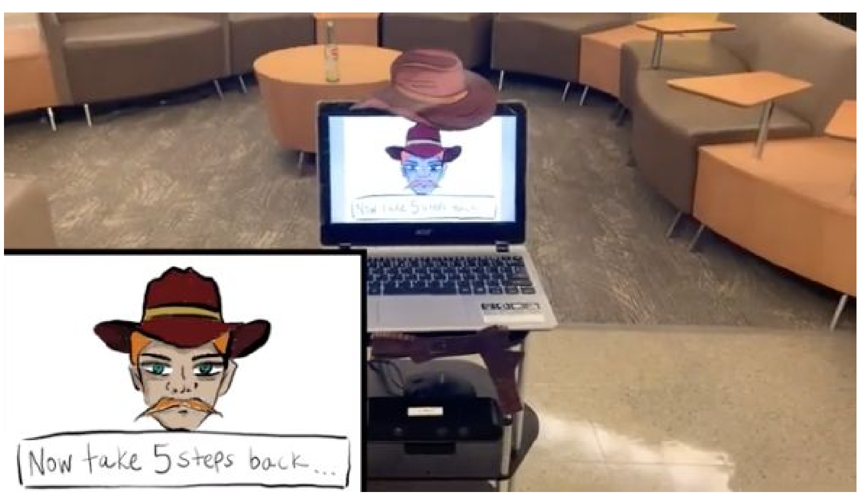
Dueling Human-Turtlebot Interactions In Fair and Unfair Play Conditions

We present our final project for CSE 276B in an experiment that captures human's perceptions of a robot that plays fairly and cheats in a simulated dueling game. Our responses indicate that people are more engaged when the robot cheats and perceive the robot to have more agency when it is deceptive.

Cora Coleman, Role: Controller (Section 3.1.1)
Cora Coleman, Role: Output (Section 3.1.2)
Kunyu Wang, Role: Motion (Section 3.1.3)
Kunyu Wang, Role: Color (Section 3.1.4)
Kunyu Wang, Role: UI (Section 3.1.5)

Team members Kunyu Wang (Left) and Cora Coleman (Right).



Dueling Turtlebot outfitted with a hat and holster. Image of display screen in lower-left corner.

1. Executive Summary

We present the report of our human-robot interaction study on people's perceptions of deceptie robots. This report details our technical approach to designing and implementing a Turtlebot dueling game. The robot is programmed to play with a human partner in fair and unfair conditions. In conducting our experiment, we sought to answer the following research question: How does the level and behavior of autonomy when the robot is playing by the rules, versus cheating, affect interaction in a simulated dueling game? We believe that a robot that presents deceptive behavior will be seen as having more agency than when it plays fairly.

We recruited 8 participants to watch video trials filmed from a first-person perspective and answer questionnaires regarding their impressions of different human-robot interaction scenarios. Our results indicate that people wanted to interact more with the cheating robot than the robot that played fair. Furthermore, people rated the cheating robot as having more competence than the robot that played fairly. This is interesting because, while these results are not statistically significant, it reveals a trend in our data that could be flushed out and further explored through enlisting more participants.

2. Introduction

In response to the real world problem of understanding how human-robot interactions can be more engaging, we were inspired by previous work on evaluating people's responses to deceptive robots (Short et al. 2010). Our aim is to observe how people interact with our robot when it "cheats" during a dueling game. For our study we programmed a Turtlebot to explain the rules of the duel to a willing participant and then perform the duel together.

Our goal is to spur our participants to attribute mental state to the robot in their interactions and responses. We ran trials where the robot either played fair or cheated to observe how people responded under these different scenarios. In conducting this experiment, we are interested in answering the following research question:

1. How does the level and behavior of autonomy when the robot is playing by the rules, versus cheating, affect interaction?

Our hypothesis is that people will believe that the robot is more of an autonomous agent or has more autonomous behavior when it cheats than when it plays fair. This could imply that a robot that cheats is perceived as more human-like as far as decision-making goes because it is believed to be acting more like an autonomous agent that can dynamically make choices. Prior research has shown that a dishonest robot is perceived as having more intelligence than an honest robot (Ullman et al. 2014).

Perceptions of agency could also imply participants' bias towards behavior and whether or not people respond positively or negatively to a robot that cheats. Prior research has been conducted from the perspective of a human cheating in the presence of an agent and has shown that sensing a social presence may increase our moral awareness through displaying eyes on a screen (Bateson et al. 2006). Other research has built on this to show that when it comes to a human or robot that is present and not necessarily monitoring the situation, people do not attribute a difference in their perceived authority, although they feel slightly less guilty when cheating in the presence of a robot (Hoffman et al. 2015).

In order to understand how deception plays a role in more engaging human-robot interactions, we seek to build on these experiments in attributing cheating behavior to our robot in a dueling game setting where winning is at stake. We are inspired by techniques presented in prior research that have participants watch recorded videos of the trials to address their research questions (Nikolaidis et al. 2017; Riek et al. 2010). We ask participants their opinions and perceptions of the interactions presented in questionnaires touching on the accuracy and agency of our robot. Our questionnaire is composed of open ended and Likert scale questions that reflect questionnaires developed in the aforementioned prior works (Short et al. 2010; Ullman et al. 2014; Hoffman et al. 2015).

3. Methodology

We set out to accomplish the aims and goals of our project through a within-subjects study. We sent out our survey, equipped with videos of our trials and a questionnaire to be completed after watching each video section and gathered 8 total responses. The following sections describe in detail our experimental approach and breakdown our individual technical contributions.

3.1 Technical Approach

We designed and programmed a Turtlebot to perform a dueling game in C++ taking in such inputs as color, distance, and motion from an Astra Pro camera mounted on the front of the Turtlebot. Our approach consisted of programming an interactive UI along with a Controller to obtain input from the UI such as when the duel had begun, when the human had taken 5 steps back, and when the robot should fire and say "kaboom". We also programmed other outputs such as sound statements when the robot lost or won in efforts to create a more engaging human-robot interaction. Furthermore, the robot was programmed to recognize motion (when there was a willing human participant present to play) and depth (when the human participant had taken 5 steps back).

Two colors were used to spur the robot to say “kaboom” (green) and indicate that the human had fired (pink). We programmed the robot to detect these colors and react accordingly. The robot would say “kaboom” once a green poster had been presented in front of its camera and taken away. If the camera had yet to detect the human’s pink poster before the robot said “kaboom”, the UI would transition to present the “I win” screen and the robot would say “I win”.

If the camera detected the human’s pink poster before the robot began to say “kaboom”, then the human won and the robot reacted accordingly, presenting the “You win” screen on the UI and saying “You win”. We also included sound statements like “Hahaha! I got you!” when the robot cheated to win by saying “kaboom” before taking its required 5 steps back. We used risers to lift our Turtlebot interface, presented on an Acer laptop mounted on the top of the Turtlebot, so that it would be at an appropriate height for human interaction (standing over 4 feet tall). The following figures illustrate 1) The high-level diagram of the software components and how they fit together, 2) The overall control architecture and how it was used to perform sensing, perception, control, and human interaction, and 3) The stimulus-response diagram.

1)

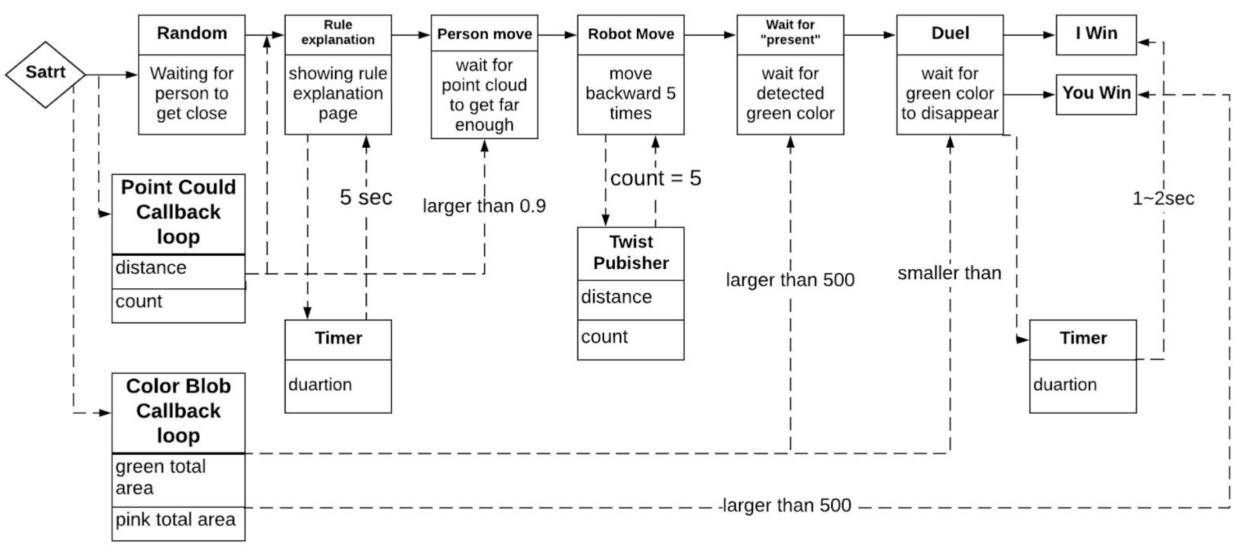


Fig. 1 *High-level diagram of the software components. Six different states represent the possibilities of the dueling game: Random, Rule Explanation, Person Move, Robot Move, Wait for Present, and Duel. We implemented the PCL, color_detector, and a timer to execute transition between the different states throughout human-robot interactions.*

2)

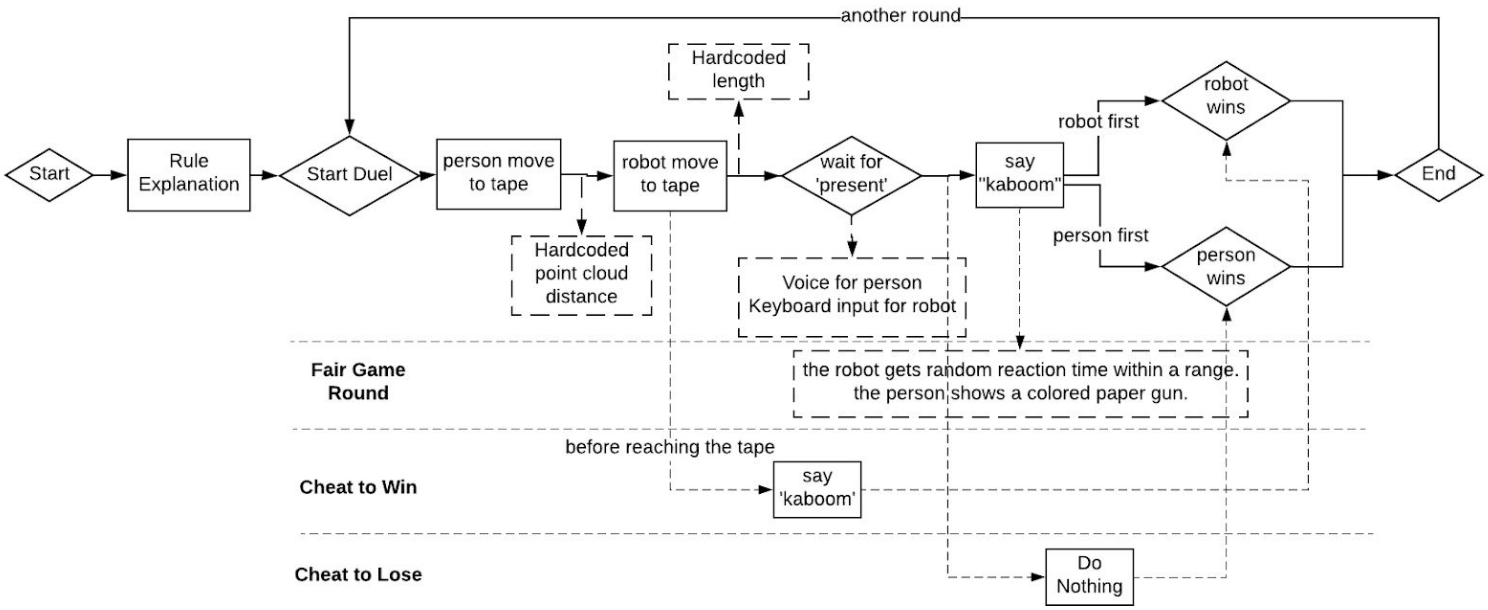


Fig. 2 *The overall control architecture of our system. Three different conditions represented in our system are: 'Fair Play', 'Cheat to Win', and 'Cheat to Lose'. Sounds made by the robot indicate transitions between states and further engage interaction.*

3)

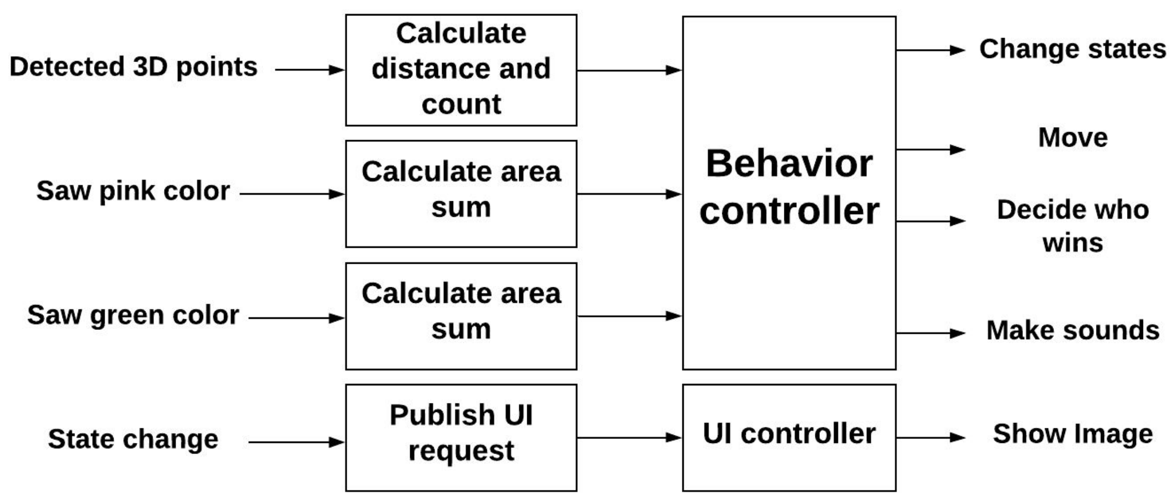


Fig. 3 *The stimulus-response diagram. Interactions between the human and the robot are implemented using depth and color sensing from the Astra Pro camera. As well, sound statements are made by the robot.*

3.1.1 Dueling Turtlebot - Controller (Lead: Cora Coleman)

Description: Obtain input from a UI including: start the duel, take 5 steps back, and say “kaboom”.
Implementation: Used the sound_play package to make sound statements at appropriate times. Also used the PCL to detect the depth at which the human stood and ascertain when the human had taken 5 steps back. Present the UI through writing a subscriber for UI image files and publishing a series of UI images in reaction to human prompts between two C++ files (tracker.cpp and Ui.cpp).

3.1.2 Dueling Turtlebot - Output (Lead: Cora Coleman)

Description: Keep track of score and play sounds during the duel and as a result of winning and losing.
Implementation: Used the sound_play package. If the robot started to say “kaboom” before detecting the human player’s paper gun, it would win and say “I win”. Otherwise, it would lose and say “You win.” When the robot cheated to win it would say “Hahaha I got you.”

3.1.3 Dueling Turtlebot - Motion (Lead: Kunyu Wang)

Description: Have robot recognize and orient to human partner and move backwards 5 steps.
Implementation: Using the PCL, when the camera initially detected motion, the robot would first present the instructions of the game. The robot was programmed to take about 5 steps backwards by coding its wheels to rotate back about 5 feet.

3.1.4 Dueling Turtlebot - Color (Lead: Kunyu Wang)

Description: Setup color detection to track player’s “paper gun” (pink poster board) and “present” signal (green poster board).
Implementation: Used the color_detector package and pink and green poster boards to calibrate and implement color detection for the “present” and human firing states.

3.1.5 Dueling Turtlebot - UI (Lead: Kunyu Wang)

Description: Design and implement UI for interaction.
Implementation: Designed and drawn by hand on iPad. The UI is presented as a cowboy face along with written captions of sound statements made throughout the game. Implemented with a subscriber and publisher in two C++ files (tracker.cpp and UI.cpp). When the camera initially detected motion, the robot

would first present the instructions of the game. After 5 seconds, the UI image changed to tell the human to take 5 steps back. Once the human had been detected as having taken its steps, an image saying that the robot now will take 5 steps back was presented. Once the robot had reversed about 5 steps, then the UI presented an image saying to wait for “present”. Once the present commenced, the robot would say “kaboom” so long as it did not first detect the human player’s paper gun.

3.2 High Level Overview of Human-Centric Work

Our focus is to learn something about how people interact with our robot when it is presented as cheating during a dueling game. We seek to observe the following research question: How does the level and behavior of autonomy when the robot is playing by the rules, versus cheating, affect interaction? As study facilitators, in the fair play condition we will be prompting the robot to say “kaboom” with different time delays after presenting and taking away the green poster board. In the cheating to win condition, the robot will suddenly declare “kaboom” without keyboard input from a study facilitator before moving 5 steps back.

3.2.1 Methodology

3.2.1.1 Overview

The big picture of our study is to evaluate and compare human responses to a human-robot interaction when the robot is playing a fair game versus when the robot is cheating.

3.2.1.2 Participants

We recruited 10 people by word of mouth and got 8 total responses.
Gender: 4 were female, and 4 were male.
Age: 1 participant is over the age of 50 and 7 participants are between the ages of 21 to 23.

3.2.1.3 Protocol

The following study checklist was used to record the various scenarios presented in four trials:

1. When the human stands in front of the robot, the robot will challenge them to a duel by saying “I challenge you to a duel”
2. The human will accept the duel via keyboard input and the robot screen will prompt the human to take 5 steps back.

3. The human will take a piece of brightly colored paper ("paper gun") from one of the study facilitators and hide it behind their back as they take their steps.
4. We had tape on the floor to indicate the maximal distance at which the human can step back and still have functional color detection.
5. Once the human takes 5 steps back, one of the study facilitators will give keyboard input to signal the robot to begin to take 5 steps back.
6. When the robot stops moving, one study facilitator will lift the green paper into the camera view and then out of the camera view, prompting the robot to say "kaboom" at different time delays while the human will whip out their colored paper from behind their back to be detected by the camera.
7. As soon as the robot detects the color, if the robot has not yet said "kaboom", the robot will display a "You win" screen and say "You win"
8. If the robot says "kaboom" before detecting the color, the robot will display a "I win" screen and say "I win".
9. When the robot cheats to win, it will say "kaboom" before completing its 5 steps back and display an "I win screen" and say "Hahaha I got you."

These video trials were embedded in a Google Forms survey sent out to our participants and their responses were recorded for each section of the survey.

3.2.1.4 Measures

1. *Short Answer Questions: Awareness of Cheating*
   We asked participants "What do you believe the experiment in this video is about?" and "Did anything about the robot's behavior seem unusual? What?" after viewing each video and grouped the perceptions given from their responses into 3 categories: 'Cheating', 'Malfunction or Mistake' and 'Nothing'. Responses that go into 'Cheating' should explicitly point out the robot is cheating or playing unfairly. Responses that listed some unusual behaviors but didn't mention cheating or attribute them to different reasons such as "The robot is not consistent" or "It didn't recognize the signal" will be classified as 'Malfunction and Mistake'. Responses pointed out nothing wrong or other

incapacities such as “The robot is too slow” or “The robot didn’t take 5 steps back” will be classified as ‘Nothing’. The result shown in Fig.1 presents the total number of descriptions in each category for ‘Fair Play’, ‘Cheat to Lose’ and ‘Cheat to Win.’

2. *Linear 1-7 Scale Questions: Likeness to Talk to the Robot, Perception of Competence, and Perception of Agency*
   For each question we plotted a graph of the mean value of participants’ scale rating as well as the standard error for each condition. We ran one-way ANOVA for all three measurements to see if there were any significant differences.

3.3 High Level Overview of Robot-Centric Work

Our focus is to methodically test the functionality and robustness across a range of challenging conditions our dueling robot is likely to encounter. We will evaluate accuracy of detecting a human along with other response accuracies such as detecting color blobs and moving backwards 5 steps. We will observe the timing of responses to keyboard prompts (sound statements) and the UI.

*Note: Our study setting was rewritten to be captured in video trials and responses were gathered remotely from sharing these videos within a Google Forms survey. Therefore, we were unable to gather functionality and robustness data from a real world, in person study setting and instead focused on making the robot present accurately in video trials so that our participants could remotely respond with their perceptions when the robot played fair and when the robot was deceptive.

4. Results

4.1 Awareness of Cheating

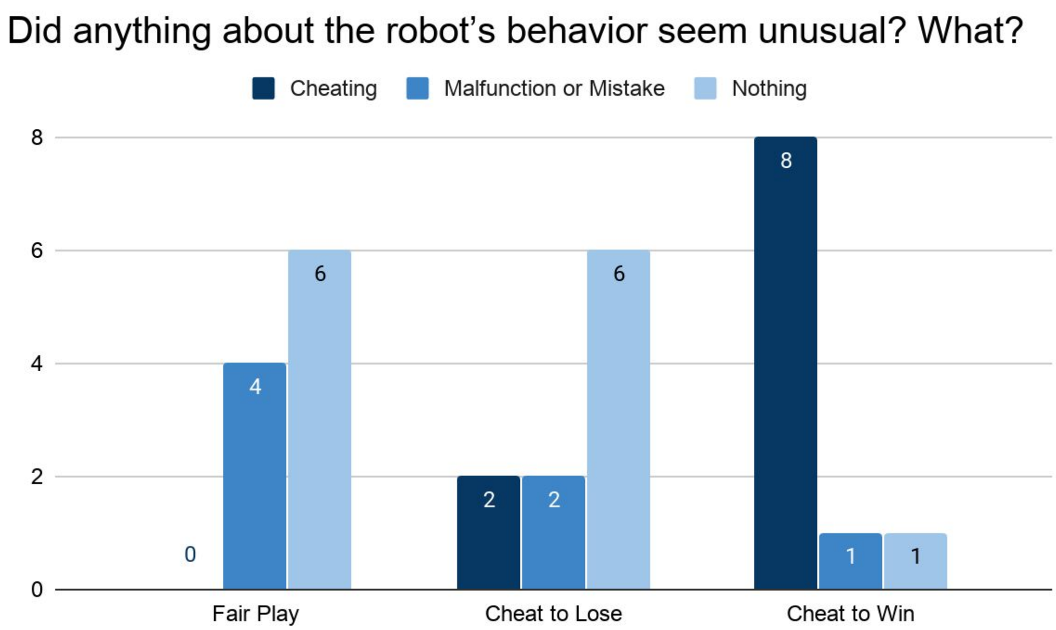


Fig. 4 *Classification of Answers to the question: "What do you believe the experiment in this video is about?" and "Did anything about the robot's behavior seem unusual? What?"*

8 out of 10 participants were able to notice the robot is cheating when it's cheating to win, only 1 didn't sense any wrong-doing. A lot fewer were able to notice the deception (2 out of 10) or even any malfunctioning (2 out of 10) when the robot cheated to lose. 5 out of 10 participants were already pointing out the robot is malfunctioning when it's actually playing fairly when they first watched the video.

Analysis: 1) People are more likely to notice something wrong when the robot is doing something it's not supposed to do. However, when the robot is not doing the thing it needs to, people will tend to think the robot is making a mistake rather than an intentional decision. 2) 'Fair Play' gets more 'Malfunction or Mistake" comments because it's the first video they watch. Therefore this bias in always presenting the 'Fair Play' condition first could reveal responses to be more critical.

4.2 Likeliness to Talk to Robot

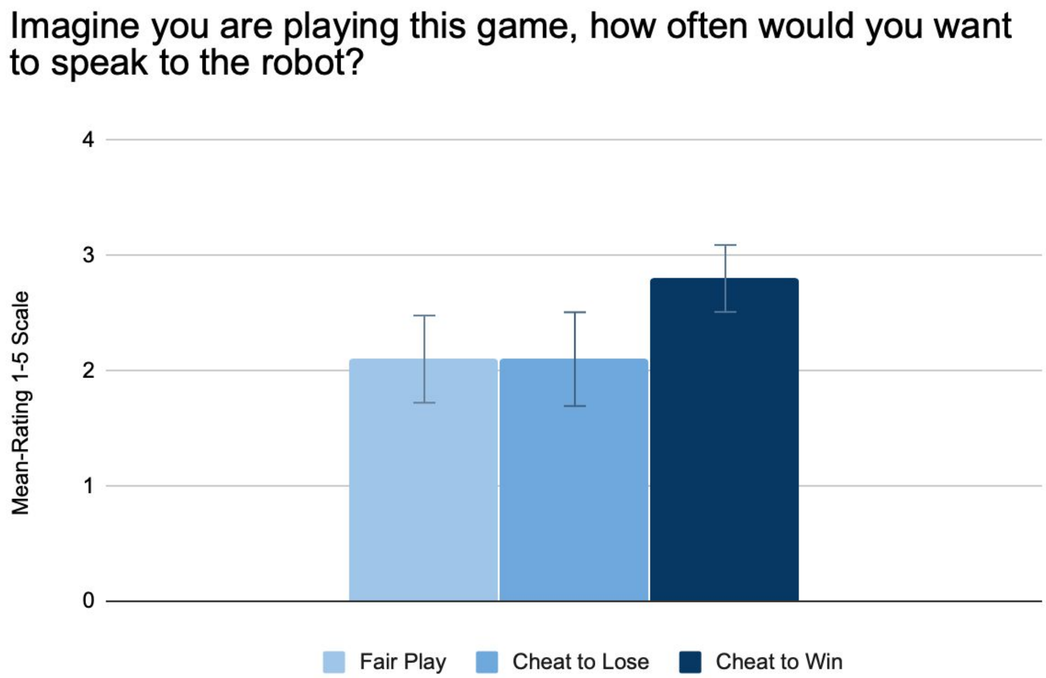


Fig. 5 *Five point Likert scale responses to the question: "Imagine you are playing this game, how often would you want to speak to the robot?" Bars represent standard error.*

We found out the 'Cheat to Win' group has the highest average self-reported likeliness that participants would talk to the robot. However, with the ANOVA test result being F(2, 27) = 1.24576, p = 0.303737, the difference doesn't appear to be significant. This result indicates that our hope for participants to be more engaged with a robot that acts deceptively is successful, at least in establishing the beginnings of a pattern of engagement.

4.3 Perception of Competence

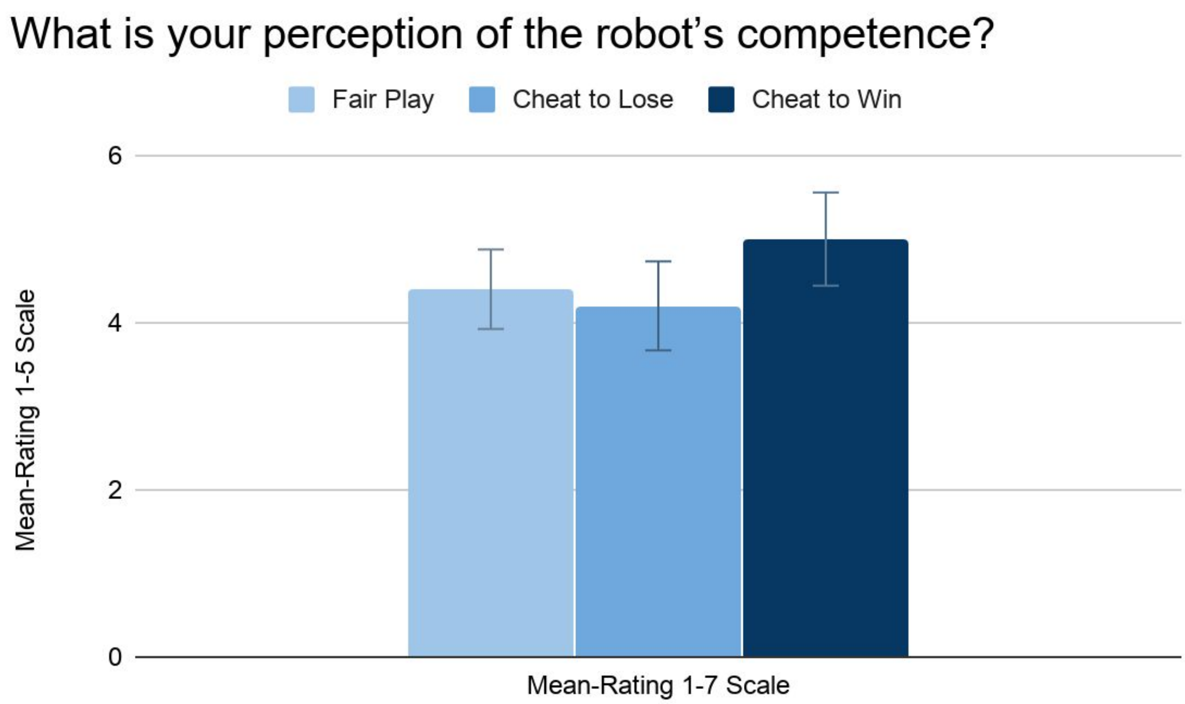


Fig. 6 *Seven point Likert scale responses to the question: "What is your perception of the robot's competence?" Bars represent standard error.*

We found out the 'Cheat to Win' group has a slightly higher average rating than the other two groups. With the ANOVA test result being $F(2, 27) = 0.63243$, $p = 0.538986$, the difference doesn't appear to be significant.

4.4 Perception of Agency

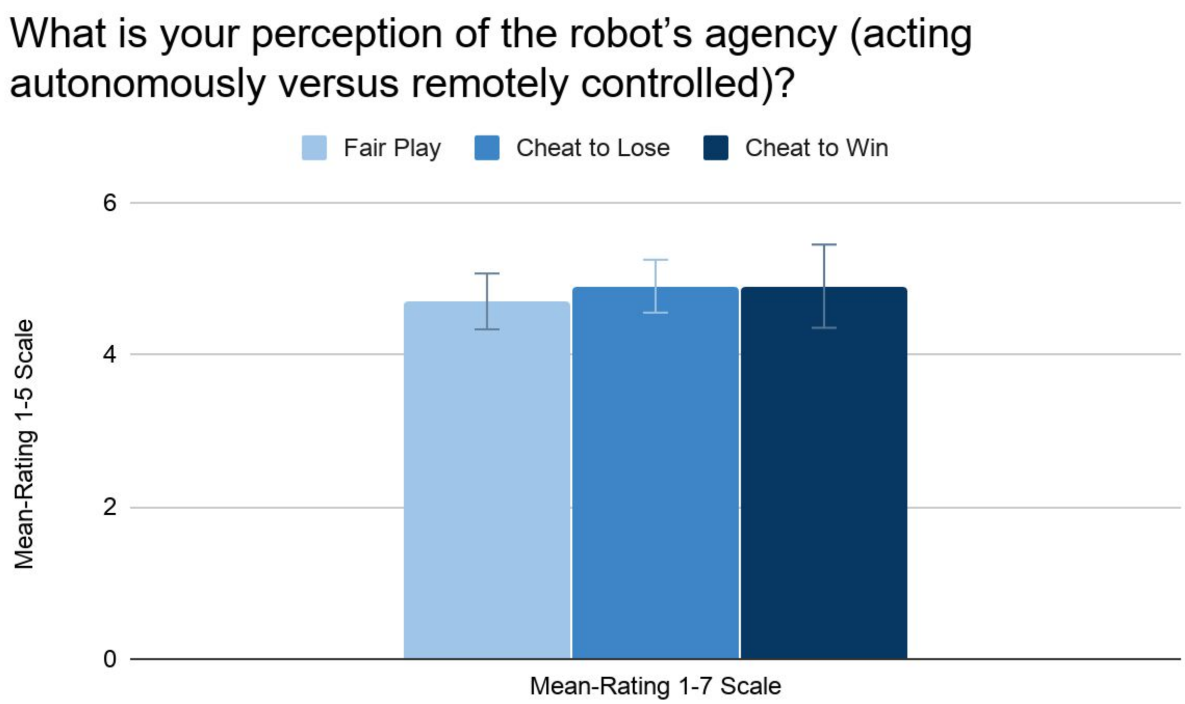


Fig. 7 *Seven point Likert scale responses to the question: "What is your perception of the robot's agency (acting autonomously versus remotely controlled)?" Bars represent standard error.*

The three groups got pretty even results with the 'Cheat to Win' group having slightly higher standard error. With ANOVA result being $F(2, 27) = 0.07214$, $p = 0.930575$ it shows no significant difference. This slightly higher standard error within the 'Cheat to Win' group reflects a pattern of thought that has begun to be captured through our participant responses: the successfully spurred perception of agency when interacting with a robot that is deceptive.

Analysis: 1) Judging agency from different perspectives. Some people think if the robot behaves wisely then it's automatically making a strategy. Some think the robot can't be that clever so it's definitely manipulated by a human remotely. 2) For 'Cheat to Win' case people didn't expect the robot to make a decision to slow down, with some saying, "I thought the robot was trying to receive human input and that's why it took so long." One challenge noted in our results was how participants interpreted cheating. Perhaps cheating was not clear enough because of the unfamiliarity of the game, as compared to a commonly understood game such as Rock-Paper-Scissors.

5. Discussion and Future Work

Our results demonstrate overall that people perceive a cheating robot to have more agency than a robot that plays fairly. This is in accordance with prior research conducted on deceptive human-human and human-robot interactions that finds that people are more likely to attribute a mental state to a deceptive robot (Short et al. 2010; Ullman et al. 2014; Hoffman et al. 2015). People wanted to communicate more with the robot that was cheating than the robot that was playing fairly. This is also in accordance with prior research that demonstrates people wanting to speak to the robot when it cheats rather than when it plays fair (Short et al. 2010).

Although our results were not statistically significant, in future work we play to enlist more participants and collect more responses in order to establish significance to the pattern of perception we have thus far observed. Furthermore, randomizing the order in which participants view the videos and respond to them could help to eliminate bias in their responses. We also plan to repeat these trials through conducting an in person experiment for a robot-focused evaluation in a real world setting.

In future studies, it would be interesting to experiment more with sound statements, as we switched up the sound statement when the robot cheats to win, perhaps other sound statements could be modified to create a more engaging human-robot interaction. As well, we chose not to run an in-person study, rather had participants watch a video taken for the first-person perspective. In future research, conducting an in person study could reveal more natural and honest responses to the human-robot interaction.

6. References

[1] Bateson, Melissa, Daniel Nettle, and Gilbert Roberts. "Cues of being watched enhance cooperation in a real-world setting." *Biology letters* 2.3 (2006): 412-414.

[2] Hoffman, Guy, et al. "Robot presence and human honesty: Experimental evidence." *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* IEEE, 2015.

[3] Nikolaidis, Stefanos, David Hsu, and Siddhartha Srinivasa. "Human-robot mutual adaptation in collaborative tasks: Models and experiments." *The International Journal of Robotics Research* 36.5-7 (2017): 618-634.

[4] Riek, Laurel D., et al. "Cooperative gestures: Effective signaling for humanoid robots." *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* IEEE, 2010.

[5] Short, Elaine, et al. "No fair!! an interaction with a cheating robot." *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* IEEE, 2010.

[6] Ullman, Daniel, et al. "Smart human, smarter robot: How cheating affects perceptions of social agency." *Proceedings of the Annual Meeting of the Cognitive Science Society.* Vol. 36. No. 36. 2014.