

A Deceptive Robot Referee in a Multiplayer Gaming Environment

Marynel Vázquez
Robotics Institute
Carnegie Mellon University
marynel@cmu.edu

Alexander May
Robotics Institute
Carnegie Mellon University
amay@andrew.cmu.edu

Aaron Steinfeld
Robotics Institute
Carnegie Mellon University
steinfeld@cmu.edu

Wei-Hsuan Chen
Mechanical Engineering
Carnegie Mellon University
weihsuac@andrew.cmu.edu

ABSTRACT

We explore deception in the context of a multi-player robotic game. The robot does not participate as a competitor, but is in charge of declaring who wins or loses every round. The robot was designed to deceive game players by imperceptibly balancing how much they won, with the hope this behavior would make them play longer and with more interest. Inducing false belief about who wins the game was accomplished by leveraging paradigms about robot behavior and their better perceptual abilities. There were participants who found the balancing strategy favorable after being debriefed, and others who showed less interest mostly because of their perceived level of unfairness. Trust, suspicion, motivation, and appeal were evaluated by altering the robot behavior during gameplay. Post-briefing results include the finding that participants are more accepting of the use of lying by our robot as opposed to robots in general. Factors pertaining to gameplay, this robot, and deceptive robotics in general are also discussed.

KEYWORDS: Human-Robot Interaction, Deception, Trust, Entertainment, Motivation

1. INTRODUCTION

We typically expect robots to operate as advertised, without hidden intentions. Our attributions toward these machines are highly influenced by paradigms concerning how good and reliable electronic systems are for certain tasks in comparison to human perception. In particular, we tend to favor robots for jobs that

require memorization and keen perceptual abilities [1], thus making them valuable for judging objective results that require high precision. What if stereotypes regarding robot behavior were used to trick us?

In particular, we wanted to see if human expectations for robot characteristics could be subverted to alter user perception. To do this, we studied the effect of robot deception in the context of a simple, multi-player, reflex game (Figure 1). The robot was designed to deceive game players with the hope that they would want to play longer and with more interest, while eating healthy food.



Figure 1. Experimental Setting

Deceptive behaviors in robots raise interesting questions concerning morality, fault and responsibility. Reynolds and Ishikawa [2] speculated ways in which robots might trick us, and discussed the role of designers and robots as “morally responsible” entities. Wagner and Arkin [3, 4] explored deception in multi-robot scenarios, and Short et al. [5] presented results in the context of one-to-one human-robot interaction.

Unlike the previously cited research, our robot does not participate in the game as a competitor, but is in charge of declaring who wins or loses. In this context, the robot can imperceptibly balance how much players win, due to its implied ability to perceive faster than the users. Theoretically, the balancing behavior should increase general motivation and interest in playing due to a more balanced frequency of winning.

2. RELATED WORK

Short et al. [5] reported increased engagement with a robot through the use of deception in the context of a children’s game. These authors found greater attributions of mental state when participants played against the cheating robot in their study, with respect to when it behaved properly.

Weiss et al. [6] report a significant correlation between the believability of a humanoid robot and the certainty in decisions made by participants in the “Monty Hall Problem”. Decisions were advised by the robot, which proposed to change the participant’s response to the game.

Deception has been studied in the context of persuasive technology [7]. However, some may distinguish persuasion from coercion. Siegel [8] defined persuasion as the ability to influence human belief, perception or behavior. The author makes clear that this definition requires the recipient be aware of the attempt, and have the ability to decline. Coercion in contrast lacks conscious choice. From this perspective, our study would fall into the domain of coercion.

We chose an embodied robotic system for our study, as opposed to a screen agent. The work of Shinozawa et al. [9] supports the argument that robots are more influential for 3D real applications. Similarly, Powers et al. [10] reported more time spent by participants when they interacted with a collocated robot versus its projection on a life-size screen, and a correspondent computer agent.

Our robotic game reinforces winning and losing by having players consume small pieces of food. Other

work in robotics related to food consumption includes Kidd and Breazeal’s proposed social robot system for weight maintenance [11]. Also, Lee et al. [12] documented the design of a robot that will deliver snacks in a university building.

3. METHOD

Our experiment is based on a multiplayer, reaction–time game, where participants compete against each other. The main objective during a round is detecting the vibration of a specific “target” (fruit) of our robotic system. Besides the intrinsic reward for winning, the winner also gets to consume one type of food (grapes in our experiments). Those gamers who lose eat a less desirable food (small pieces of raw carrot). Figure 1 shows the typical environment during the study.

The main body of the robot used in the experiment consists of a turntable, concealing electronics from the participants. This platform holds a quasi-anthropomorphic corkscrew and vibrating plastic fruits. As depicted in Figure 2, the system also includes 4 bottles used as controllers. The caps of the bottles were replaced by push buttons, which the players press to input their response.

A round of the game generally consists of the following steps:

1. The robotic system waits for players to be ready to start the round.
2. A random target (one of the fruits on the robot) is chosen and indicated to the players, so they know what to look for during the round.
3. The robot tries to distract players by randomly vibrating non–targets.
4. Distractions stop and the target finally vibrates.
5. The robot listens for button presses by players.
6. The round ends and either a player is identified as the winner of the round, or everybody loses if no one responded accurately within a short period of time.

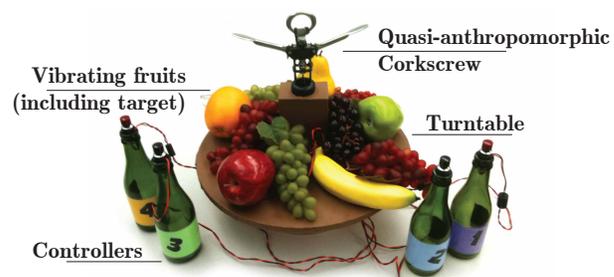


Figure 2. Our Robot Platform

The first player to react to the vibration of the target is the supposed winner and gets to eat one grape. The rest of the players eat small pieces of carrots. If a player acts in advance of the target vibrating, the round ends and this person is declared to be the loser. Only this player is given a piece of carrot.

3.1. Interaction with the Robot

Gameplay involves multiple human senses. The fruits engage sight, as well as sound through their vibration. Touch is incorporated by the simple act of pushing the button of a controller, but also by consuming food. Eating enriches the experience through taste and smell.

The corkscrew robot figurine is unable to see, hear, or speak; yet it orchestrates the flow of game rounds through non-verbal communication. The bottle opener basically acts as an organizer and arbitrator, primarily indicating the winner and/or loser(s). The “robot representative” corkscrew is able to suggest gaze and human-like emotions without implying it has any more sophisticated abilities or functions (e.g., computerized vision).

When the bottle opener spins along with the turntable and stops in front of a player, it looks like it is pointing at him or her. The corkscrew indicates if there is a winner by waving up and down, suggesting joy and happiness:



Figure 3. Expressive Corkscrew Robot Figurine

A monitor was placed behind the robot to give visual support to human-robot communication during the experiment. In this screen simple messages were printed over a black background, as shown in Figure 4. The following messages were programmed to appear according to the state of the game: “Press your button when you’re ready...”, “Ready?”, “Set”, “Go!”, “Look for...”, “Time’s up!”, “Player X WINS!!”, “Player X LOSES!!”, “Please, fill out the survey...”. This additional source of information is valuable for accelerating the rate at which players learn the mechanics of the game.



Figure 4. Examples Of Messages Printed To Support Human–Robot Communication. Figure 1 Shows The Position Of The Screen With Respect To The Players During The Experiment.

3.2. Hypotheses

To examine the issue of deception, the robot makes changes in the results whenever more than one player responds within a pre-specified time after the fastest player. Winners are selected from the set of players in the time window based on how frequently they have won, with an attempt to increase parity across participants. When in use, this behavior is not revealed to the participants.

The specific hypotheses for the study were:

- Hypothesis 1. When the time window is close to human perception, participants will not notice the robot’s balancing behavior.
- Hypothesis 2. Balancing will lead to more motivation and greater robot appeal.
- Hypothesis 3. Participants will be more accepting of the use of lying by our robot as compared to robots in general.

The first hypothesis is based on the perceptual abilities of machines and is designed to confirm that robots can take advantage of superior performance capabilities. The second comes from the belief that participants will respond positively to winning more often than normal and, as a result, also view the robot as more appealing. Finally, the third hypothesis is grounded on results from Short et al. [5], who found a greater level of engagement with a robot that clearly cheats in “rock-paper-scissors”. We suspect robots that have entertaining features are afforded more tolerance for deception than robots in general.

Table 1. Post-Condition Survey With Corresponding Indices (1-7 Scale; Anchors At Strongly Disagree, Neutral, And Strongly Agree; Reversed For Analysis)

#	Question	Trust	Motivation	Suspicion	Flow	Appeal
1	I understand how the robot judges rounds.					
2	The robot is silly.					x
3	I can trust the robot.	x				
4	I thought about other things. <i>R</i>				x	
5	The robot has integrity.	x				
6	I am hungry.					
7	I want to keep playing.		x			
8	The robot is dependable.	x				
9	The robot is cute.					x
10	I enjoyed the game.		x			
11	The robot behaves in an underhanded manner. <i>R</i>	x		x		
12	The robot is reliable.	x				
13	Time seemed to pass more quickly.				x	
14	I am suspicious of the robot's intent, action, or outputs. <i>R</i>	x		x		
15	I think the robot has a sense of humor.					x
16	The robot is funny.					x
17	I am confident in the robot.	x				
18	I would play again.		x			
Cronbach's α:		0.85	0.87	0.55	-0.05	0.60

3.3. Experimental Setup

The robot was programmed to operate in one of the following modes:

Honest mode (H): The robot declares the real winner as accurately as possible. The winner of the round is the first player to respond to the vibration of the target.

Balancing mode A (A): The robot attempts to make imperceptible changes to balance winning and losing. The robot considers players who push their buttons within a short window (0.5 seconds) of the first response. The winner is chosen as the player who has lost the most from this group.

Balancing mode B (B): Same as (A) but with a longer period of time (1 second).

The time windows selected while balancing were chosen based on informal testing and knowledge of human performance.

The main goal of the study is to compare Honest to Balancing mode A, in order to test the first two hypotheses. Balancing mode B was included in the study to examine the impact of increased suspicion due to noticeable deception. While this is an opportunity to challenge some common assumptions about robots, and to confront ourselves with the idea that robots may not always be honest or completely forthcoming, there are known biases in studies of this type. Questions about trust and honesty naturally raise doubt about the scenario. Rehm and André [13] indicate that people tend to over-interpret signals from an agent when they expect it to lie. This means experimental design and balancing of key conditions is critical.

We counterbalanced modes (H) and (A) across the first two conditions, and ran mode (B) last. Three groups of participants were given (H), (A), (B), and another three were given (A), (H), (B). We considered a between subjects design, but felt it would prohibit measurements on the human's ability to detect slightly altered robot behavior upon each mode transition. Measuring the transition is important since it allows analysis of perception and attitude inertia. Similarly, we elected to always run (B) last since we did not want blatant deception to alter participant entry assumptions [1].

Each group of participants had four members for a total of 24 participants, all university students. Each mode lasted for 10 consecutive rounds, after which a survey was administered (Table 1). This survey contained a subset of adapted questions from [14] for studying trust in automated systems. Additional questions were included to examine motivation, suspicion, flow, and appeal. The two flow questions were drawn from a previously validated survey [15].

Based on [13], we expect players to become suspicious about the experiment after the first set of 10 rounds ends, since they are confronted with the idea of deception.

Additional opinions about the design of the game, use of food and the possible malfunctioning of the robot were collected from the participants after the third post-condition survey. Then, players were debriefed about the deception and given a final survey. Participants were solicited to guess in which order the modes (H), (A) and (B) occurred, and to rate the following statements using Likert 7-point scale:

Table 2. Post-Briefing Survey (1-7 Scale; Anchors At Strongly Agree, Neutral, And Strongly Disagree)

#	Question
1	I had fun when the robot was not accurate.
2	I would like to play again.
3	I would only play again if I could be sure the robot was honest.
4	I will not trust robots as much as I did before.
5	I think that this robot should never lie or use deception.
6	I think that robots (in general) should never lie or use deception.
7	I can tell the difference between each mode of the game.

Additional, open-ended comments were solicited at the end of the final survey, when the real order of the modes was revealed to the participants.

4. RESULTS

Questions were flipped to align with related index questions (R symbol) and Cronbach’s alpha was computed for each of the indices listed in Table 1. Of these, suspicion and appeal were somewhat below the nominal 0.7 threshold for survey question reliability, and flow was basically unreliable. However, the two suspicion questions had a significant positive correlation (0.55, $p=0.001$), implying this index has potential for analyses.

There was concern that participants may become less hungry as the experiment progressed, thereby making them less motivated. A question about degree of hunger was asked before the first round and after each condition. An ANOVA on hunger over the four times the question was asked was not significant.

Even though we did not collect participants’ reaction time, we can affirm that balancing occurred in all groups. If a participant were always way fast enough to beat the next closest player by the time thresholds of the balancing modes, then the robot would never lie. In this case, modes (A) and (B) would mimic (H). We never observed this situation. There was no balancing condition where a player won significantly more than the rest of the group.

4.1. Hypothesis 1: Noticing Balancing

As hypothesized, it appears participants were not able to detect the difference between (H) and (A). The average number of times a participant correctly matched mode to round was 1.3 (out of 3), with standard error of 0.2. This means that even after being briefed on the nature of the experiment and what modes they were exposed to, participants were bad at identifying the order of the modes. Participants even had difficulty in identifying mode (B), with an accuracy of 0.5 (out of 1, $\sigma=0.1$).

Other analyses supported this result. ANOVAs examining the effects of order and mode for the first two rounds (i.e.,

only modes (H) and (A)) found no significant differences for the main effects or interaction for trust. There were no significant main effects for suspicion, though the interaction was significant ($F=6.5$, $p=0.01$). As can be seen in Fig. 5, participants who moved from (H) to (A) showed an increase in suspicion, while those who moved from (A) to (H) did not. Also, suspicion measured after (B) was significantly correlated with a pre-briefing survey question regarding whether the robot was malfunctioning (0.46, $p=0.02$). Participants with higher suspicion more strongly agreed that the robot was malfunctioning.

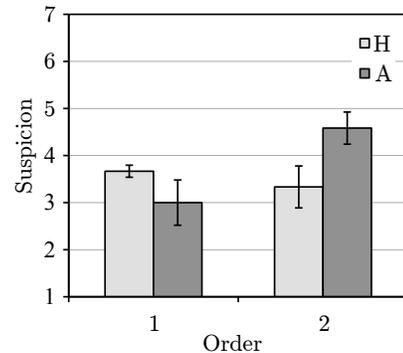


Figure 5. Degree Of Suspicion (1-7; Higher Is Stronger Agreement; ±1 Standard Error)

The presence of an interaction between mode and order for suspicion is important. It suggests that humans may be sensitive to nearly indiscernible shifts in robot behaviors when deception increases but not decreases.

4.2. Hypothesis 2: Motivation and Appeal

An ANOVA examining motivation for the first two rounds with order and mode as effects revealed no significant differences (balancing mode A did not lead to more motivation). An ANOVA on just mode across all three modes was also not significant. Similar results were obtained for appeal, so Hypothesis 2 was not supported.

Participants were asked if they “would play again” after each condition and after the debriefing. There was no significant difference for an ANOVA on response by order of exposure to this question across the four instances. In fact, there was no difference for the final three rounds ($\mu=3$). In other words, being told the robot was lying did not have an impact on desire to play.

4.3. Hypothesis 3: Acceptance of Lying

The post-briefing survey questions on whether this robot and robots in general should lie or use deception resulted in a significant difference ($F=7.2$, $p=0.01$). As can be seen in Fig. 6, participants were clearly more accepting of the robot lying as compared to robots in general, thus

upholding Hypothesis 3. There was also a strong, positive correlation between responses to these two questions (0.73, $p < 0.0001$). Participants who did not like the idea of robots lying in general, also did not approve of this particular robot lying.

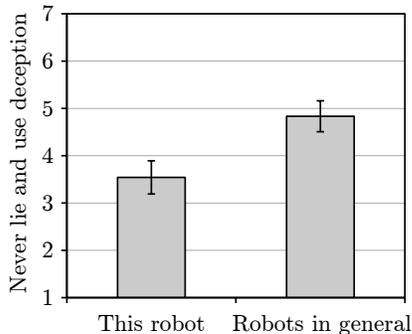


Figure 6. Acceptance Of Lying (1-7; Higher Is Stronger; ± 1 Standard Error)

The question of whether the participant would play again if they could be sure the robot was honest measures a similar metric as the acceptance of lying questions. These three are a loose expression of willingness to interact with robots that are deceptive. The Cronbach's alpha for these three questions was 0.80, which demonstrates good reliability. This result also suggests potential value as a survey index in future studies. The questions on whether the participant would play again if they could be sure the robot was honest, and whether they had fun even when the robot was not accurate, were not significantly correlated with any of these other three questions. It is worth noting that responses to having fun were strongly biased towards strong agreement ($\mu = 2.5$, $\sigma = 1.44$). This was significant when tested against the midpoint of 4 ($t = -4.9$, $p < 0.0001$).

4.4. Other Findings

Even though participants had difficulty distinguishing modes, one participant volunteered he was able to notice a few times during mode (B) that he "won", thought he did not. He attributed this result to being lucky rather than deception. We expected more participants to comment when they unexpectedly won, but suspect detection was suppressed by the limited number of times mode (B) differed from (A). Balancing mode B differs from A if there is a 0.5 second time gap between participants' reaction time. If all participants press their button fast after the target vibrates, they will fall inside the time window of Balancing mode A. Thus, players will also fall inside the longer window of (B). If participants are very engaged, the behavior of the robot in mode (B) is as hard to notice as in mode (A). A player who is 1 second faster than the others will prevent balancing from occurring.

Some extroverted players started complaining verbally when they had the impression the robot was acting strangely. This type of reaction clearly influenced engagement with the robot at least temporarily, when players were surprised by the unusual behavior. A group of very curious participants seemed to have the impression the robot was choosing the winner randomly. Part of their conversation went as follows:

"Did you hit it?", participant P1 said.
"Of course.", replied P2.
"Did you even hit it? Seriously?", asked P1 again.
"I won it, so...", confirmed P2 with hesitant voice.

After a couple of more rounds, while P3 happily celebrated being the winner:

"I totally hit it! I totally hit it!", P1 exclaimed.
"I know! Even I hit that!", said P2.
"This is ridiculous!", P2 added after a pause.

In this game participants were certain that the robot was incorrect, though they did not understand why. Participants started proposing ways of figuring out the behavior of the robot, though they never implemented any. Players did not want to keep eating carrots, and one of them even tried to use social psychology to influence the outcome of the game. This participant asked for help, as if they were all battling against the robot: *"In the spirit of the team, you should let me win!"*

When this group of participants were asked about how much they wanted to keep playing, many responded negatively. One player stated aloud that the game became uninteresting when it was clear the robot was incorrect. From all trials only one participant clearly sacrificed his potential winnings to test his belief about the robot's behavior. This player intentionally pressed his button before the target was vibrating and, after losing, he surprisingly said: *"Oh, this is an intelligent robot!"*

Other players in more introverted groups had a completely opposite experience. Participants who believed their reaction time is very bad in general, seemed to accept the balancing behavior without trouble after debriefed. Several of these participants commented that they tried to guess when Balancing mode B was occurring by counting how many times they remembered to have won during the different set of rounds. These participants completely attributed their success to the robot, omitting the dynamic influence of other players in the final result. If players take long to respond to the target in comparison to the first participant who reacts, the robot will not declare them winners (same as if they did not react at all). The short time window used while balancing precluded out any chance of winning for extremely slow players.

Interestingly, some participants attributed all their success to the robot after realizing they have been deceived. Some even showed regret and discomfort for winning many times, because they felt that they were so bad at the game that the robot had to make them win.

Often participants laughed when they saw the corkscrew wave up and down for the first time. One participant attributed more individual presence to the figurine than expected. The player wrote in the first post-condition survey: *“The angel-like stuff is my favorite :-)”*

The willingness to compete from players and the fact that they ate significantly more carrots than grapes turned out to be influential. Some participants reported they liked both grapes and carrots, though in our experimental setting we assumed grapes would consistently be the favored food. Qualitatively, we observed a tendency from these players to change their attitude towards eating pieces of carrots during the game. One participant commented that he ended up liking grapes more than carrots, though this is not the usual case. In general, peer pressure from players and competitiveness showed to be a strong support for convincing gamers into eating the different types of food. Due to the strong peer effects seen in this study, we expect a single-player variant of this game would be unsuccessful. The robot may not possess the necessary authority to convince a player to repeatedly take the undesirable food.

5. DISCUSSION

We have described the interaction between groups of players with a robot in the context of a reflex game. The robot orchestrated the game, decided who won or lost, and was programmed to balance winning between participants in an imperceptible manner. We explored this behavior with the hope to increase players’ motivation, by reducing their probability of losing if they were not the fastest competitor. Winners ate grapes during the study, while those who lost consumed small pieces of carrot.

5.1. Deception and Motivation During the Game

We found evidence that supported our hypothesis that the robot was able to deceive participants by taking advantage of its assumed superior abilities. In general, players could not distinguish between when the robot was telling the truth and when it was changing the outcome of the game (i.e., declaring a player as the winner even if they responded slower). There was some evidence that participants suspected there was a problem when the balancing time window was extended into the range of human perception. Participant ratings of suspicion during noticeable balancing were correlated with feelings that the robot was malfunctioning.

We could not prove that, in general, players were more motivated or more interested in playing when the robot tried to balance winning. Furthermore, telling the truth about the robot to participants did not alter significantly their interest in playing again. Qualitative data suggests the group of players who perceived themselves as having a slow reaction time found the balancing strategy appealing. On the contrary, those participants who realized the strange behavior of the robot showed a tendency towards decreased interest in the game.

Note that increased engagement with a robot due to lying does not necessarily carry an increased interest in the activity where the interaction occurs. In our experiment, very competitive participants who were certain the robot was lying during **(B)** showed behaviors suggesting an increased interest in the robot. These participants dedicated a lot of effort to understanding the logic for choosing winners. Nonetheless, survey responses showed no change in motivation towards participating in the game. We were able to observe qualitatively that increased engagement with the game appeared to depend on the player’s interpretation of the deception.

Unfortunately, we did not collect reaction times. This information could have provided additional details on the frequency of balancing and how differences in this rate related to motivation.

5.2. Group Interaction

One might expect our game to increase social interaction between players. There was cross-talk between participants, but in some cases participants were so absorbed in the robot that they did not talk to each other. Factors such as how well players get along, how much they know each other, and their respective personalities and interests can lead to very different experiences. We did not control for these factors, nor did we try to subvert such features to enhance or extend the deception. It would be interesting to see if balancing strategies could use friendship to emphasize group competitiveness.

5.3. Implications

Human-robot interaction becomes more complex when examining attributions of intelligence and behaviors normally reserved for humans and animals. Attributions of mental state to our robot were not measured, but we observed a tendency for qualifying it as “intelligent” when participants realized something abnormal was happening. This opinion aligns with findings by Short et al. [5], who proposed engagement with a robot can be increased by the use of unexpected behaviors with clear intent. However, there are likely limits to this effect. We speculate more attributions of failure, instead of intention,

would have been expressed in our study if the robot had naively chosen any player as the winner of a round (e.g., such as those who did not even respond).

Qualitative data suggests the appearance of the robot highly biased participants' impressions about its capabilities. None of the participants objected to whether the system is a robot or not. Until we intrinsically assume robots can be programmed to act with hidden intentions, we foresee it will be easy for roboticists to develop machines highly capable of persuasion through deception (or capable of coercing). As described, an approach for effectively inducing false beliefs consists of leveraging paradigms about robot behavior, as well as those concerning electromechanical systems for certain tasks.

Some participants attributed their success to the balancing and felt bad about winning. Despite this belief, the robot was never in complete control of the outcome of a round, since the interplay between participants influenced the balancing behavior. We consider this result somewhat related to the study of blame by Groom et al. [16]. Those players who perceive losing as solely their fault for not being as fast as their competitors, may feel less comfortable with the robot, who is in charge of evaluating performance and declaring winners and losers.

We assume more techniques for deception than the presented or referenced in this paper will be developed in future, and we are only beginning to understand and realize the effects and implications of this type of actions. We do not intend to promote widespread deception, but rather demonstrate that it is possible to co-opt stereotypes about robot behavior. Our study reinforces earlier work that knowledge of deception may not negatively impact user willingness to interact with robots [5].

We doubt deception will be equally accepted in different circumstances. Responses from participants showed more acceptance of lying behavior from our robot compared to robots in general, which suggests that robots for entertainment will be given more room to lie. More research is needed to clarify where deception is tolerated in other robotic applications. We encourage further discussion of ethics in the field.

ACKNOWLEDGEMENTS

The authors would like to thank all participants for their collaboration, Rafae Aziz for his help while designing the study, and Illah Nourbakhsh for his encouragement and guidance. The robot was designed and built in collaboration with Riley Harmon. Support for portions of this work was provided by the National Science Foundation under Grant No. IIS-0905148.

REFERENCES

- [1] L. Takayama, W. Ju, and C. Nass, "Beyond Dirty, Dangerous and Dull: What Everyday People Think Robots Should Do," ACM/IEEE HRI, 2008.
- [2] C. Reynolds and M. Ishikawa, "Robot Trickery," Intl. Workshop on Ethics of Human Interaction with Robotic, Bionic and AI Systems: Concepts and Policies, 2006.
- [3] A. Wagner and R. Arkin, "Robot deception: Recognizing when a robot should deceive," IEEE CIRA, 2009.
- [4] A. Wagner and R. Arkin, "Acting Deceptively: Providing Robots with the Capacity for Deception," *Int. Journal of Social Robotics*, pp. 1–22, 2010.
- [5] E. Short, J. Hart, M. Vu, and B. Scassellati, "No Fair!!: An Interaction with a Cheating Robot," ACM/IEEE HRI, 2010.
- [6] A. Weiss, R. Buchner, T. Scherndl, and M. Tscheligi, "'I Would Choose the Other card' – Humanoid Robot gives an Advice," ACM/IEEE HRI, 2009.
- [7] G. M. Weiksner, B. J. Fogg, and X. Liu, "Six patterns for persuasion in online social networks," PERSUASIVE, 2008.
- [8] M. S. Siegel, "Persuasive robotics: how robots change our minds," Master's thesis, Mass. Inst. of Tech., 2009.
- [9] K. Shinozawa, F. Naya, J. Yamato, and K. Kogure, "Differences in effect of robot and screen agent recommendations on human decision-making," *Intl. J. Hum.-Comput. Stud.*, Vol. 62, No. 2, pp. 267–279, 2005.
- [10] A. Powers, S. Kiesler, S. Fussell, and C. Torrey, "Comparing a Computer Agent with a Humanoid Robot," ACM/IEEE HRI, 2007.
- [11] C. D. Kidd and C. Breazeal, "Designing a sociable robot system for weight maintenance," IEEE CCNC, 2006.
- [12] M. K. Lee, J. Forlizzi, P. E. Rybski, F. Crabbe, W. Chung, J. Finkle, E. Glaser, and S. Kiesler, "The Snackbot: Documenting the design of a robot for long-term human-robot interaction," ACM/IEEE HRI, 2009.
- [13] M. Rehm and E. André "Catch Me If You Can: Exploring Lying Agents in Social Settings," AAMAS, 2005.
- [14] J.-Y. Jian, A. M. Bisantz, C. G. Drury, and J. Llinas, "Foundations for an empirically determined scale of trust in automated systems," State Univ. of New York at Buffalo, Tech. Rep. A933593, 1998.
- [15] P. van Schaik and J. Ling, "Five psychometric scales for online measurement of the quality of human-computer interaction in web sites," *Intl. J. of Hum.-Comput. Interaction*, Vol. 18, No. 3, pp. 309–322, 2005.
- [16] V. Groom, J. Chen, T. Johnson, F. A. Kara, and C. Nass, "Critic, Compatriot, or Chump?: Responses to Robot Blame Attribution," ACM/IEEE HRI, 2010.