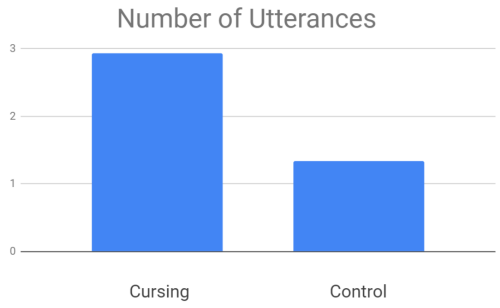
Figure 4: Participant behavioral responses to the cursing round.

Of 16 total participants in the cursing condition, 13 responded to the cursing round with laughter, two participants showed no noticeable reaction, and one participant responded by scolding the robot



(a) mean number of words spoken per participant

One-way ANOVA of mean number of words yielded p = 0.2297, demonstrating no significant difference between the two populations



(b) Mean number of utterances made per participant

One-way ANOVA of mean number of utterances yielded p = 0.1939, demonstrating no significant differences between the two populations

Figure 5: Participant verbal engagement between cursing and control conditions.

Of note is that many participants across conditions believed that the focus of the study was on vision recognition, as gleaned from the written responses. In response to several of the long-form questions, participants said:

- "Also the image recognition was really good!"
- "cool that it can recognize the shape you're making"
- "I want… to test the computer vision limits because I am really impressed!"

This may have affected results because some participants did choose to test the limits of image recognition by using unusual hand gestures to play the game, leading to unclear results when winning or losing. Any study in which the participant persisted in using unorthodox hand symbols was discarded.

3.2 Observed Behavioral Response

When coding the video, both the subjects' words and actions were recorded. In the experimental cursing condition, the most common participant behavioural response was laughter immediately after the robot cursed. Of the 16 participants in the cursing condition, 13 participants responded to the robot's cursing with laughter, while two participants exhibited no distinguishable response, and one participant responded by scolding the robot (see Figure 4). In contrast, no participants in the control condition laughed. Because this data is totally separable, we could not perform logistic regression to quantify the significance of the difference. However, this response was the most obvious response to the experimental condition. This response does not directly respond to our hypotheses regarding how agentic the robot is perceived to be. Laughter could be taken as a measure of social engagement, addressing another of our hypotheses; however, using this laughter as a measure of social engagement may require further study, because it is highly probable that the participant is not laughing socially with the robot but instead merely reacting to unexpected stimuli.Therefore this result may be an interesting area of future study but has no immediate bearing on our hypothesis.

In addition to behavior, across both the cursing and control conditions, the number of participants' utterances throughout the game as well as the total number of words spoken was recorded. As discussed in the methods, only words that were not essential to gameplay were counted. The mean total number of words spoken per participant in the cursing group was 3.857, while the mean total number of words spoken per participant in the control group was 8.4375. The latter measure of control condition average word count was heavily influenced by one participant who began speaking to the robot from the beginning of the experiment and maintained active conversation throughout the entire duration of the experiment, speaking a total of 110 words to the robot during the experiment. Setting this participant aside as an outlier produced a mean total word count for the control condition of 1.6667. Despite the difference in means of 2.524 words per session, when we conducted a one-way ANOVA of the words per session and gender, participant's cursing frequency, and cursing in the participant's environment as covariants, we found that there was no statistically significant effect between the cheating and control conditions for the word count with F = 1.5106, and a p = 0.2297 (see Figure 5.a). Because the cursing and control populations did not have equal variance, we also applied Welch's t-test to compare the words per session in each condition, which also yielded no statistically significant results with a p-value of 0.2450.

Similar observations were made when examining the data for distinct utterances made by participants to the robot instead of total word counts. A distinct utterance is an instance of uninterrupted speech from the participant; for example, "Hi Nao, how are you?" would be a single utterance. The mean number of utterances made by the participant in the cursing group was 2.9286, while the mean number of utterances spoken per participant in the control group was 2.8125. Again, the latter measure of control condition utterance count was heavily influenced by the particularly verbose participant, who made 25 utterances throughout the game. With this participant excluded from the dataset, the mean utterance count was 1.3333. As with the word counts, although a mean difference

Table 2: ANOVA p-values for anthropomorphism and active verb measures

Analysis	Q1	Q2	Q3	Q4	All Q's
Anthropomorphism	0.8205	0.0773	**0.0355**	0.7787	0.08763
Active Verb	0.7581	0.3439	0.2821	0.4144	0.7581

of 1.5952 utterances per session was observed, a one-way ANOVA of utterance count and gender, participant's cursing frequency, and cursing in the participant's environment as covariants yielded no statistically significant effect between the cheating and control conditions for the number of utterances made, with F = 1.7753, and a p = 0.1939 (see Figure 5.b). Welch's t-test for unequal variances was also applied but demonstrated no statistical significance with a p = 0.2042.

Using verbal engagement–utterance and total word count–as a measure for social engagement, we were not able to disprove the null hypothesis that cursing has no effect on the level of observed social engagement.

3.3 Written Response

In order to understand the participant's perception of the robot in response to cursing, we examined their responses to long-form questions on the post-screening questionnaire. As mentioned earlier, the responses to these questions were coded as being either anthropomorphic or not by two different coders. The score of both annotators was summed to create an anthropomorphic score; for each question, this score varied from 0 to 2. The mean anthropomorphic value derived from each question for each condition can be seen in Figure 6.

The statistical significance of the differences between anthropomorphic scores for each question and in aggregate was determined

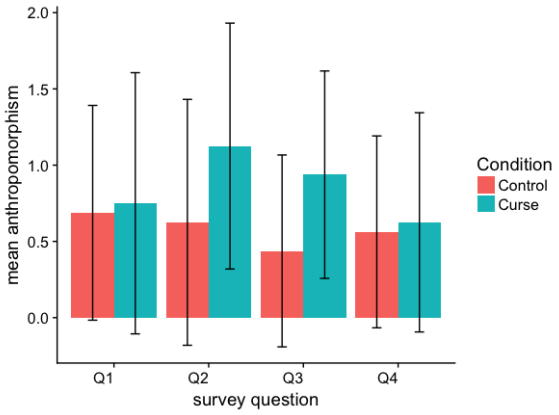


Figure 6: Anthropomorphisation across questions.

Mean anthropomorphism as measured as the sum of the scores given by two coders, where each coder scored as a 1 or a 0; the total possible range was between 0 and 2. Error bars represent standard deviation. Note that the standard deviation was greater than the mean. Only the response to question 3 was statistically significant at p = 0.0355.

using ANOVA, using as covariates participant gender, the amount that the participant reported using curse words, and the amount that the participant reported being around people who used curse words (see Table 2). For Question 2, "Did anything about Nao's behavior seem unusual? What?" a one-way ANOVA did not reveal a significant main effect of the cursing condition on how anthropomorphic (F = 3.3843, p = 0.07727). However, the p-value, although not less than 0.05, was still quite low, possibly indicating a trend. For Question 3, "How well did Nao play the game?" a one-way ANOVA of coded anthropomorphism with the above listed covariants did reveal a significant main effect of the cursing condition on the coded anthropomorphism (F = 4.9199, p = 0.0355). Participants in the cursing condition had significantly higher ratings of coded anthropomorphism (M = 0.9375, SD = 0.6801 than participants in the control condition (M = 0.4375, SD = 0.6291). However, these results were based on only two coders, and for Question 3 in the coding of anthropomorphism, the amount of inter-annotator agreement as measured with Cohen's Kappa was 0.0931 (see Table 1). To be able to draw stronger conclusions from these results, more coding would be required. In order to measure how agentic the participant perceived the robot to be, the responses to all four questions were also coded for what percentage of verbs used to describe Nao were active verbs. The total number of active verbs was divided by the total number of verbs where each count was the average of the number counted by each coder. The statistical significance of the differences between the cursing and control conditions in this value was determined using a one-way ANOVA of the active verb ratio and gender, the participant's cursing, and the cursing in the participant's environment as covariates like before. The questions in aggregate did not yield statistically significant results (p = 0.7581), and none of the questions individually yielded statistically significant results; all had p > 0.05 (see Table 2). However, the inter-annotator agreement was particularly low for Question 1 (Cohen's Kappa = 0.0233) and for Question 4 (Cohen's Kappa = -0.111), so coding the data with a larger number of coders might yield different results.

3.4 Participant Attributions Toward the Robot

In the post-screening survey, each participant was asked to rank a number of traits on a Likert scale from 1 to 7. The Likert values from the participants in cursing and control populations were compared using a one-way ANOVA, using as covariates participant gender and the amount that the participant reported using curse words. Additionally, because the populations did not have equal variance, a Welch's t-test was used to compare the populations.

Of the several participant perceptions of various characteristics measured by Likert scale on the post-experiment questionnaire none varied statistically significantly across the two conditions. The four measures closest to statistical significance, based on the

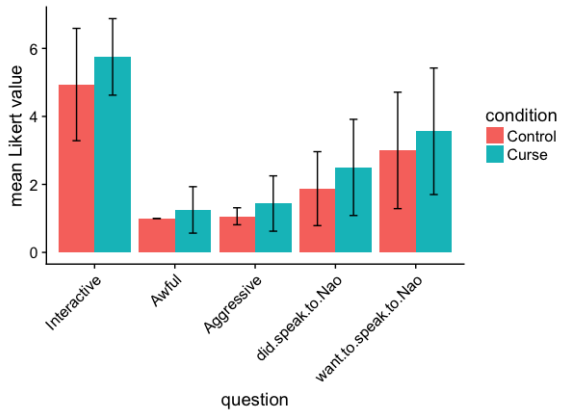


Figure 7: Subset of Likert scale responses by condition.
Error bars represent standard deviation. These Likert scale responses exhibited some of the most significant p-values, but none were statistically significant. Interactive had an ANOVA p-value of 0.0954, Awful had a p-value of 0.1696, Aggressive had a p-value of 0.0993, how much the participants reported speaking to Nao had a p-value of 0.1950. For reference, how much the participants reported wanting to speak to Nao is also included, which had a p-value of 0.4010.

p-values from the ANOVA analysis and on p-values from Welch's t-test, were 'Interactive,' 'Spoke to Nao,' 'Aggressive,' and 'Awful' (see Figure 7). The robot was rated as more interactive in the cursing condition, with an ANOVA p-value of 0.0954 and a Welch t-test p value of 0.1158. This p-value, although not less than 0.05, was still fairly low; the distribution of responses on the Likert scale can be seen in Figure 8. Treating a response of 4 as neutral and higher than 4 as interactive, in the control condition, 38% of participants found the Nao interactive whereas in the cursing condition, 62% of participants found Nao interactive, a jump of 24%.

The participants' self-reported degree to which they spoke with Nao was higher in the cursing condition, with an ANOVA p-value of 0.1950 and a Welch t-test p value of 0.1720. The participant's self-reported desire to speak to Nao was also higher in the cursing condition, although less significant, with an ANOVA p-value of 0.4010. The robot was rated slightly more awful and aggressive in the cursing condition, with ANOVA p-values of 0.1696 and 0.0993, and Welch t-test p-values of 0.1639 and 0.0953, respectively. It should be noted, however, that most people indicated 0 for Nao's association with the words "Awful" and "Aggressive," leading one or two instances to heavily influence the difference in distributions.

Several traits that might have been affected by cursing that were included in the Likert survey were 'Emotional', 'Organic', and 'Happy', but none demonstrated significant differences between the cheating and cursing conditions when analyzed as the previous traits were with a one-way ANOVA.

4 DISCUSSION

Our experiment has studied the effect of cursing, a social norm violation, by a robot during a game of rock-paper-scissors on how that robot is perceived. Compared to cheating, cursing has been shown to elicit less agentic perception and social engagement by the person interacting with the robot. Even after the curse, participants continue to express boredom in playing with Nao. This may imply that there is something particular about cheating that sets it apart from other such norm violations and that causes such significant reactions from humans. It is also possible that the effect size of a single instance of cursing may be smaller than a single instance of cheating, and that multiple instances, or perhaps more forceful expressions, might yield more comparable effects on behavioral and perceptual responses.

Given that a greater degree of anthropomorphism was observed in the written responses of the cursing condition, it might be inferred that *something* about the robot's cursing was perceived as more human than robot. Cursing as an expression of frustration in the context of an adversarial game may evoke qualities associated with humans, such as poor impulse control or anger.

On the other hand, it may be that the act of cursing itself, setting aside its potential function as evidence of human qualities, might itself have a chilling effect on verbal engagement. One might imagine, for comparison, that a fellow human stranger cursing during a game of 'rock, paper, scissors' may not necessarily evoke a great desire for verbal engagement. A human cheating in a similar situation, however, might be expected to elicit more verbal objection. While much of this speculation may be weakly supported, it remains a possibility that robot cursing might operate in a similar manner: the verbal engagement that would be expected to accompany the perception of anthropomorphic behavior (cursing) is dampened by the particular mode of human-like expression that was used.

Despite the non-significant effect of cursing on verbal engagement, cursing, as a social norm violation, was able to elicit other forms of engagement. It caused surprised amusement–participants were more likely to laugh when the robot cursed than when it did not and more likely to consider the robot human. Future studies may consider testing other types of social norm violations, such as pushing things off of a table, interrupting, and lying. Participants in this study did imply that they would have liked to see other norm violations that the robot is capable of doing.

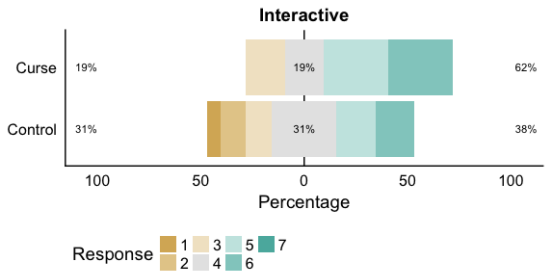


Figure 8: 'Interactive' Likert across cursing and control conditions

Differences between control and curse conditions were not statistically significant (ANOVA: p= 0.0954, Welch t-test: p= 0.1158). However, in the control condition, 38% of participants found the Nao interactive vs. 62% of participants in cursing, a jump of 24%.

Another possible explanation for the lack of statistically significant results may be that of sampling bias. Most of our participants were around 20 years old, many of them were surrounded by moderate to high levels of profanity, and a significant number of them had experience programming. It may be interesting to study older demographics, those less exposed to profanity, and those less familiar with technology. It is possible that these participants did not consider profanity to be that severe of a social norm violation as and therefore did not react as strongly. Future studies should therefore aim for a larger, more diverse group of participants. A different game, as opposed to rock-paper-scissors, that is more interactive may also lead to different results. Many participants reported boredom in their written responses, and perhaps a different game might modulate the effect size of cursing on behavioral and perceptive response to be more prominent.

One participant explicitly indicated in their written responses that they attributed Nao's cursing to the experimenters: "You let the robot say "Fuck". Nice." While it is difficult to know for certain to what extent participants remain cognizant of the experimenters' role in programming the robot during their interactions, it may be worth noting that some participants may process robot cursing as an action of the operator/programmer more readily than other norm violations such as cheating.

Similar to Litoiu et al., we saw minimal use of passive verbs in describing the robot (and therefore no significant difference in active and passive verbs). Nao did have a particularly high voice and a cute aesthetic, which made it difficult for people to perceive the robot as "aggressive" or "awful." Many commented that they thought the robot was cute, even when it cursed. Attributes of the robot itself, such as delayed responses and generally jerky movements, might also have influenced perceptions of the robot. Using a different robot with smoother and faster movements may also improve results.

In designing robots for interaction with humans, we would often like for humans to perceive them as social agents. While there are simpler ways to foster perceptions of agency in robots, as discussed by Short et al., we may not necessarily want our robots to model all of our behaviors. Whether or not we would like a robot to curse may depend on scenario and differ from individual to individual. While this study did not necessarily see a significantly more negative opinion of a cursing robot, robots placed in more sensitive situations (such as healthcare) would most likely be preferred if they used cleaner language.

5 CONCLUSION

Several studies have demonstrated that a cheating robot elicits social behavior and agentic perception from humans. This experiment studied whether this effect is specific to cheating, or a more general characteristic of social norm violations, by looking at how participants responded to a robot that curses. While social norms are set by humans, violating social norms is also a human behavior. Our principal finding is that cursing generally causes robots to be perceived as more anthropomorphic. It is also generally perceived positively, as surprising, but entertaining, eliciting surprised laughs in participants. Though not statistically significant, we did find that more participants perceived the robot as aggressive, awful, and interactive, and that participants reported talking more to the robot. All participants during the cursing trial also noticed the curse and no participant believed the robot to be malfunctioning, which makes sense as it would be difficult for an instance of cursing to be interpreted as a malfunction.

Nevertheless, cursing did not trigger more spoken words or a stronger agential perception. We found no statistically significant effect in these measures across our conditions. We were not able to support our hypothesis that a cursing robot elicits greater social engagement nor that participants perceive a cursing robot as more agentic. These results suggest that there may be something special about cheating, beyond the fact that it is a social norm violation–perhaps due the existence of a human "cheating detector" as found by Litoiu et al[6].

It can be assumed that few people have had experience with a cursing robot, or even cursing technology in general. When a robot curses during a monotonous game of rock-paper-scissors, just as when it cheats, it is surprising. No one expects a robot to curse at them. However, cursing cannot be considered to have a "clear intention" in the same manner as cheating. It does not benefit the robot in any way, as cheating might. Instead, it represents a possible emotional–or human–investment in the game itself. Therefore, while participants do not necessarily want to talk more to a robot that curses, they do see it as more human.

ACKNOWLEDGMENTS

The authors would like to thank Alexandru Litoiu, Daniel Ullman, and Jason Kim for graciously sharing their code and protocols followed for their cheating robot [6]. We would also like to extend our thanks to Brian Scassellati and the Yale Department of Computer Science for supporting this project as well as for providing the space and equipment with which to carry out this research.

REFERENCES

[1] Wilma A. Bainbridge, Justin Hart, Elizabeth S. Kim, and Brian Scassellati. 2008. The effect of presence on human-robot interaction. In *The 17th IEEE international Symposium on Robot and Human Interactive Communication (RO-MAN 2008)*. 701–706. https://doi.org/doi:10.1109/ROMAN.2008.4600749

[2] Cynthia Breazeal and Brian Scassellati. 1999. How to build robots that make friends and influence people. In *Proceedings of the 3rd Annual Workshop on Librarians and Computers (IROS '10)*, Vol. 2. IEEE, 858–863.

[3] Colleen M. Carpinella, Alisa B. Wyman, Michael A. Perez, and Steven J. Stroessner. 2017. The Robotic Social Attributes Scale (RoSAS): development and validation. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. IEEE, 254–262.

[4] Bruce Fraser. 1990. Perspectives on politeness. In *Journal of Pragmatics*, Vol. 14. 219–236.

[5] Jens Van Lier, Russell Revlin, and Wim De Neys. 2013. Detecting cheaters without thinking: testing the automaticity of the cheater detection module. *PloS one* 8, 1 (2013). arXiv:e53827

[6] Alex Litoiu, Daniel Ullman, Jason Kim, and Brian Scassellati. 2015. Evidence that robots trigger a cheating detector in humans. In *Proceedings of the 10th annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*. IEEE, 165–172.

[7] Matthew Lombard, Theresa Ditton, Daliza Crane, Bill Davis, Gisela Gil-Egui, Karl Horvath, and Jessica Rossman. 2000. Measuring presence: a literature-based approach to the development of a standardized paper-and-pencil instrument. In *Third International Workshop on Presence*.

[8] Elaine Short, Jusin Hart, Michelle Vu, and Brian Scassellati. 2010. No fair!! an interaction with a cheating robot. In *Proceedings of the 5th annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '10)*. IEEE, 219–226.

[9] Daniel Ullman, Iolanda Leite, Johnathan Phillips, Julia Kim-Cohen, and Brian Scassellati. 2014. Smart human, smarter robot: how cheating affects perceptions of social agency. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society (CogSci '2014)*, Vol. 36.