Effect of Hesitation on Perception of Intelligence

Colin Dunn, Nicholas Gonzalez, Danielle Zucker, Brian Scassellati
Department of Computer Science
Yale University
New Haven, CT 06511
colin.dunn@yale.edu, nicholas.gonzalez@yale.edu, danielle.zucker@yale.edu, scaz@cs.yale.edu

***Abstract*—We describe the ways in which the presence or absence of thought-like behaviors in a robot opponent affect's a human's perception of the robot. Participants play a simple game, Take Two, against the NAO robot. In each trial, the robot demonstrates one, both, or neither of two actions associated with thought: verbal thinking noises and hesitation. We find that robot hesitation has an effect on human's perception of robot intelligence, and that the robot that acted most naturally was perceived as less distracting and more connected. We find no effect of hesitation or thought noises on participants' enjoyment of the game or the effort they exhibited while playing.**

I. Background and Related Works

As autonomous robots become more prevalent in today's society, it is important to understand the perceptions that humans have of different robotic agents and their actions. In the past, much work has been done to examine attribution of agency and animacy to robotic agents by humans [1]. From these studies, we have learned that robots that engage and interact with their environments are more likely to be perceived as real, intelligent agents. It has also become apparent that the design of a robotic agent also plays a role in how humans perceive its actions.

Following these studies, there has been a large shift of focus in the robotics field to make human–robot interactions more fuid, natural, and comfortable. In fact, there is a large field of robotics dedicated to conducting studies known as Human-Robot Interaction (HRI) studies. Among the many findings that these HRI studies have uncovered is the fact that people perceive robot hesitation as an implicit mode of communication for the robot. Specifically, studies have found that a robot that hesitates more at the beginning of a "learning cycle" than at the end is perceived as "teachable" [3], that hesitation implies uncertainty of action while immediate action implies certainty [3], and finally, that human perceivable hesitation is modeled as a decceleration or stopping of a portion of the robot's velocity or movement [2].

Within the HRI field there have also been experiments for the purpose of studying social robots in the context of gameplay. Recently, a number of studies have focused on the effect of cheating by a robotic agent on human perception of the robot's agency. These studies have found that cheating robots are more likely to evoke the perception of the robot as an intelligent agent capable of making concious decisions [4].

Given this background, we were surprised to see that there had been no studies examining the connection between hesitation and the perception of a robot's actions as intelligent. As a result, we decided to study precisely that: the effect of hesitation on human perception of a robotic agent's intelligence and ability to engage in conscious thought. Although hesitation had previously been found to show uncertainty, we hypothesized that hesitation coupled with "thought sounds" during a strategy game would actually make the robot be perceived as making concious, well-informed decisions. In turn, this could effect not only a human's own decision-making speed, but may also effect a human's connection to their robot opponent.

II. Implementation

A. The Game

In order to study hesitation's effect on perception of robot's intelligence during gameplay, we decided to use the NAO, a small, humaoid robot, as the robot opponent for the experiment. Among other reasons, we chose to use the NAO because its humanoid design allowed it to be perceived as a human-like agent more easily, and therefore it would reduce external effects on the independent variable of hesitation throughout the trials.

After deciding on the NAO, we needed to select a suitable game. Initially, we considered having participants play games of checkers against the robot. Due to the NAO's size, its limited degrees of freedom, and the noise in its motors, however, we realized that implementing such a game was infeasible without making a large amount of modifications to the game, the pieces, and the board. Consequently, we realized that we needed a simpler game, in which the NAO's motion and sensing capabilities could be simplified.

This led to the creation and use of a game that we call Take Two.

1) How To Play: There are 2 players, in this case the NAO and a human participant. In front of the players is a row of colored blocks, with each color corresponding to a point value. (For example, during our trials the color-value pairs were as follows: [Blue, 2], [White, 5], [Green, 10], [Red, 25]). The players alternate turns, and on each turn the player decides between selecting one or two blocks. The blocks are chosen in order, meaning that a player decides to take only the first block of the remaining blocks in the row, or the first and the second blocks that are remaining. The game ends when there are no longer blocks remaining for a player to choose. The winner of the game is the player with the most points at the end; a player's score is determined by summing the point values corresponding to the colors of the blocks chosen by that player.

Given this structure, the simple strategy for Take Two is to maximize the number of the greatest point-valued blocks you can attain.

B. Hardware Implementation

1) The Blocks: For our trials, we 3D printed 1-inch cubes. Initially, we printed blocks with solid faces. After testing the NAO's grabbing ability with these cubes, however, we realized that the robot was prone to losing its grip on the blocks. To fix the issue, we reprinted the blocks to have one-fourth inch deep indents on opposing faces. When the NAO grabbed these indented blocks, its "thumb" would then lock into the indented region, ensuring that it would not lose its grip when pulling the block out of the ramp. The NAO's grabbing motion will be discussed in more detail below.

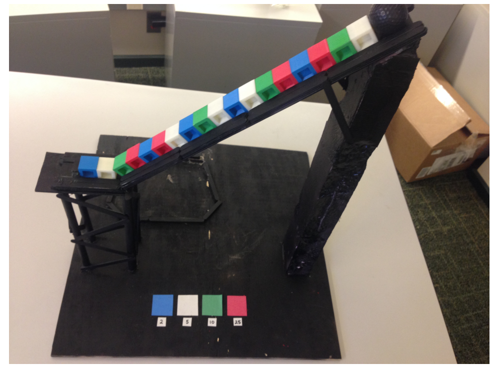

Fig. 1. Blocks on the Ramp

2) The Ramp: As mentioned above, the purpose of creating Take Two was to simplify the NAO's perception and motor responsibilities. The structure of the game allowed for us to do this given the right layout, of course. We discovered that building a ramp to hold the "row" of blocks, allowed us to solve these issues. By having the blocks on a ramp, the block to be taken would always be at the front of the ramp (given that players could only take one block from the ramp at a time), so we only needed to create the motion of grabbing the "first" block. It may not seem obvious at first why this is the case, but essentially the weight of the remaining blocks would push the blocks to the front of the front of the ramp. In order to ensure that there was always enough weight to push the blocks to the front, we placed a golf ball at the end of the ramp, on top of the last block in the row, to supply the extra weight. The golf ball was painted black to distinguish it from the blocks and to indicate the end of the game.

The ramp also allowed us to simplify the robot's perception. Following the same logic as above, the blocks that the robot could choose on its turn would always be at the front of the ramp. This means we could have the NAO look at the pixel positions corrsponding to where the first six, remaining blocks would be on the ramp. The robot's perception will be discussed in more detail below.

C. The NAO's Positioning and Posture

In order for the NAO to grab and perceive the blocks on the ramp correctly and consistently, we had to position it quite carefully. The NAO faced the ramp at a slight angle, near 80 degrees. As will be described below, this slight angle allowed for the NAO to grab blocks without largely effecting the position of the remaining blocks on the ramp. In order to replicate this position for every trial, we built footholds into the ramp that would lock the NAO's feet into the correct position each time.

It is important to note here that we had the NAO in its standard crouching position. We chose this position over the others because it seemed to be the most natural, and it allowed us to see the first six blocks easily. One issue that we faced with this position, however, was that the NAO's knee joints would heat up after prolonged use. In order to fix this, we had the NAO relax its motors when it was no longer its turn. When it was the NAO's turn again, we had it stiffen its motors and reset its head position (to ensure that it could see the first six blocks) before perceiving and making its move.

D. Grabbing Blocks

The grabbing of the blocks was a little more difficult than we had originally foreseen. As was mentioned above, we added an indentation to the faces of the blocks to allow the robot's thumb to stay in place. This removed the issue of the robot's hand sliding off the blocks. The other issue we had was that when the robot picked up a block, the other blocks on the ramp would occasionally flip or fall off. This was due to the fact that one of the edges of the block that we were grabbing would catch slightly on that of an adjacent block, causing it to rotate occasionally, and to sometimes fall out of the ramp. We adjusted for this problem by having the robot pull the blocks out at angle, which eliminated the issue.

E. The NAO's Strategy

Given that we were placing an emphasis on robot sound and hesitation, we needed to find an algorithm which could be used to play the game without skewing the participants' perception of the robot's intelligence. The game can be solved in its entirety by using a dynamic programming algorithm, applying reverse induction to the last block. This style of algorithm can determine which participant should be able to win the game, and, for each block, the optimal number of moves to make. There were two

Fig. 2. NAO Grabbing Motion

problems with using this algorithm in our design. First, using the camera on the NAO, we could only see up to seven blocks at a time. Given that our game has 18 blocks, the robot would have to look up and down the ramp for each move. We felt this behavior would have been too indicative of hesitation and thought, and thus chose not to take this route. The second reason, as mentioned above, was to produce a robot whose play style was believable and matched closely to that of the human. Running an algorithm that always chose the optimal move would skew the perception of the robot's intelligence.

To adjust for this, drawing for our own experience and speaking with others who served as testers for our game, we found that humans tend to only look about six blocks into the future. This was a perfect fit for the NAO, as six blocks would always fall directly into his line of vision. Looking at the first six blocks on the ramp, we designed a minimax algorithm for choosing blocks.

The algorithm considers four possible states: if the robot chooses one or two blocks, and then if their opponent chooses one or two blocks. The algorithm looks at each of these four states and finds both which state would yield maximum number of points for the robot and which state would results in the least number of points for the human. If the robot can score a greater number of points than the human, it chooses to remove the number of blocks corresponding to that state. If not, it will choose the move that will result in the lowest score for its opponent.

Overall we found this algorithm exhibited precisely the behavior we were hoping to find. We found the robot would make what often appeared to be the most optimal choices, but it was still susceptible to mistakes that would hurt it further along in the game. The outcome of games can be largely dependent on the order of blocks, but through our testing, we found there were many game setups where either player could win, and the outcome would depend largely on how well the human played. Given this algorithm, our robot now seemed to possess an intelligence equivalent to that of a decently experienced human player.

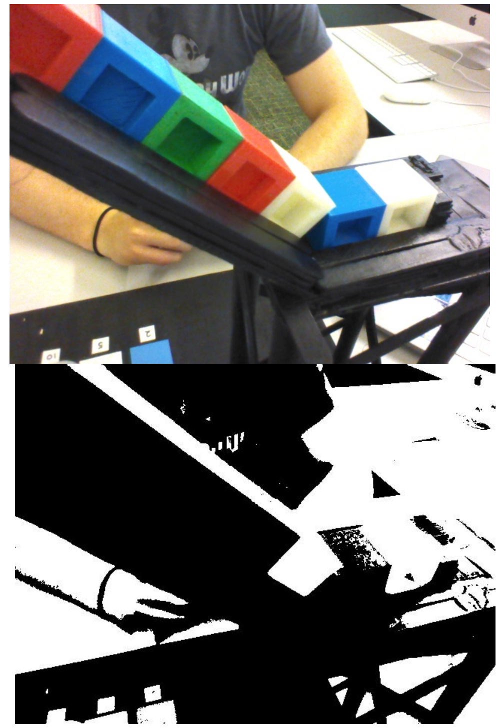

Fig. 3. Sample Images Used by the NAO to Determine Block Colors

F. Recognizing the Blocks

Now that we had an algorithm for choosing blocks, we needed a way for the NAO to be able to recognize that distinct values of the first six blocks on the ramp. This was a simple issue of color processing. Each time the player finished their move, the robot would take an image of the first six blocks on the ramp. As both the NAO's feet and its head were in a fixed location in relation to the ramp, the six blocks would always lie in the same positions on the image. Knowing this, we were able to reduce the problem of recognizing the first six blocks to analyzing the dominant color for each of six different pixel locations.

The biggest challenge with this was caused by the indents in the sides of the blocks, which were used to optimize the NAO's grabbing. These indents would be filled with shadows, making analyzing the distinction between the color of the block and the color black much more difficult. Recognizing black was important to our design as, as mentioned above, this was the technique used for understanding when the game had ended. When the NAO observed the black golf ball, it was aware that there were no blocks left, and could thus notify the participants that the game was over, and of their final scores.

To deal with ambiguity in color perception on the side of the blocks, we attempted to look at the tops, which could be seen at a slight angle given the ramp's height and the robot's position. This posed a separate issue however. Given the four colors of our blocks (white, red, green and blue) we decided the simplest way to identify white blocks was to convert to a binary, or black and white, image and check each pixel location to see if contained a majority of white pixels. Ideally, given our threshold, the colored blocks would display as black. However, when looking at the tops of the blocks, this was not the case. Too much light fell on the tops of the blocks that given any reasonable conversion threshold, there was still a possibility of colored blocks registering as white.

Our fix was a rather simple one. When trying to decide if a block was white, we chose to look at pixel locations correspond to the side of that block. The white blocks were light enough that, regardless of shadows, their indentations were still converted to white. All other colored blocks were displayed as black in the binary image, since they were much darker. After figuring out which blocks were white, we then moved to look at the pixels corresponding to the tops of the colored blocks in the unconverted image. Looking at which value, red, green, or blue, had the highest concentration, we were able to determine the color of each block. Black was determined by looking for all three values below a specific threshold. Once the colors were determined, their associated point values were fed into our minimax algorithm for choosing blocks.

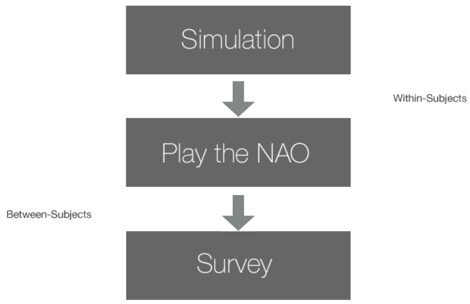

Fig. 4. Experimental Design

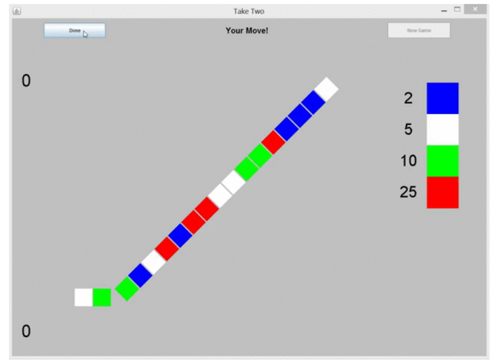

Fig. 5. The simulation

III. Experimental Design

Our experiment was divided into three different parts. The first part involved the participants playing two games on a computer simulation, as a chance for learning, and for determining a baseline of how long the participant took to make his move. The second part of the experiment was our main focus, which involved playing two games against a NAO. In this part of our experimental design we tested the effects of hesitation and sounds. The final part was a survey in which we gathered data on how the participants felt about the robots intelligence, naturalness, and other similar characteristics.

A. The Simulation

Upon first arriving at our experiment, participants were briefly explained the rules of the game, and were then asked to play two games against a computer simulation. This simulation was made with Java, and ran the algorithm that was described above, the same algorithm that was implemented on the robot. The blocks were generated randomly, but the distribution of colors was kept the same as it was in the physical game played against the NAO. The main purpose of the simulation was to teach participants how to play, but we also used it to collect data on how long a player takes in between moves. Following their move, players were instructed to touch a “Done” button, signaling to the computer to take its turn, and allowing us to record the time it took participants to make their moves. This data was used as a baseline to observe how participants’ move time changed while playing against the robot. Between each move, the simulation would pause for one second. The participants were asked to play two games, and then to play the robot.

B. Playing the NAO

Once participants moved to the robot, they were asked to tap its right foot to begin the game. The robot then gave a brief overview of the rules, and told the participants to tap its head after the completion of their move, similar to pushing the button in the simulation. Participants were asked to move first, and the game was played until no blocks remained. After completing the game, the robot would state the score of each player, and whether it won or lost. Instructions would then be given to the participant to reset the blocks on the ramp as illustrated in a diagram, and tap the robot’s head to start the second game. This game was also played to completion.

The trials that we ran were divided into four different categories, with each participant playing two games of the same type. The first of these four was the neutral sounds trial. In this trial, the robot would make one of three different mechanical sounds while making its move. The three sounds were a clock winding up, a printer scanning, and some gears clanking. The sounds were meant to simulate natural mechanical noises that might be made by a robot, and after speaking with participants, we found that they seemed to fill this role fairly well. The second trial was the thought noises trial. Here, participants would hear one of four humanlike thought sounds. Each played in a voice similar to that of the NAO’s. These sounds were variations on human thinking noises, such as “hmm” and “umm.”

The next two trials used similar sounds, but incorporated hesitation. While making the sound, the robot would hesitation for exactly three seconds. This hesitation could take place in three different ways. The first option was that the robot could hesitate and make a sound prior to removing any blocks. The other two options depended on how many blocks that robot decided to take. The NAO would remove the first block, and then reach up and pause above the second block. The robot would then pause for three seconds while making its noise and either take the block or pull back, depending on the output of the algorithm. The matrix below details our four different trials.

The hesitation time of three seconds was based on our own hesitation time when we played against the simulation. More importantly however, we noticed that hesitation time that was greater than three seconds seemed to convey that instead of thinking, the robot was just slow to process or perhaps broken. We would like to test with longer hesitation times in the future.

C. The Survey

Following the completion of the games against both the simulation and the NAO, participants were asked to fill out a survey of 20 questions. Most of these questions fell into the following five categories: intelligence, distraction, naturalness, enjoyment, and connectivity. Each of these was ranked on a scale from one to seven, and participants were additionally asked to answer two yes-no questions regarding whether they thought the robot was smarter than them, and whether they would want to play the robot again. Following the completion of the survey, the trial was complete.

IV. Results and Discussion

During each participant trial, we collected data in two particular ways: by timing the duration of each participant's turns while playing Take Two against the computer simulation and the robot, and by asking each participant to fill out a post-trial survey.

The turn duration data collection was performed by the computer simulation and by the NAO. While a participant played against the simulation, a time stamp was recorded each time the simulation finished a turn and again each time the participant clicked the button indicating he had finished a turn. The difference in these two time stamps was exported to a file for later analysis. Similarly, while the participant played against the NAO, a time stamp was recorded each time the NAO completed a turn, and again each time the participant tapped the NAO's head to indicate that his turn was complete. The difference in these time stamps was also exported to a separate file. It is important to note that for games in which the participant took the first turn, the time for completion of that turn was not recorded due to variability in the time for explanation or set-up of the games that we could not standardize.

TABLE I
Matrix for Trial Types

	No Hesitation	Hesitation
Neutral Noises	Trial 1	Trial 3
Thought Noises	Trial 2	Trial 4

After a trial, participants were asked to complete a twenty-question, online survey. The first eighteen questions asked participants to rank their experience playing against the NAO across a variety of criteria on a Likert scale from 1 to 7. The last two questions were yes-no questions that asked participants if they would play the robot again and if they thought that the robot was smarter than they were.

In our study, we looked at the effects of two independent variables: noise type and hesitation. This gave us a two-by-two matrix design, with four possible trials: no hesitation and neutral noise (no hesitation-neutral), no hesitation and thinking noises (no hesitation-thought), hesitation and neutral noise (hesitation-neutral), and hesitation and thinking noises (hesitation-thought). We ran our study on 37 participants, but had to discount five of them due to issues with the robot during experimentation that could have influenced the data. This left us with 32 participants, eight participating in each of our four different types of trials.

To analyze the turn duration data, we first calculated the average time between turns for each participant in the games against the simulation and

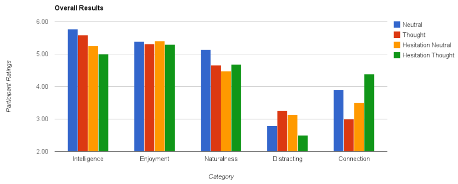

Fig. 6. Results

in the games against the robot. To do this, we found the sum of all the times in the time data files, and then divided by the number of turns the participant took. Then we calculated each participant's percent change in turn duration between games against the simulation and games against the robot. We did this by subtracting the average simulation turn time from the average robot turn time and dividing by the average simulation turn time. We recorded this value for each of the 32 participants.

To analyze the survey data, we split the eighteen Likert scale questions into five categories, each relating to a different aspect of the participants' interaction with the NAO: how intelligent the robot was (intelligence), how enjoyable their interaction with the robot was (enjoyment), how natural the robot acted (naturalness), how distracting the robot was (distraction), and how connected they felt to the robot (connectedness). For each participant, we found the average ranking they assigned the questions in each of the five categories. This allowed us to record the average score that a participant gave to each category. After compiling the data, we had six points of data for each participant: there percentage time difference in completing turns against the simulation versus the NAO and their average ranking in each of the five survey question categories.

After we compiled the data, we performed two-way ANOVA tests to find significance. We chose this type of test because we needed to look at the significance of two different independent variables. For each of the six tests we did, we found a p-value for the effect of neutral versus thought noises, hesitation versus no hesitation, and the interaction of these two variables.

We found significance in only one place in our data. With a p-value of 0.0250, we were able to conclude that hesitation did have an effect on the way in which participants viewed the intelligence of the robot opponent. When we examined the means of the groups, we found that robots were perceived as more intelligent if they did not hesitate. We hypothesized that this is because the NAO was faster when it did not hesitate, and might therefore be seen as a quicker decision-maker. A hesitating NAO would take a longer time to make its move, and may then seem less sure of whether it is correct. Because we set up the games so that the NAO would always win, participants did not seem to feel that the robot was less intelligent because it made incorrect moves. In fact, 13 of the 16 participants in no hesitation trials and 12 of the 16 participants in hesitation trials indicated in the post-trial survey that they felt that the NAO was smarter than they were. Instead, we hypothesized that participants saw the robot as more confident and a faster decision-maker if it did not hesitate.

We found absolutely no significance in participants' ranking of enjoyment when playing against the NAO. The data showed p-values of 0.7013 for the effect of thinking noises and 1.000 for the effect of hesitation and the interaction of the variables on participant enjoyment. It seemed that details like the noises made or the time to take moves did not influence how much people liked the NAO and the game. In fact, all 32 of the participants indicated in the survey that they would play Take Two against the NAO again. We hypothesize that this might be due to the novelty effect. Participants were asked to interact with a friendly robot that they had never interacted with before. The nature of the interaction was to play a game, which is inherently fun and

TABLE II
TABLE OF RESULTS

	Hesitation p-value	Noise p-value	Interaction p-value
Intelligence	0.0250	0.2555	0.8181
Enjoyment	1.0000	0.7013	1.0000
Naturalness	0.3402	0.6462	0.2314
Distraction	0.7359	0.9104	0.3156
Connectedness	0.7359	0.9104	0.3156
Decision Time	0.4474	0.2541	0.2429

entertaining. These may have combined to make the participant interested in the NAO, regardless of whether it overtly appeared to think.

After we ran the two-way ANOVA tests, we graphed the means for each of the five categories of questions. The graph demonstrated the trends that we discussed above for intelligence and enjoyment. We also found that participants seemed to feel more distracted by and less connected to the NAO in the no hesitation-thought and hesitation-neutral trials than in the no hesitation-neutral and hesitation-thought trials. We hypothesize that this is because no hesitation-neutral and hesitation-thought trials best reflect what humans would expect. It is natural for a robot to act robot-like, i.e. to make mechanical noises and to perform immediately. It is also logical to have a robot act human-like, to stop during a game and to make noises indicating thought. Thus, we could expect that participants would find these robots less distracting and more able to be connected with the robots that exhibit both human-like and robot-like thought.

Finally, we graphed the percentage change in time from the simulation games to the robot-opponent games. We found that participants took consistently longer across no hesitation-neutral, no hesitation-thought, and hesitation-thought. However, participants took notably longer during the hesitation-neutral study. We are unsure why this was the case, and we would recommend future testing to see if this result holds constant.

V. Conclusion

We have designed a robot that is able to play a simple game with blocks. In four different conditions, the robot either hesitates or does not hesitate, and either makes thought noises or neutral mechanical noises.

We find that, contrary to our expectations, there was little effect of robot thinking behaviors on participants' perception of the robot across almost all of the categories that we measured. We also found, for example, absolutely no effect of robot hesitation or noise on the participant's enjoyment of the game; participants rated their enjoyment highly regardless of the robot's behavior. We attributed this to the novelty effect. Participants were more pleased by the fact that a new robot could play an interactive game than they were concerned by the way in which the robot played it.

We find, however, that the robot's hesitation had a significant effect on participants' perception of intelligence. Participants viewed a slower robot as less intelligent, regardless of whether it sounded as though it was thinking. We also note an emerging trend in the data: a robot that either performed both or neither of the actions associated with thinking was rated as less distracting and more able to be connected with.

These results indicate that people feel more comfortable with robots that behave either robot-like or human-like, but not some combination of the two. Participants were most comfortable with the robot moved quickly with mechanical noises, as a robot would, or moved slowly with thinking noises, as a human would. However, the effect of these two variables was overshadowed by the novelty effect. Participants enjoyed the robot opponent regardless of the presence or absence of thought.

VI. Future Work

In the future, we would like to expand on this study in the following ways. First, we observed that a lot of the participants really enjoyed the game we created. When some of the participants left their trials, they asked us for the name of the game and commented on how much they enjoyed the experience. As a result, we think that the novelty effect of the game may have overshadowed the NAO and its "intelligence." To explore this issue, a future experiment with a more familiar, less involved game would be ideal. Next, as mentioned above, we had to discount five participants, a few of which were caused by perception errors. We feel that conducting the study with a better color perception algorithm and enhanced lighting could create stronger results. Finally, we noticed that when participants hesitated, they hesitated for eight to ten seconds. During this study, however, the NAO hesitated for only three seconds. We would like to explore the possibility of having the robot hesistate for longer, while at the same time ensuring that it does not seem broken by having it make gestures or sounds as well.

References

[1] Christoph Bartneck, Takayuki Kanda, Omar Mubin, and Abdullah Al Mahmud, *The Perception of Animacy and Intelligence Based on a Robot's Embodiment*. 2007.

[2] Christian Dondrup, Christina Lichtenthaler, and Marc Hanheide, *Hesitation Signals in Human-Robot Head-on Encounters*. HRI'14, March 36, 2014, Bielefeld, Germany.

[3] Tanaka Kazuaki, Ozeki Motoyuki, and Oka Natsuki, *The Hesitation of a Robot: A Delay in Its Motion Increases Learning Efficiency and Impresses Humans as Teachable*, . 2010.

[4] Elaine Short, Justin Hart, Michelle Vu, and Brian Scassellati, *No Fair!! An Interaction with a Cheating Robot*. 2010.