Analyzing Physiological Signals from Interactions with a Cheating Robot

Eric Ho, CPSC 290

Mentor: Alex Litoiu

Faculty Advisor: Brian Scassellati

Abstract:

A study by Litoiu and Scassellati found that a robot cheating in a game of rock-paper-scissors from a losing position to a winning position triggers a cheating detector within humans that caused greater attributions of intelligence to the robot. This study, aiming to eliminate subjectivity accompanying questionnaire data, analyzes electrodermal activity from the same participants as the previous study to bolster this claim, showing that a robot cheating in this way elicits a visceral response from humans.

Introduction:

Studying human-robot interaction, and more specifically, identifying perceived humanoid behavior in robots has huge implications in incorporating robots into everyday life. Humans believing a robot to be more human-like could lead to greater trust and connections between humans and robots, advancing the realm of social robotics.

A previous study by Short et al. discovered an increase in the amount of perceived intentionality of a cheating robot using gestures rather than purely aural declarations [2]. There were three conditions: a control where the robot played twenty rounds of rock-paper-scissors normally against a participant, one in which the robot declared the wrong result, “I win!” upon discovering that the human had won the round, and one in which the robot physically changed its gesture from a losing position to a winning position before declaring that it had won. The study found that in the case of the third behavior, human
participants were more likely to use action verbs describing the cheat, perhaps increasing the perceived intelligence and autonomy of the robot. Participants describing the second behavior, however, were more likely to describe the declaration as a malfunction and not the actions of an intelligent robot.

There are two possibilities, both supported by previous research, that could explain this phenomenon. The first is that the added motion during the cheat causes the participant to believe the robot is more intelligent (humans can derive intentionality from motion of even simple objects), and the second is that the robot triggers a cheat detector in humans, so that the cheat itself causes the associated increase in intelligence.

A study by Litoiu et al. then expanded upon the previous study attempting to identify the correct explanation [1]. It introduced three other cheat conditions apart from the robot cheating from a losing condition to a winning condition, further breaking down the effects of a robot’s behavior on a participant’s perception of its autonomy. There were four total conditions: the first in which the robot, as in the Short et al. study, cheated from a losing to a winning position, the second in which the robot cheated from a losing position up to a tying position, the third in which it cheated from a winning position to a tying position, and the fourth in which it cheated from a winning position to a losing position. Each participant played 30 rounds of rock-paper-scissors with the full-body, humanoid Nao robot, during which Nao performed two cheats according to one of the conditions described above.

After the experiment, participants answered a series of questions about the occurrence of cheating in the robot, and about the intelligence of the robot. If the motion itself indeed had caused the associations of increased agency of the robot, then one would
expect the questionnaire to have steady responses for all four cheat conditions (because all four conditions contained a gesture change during the cheat). If, however, the robot triggers a certain cheating detector within humans, one would expect stronger attributions of intelligence within the first condition in which the robot cheats to win.

The study eventually found that across conditions, human response to the robot changed, suggesting that a certain cheating detection within humans is triggered according to the condition and that the gesture switch itself, and by extension, motion in general, did not cause the perceived heightened agency in the robot. The greatest percentage of participants detecting and describing the cheat fell under the first condition where Nao cheated from a losing position to a winning position, followed by the tying from a losing position, tying from a winning position, and finally the cheating from a winning to losing position.

Litoiu et al. showed that the cheating to win condition was distinct from others, rejecting the hypothesis of motion as the cause of the attributions of intelligence in robots. However, it offers little insight into the nature of the cheating detector, which could perhaps be quite emotional or purely cognitive. Moreover, during Litoiu’s study, participants wore an Empatica watch, which detected their EDA (Electrodermal Activity), BVP, the acceleration of the watch, heart rate, and temperature. This paper focuses on analyzing the EDA, otherwise known as a skin conductance response, taken from the watch with respect to the four cheat conditions described above, hoping to delve deeper into the nature of the participants’ reactions. Analyzing EDA data allows an accuracy of response that cannot be captured by a questionnaire, reducing much of the subjectivity accompanying simply asking a question to the participant.
Motivation:

By sampling a person’s EDA at certain times, we can first of all develop a baseline level of EDA, then determine the participant’s EDA response relative to a cheat lying in one of the four conditions: cheating by the robot to win from a losing position, cheating to tie from a losing position, cheating to tie from a winning position, and cheating to lose from a winning position. A stimulus, in this case being the two cheats, by the robot out of the ordinary causing the participant excitement, distress, or eliciting some other response should therefore be expected to cause a spike in EDA, allowing us to quantify exactly when and to what degree this happens. Stronger EDA responses with respect to cheats from robots will reveal the emotional nature of the cheat detector in humans and strengthen the argument for a specific type of cheat detector in humans, offering insight to the increased attributions of intelligence to cheating robots.

Methods:

83 members of the local New Haven community were recruited for the study, 44 of whom had acceptable EDA data. Videos were eliminated if the Empatica watch was malfunctioning or the participant had too much movement in the hand wearing the watch. Movements (quantified by changes in the accelerometer) tend to affect EDA data, so one of the main concerns of this project dealt with correcting for spikes caused by movement. Of the original 21 videos in the first, 21 in the second, 20 in the third, and 21 in the fourth cheat conditions, 12, 11, 10, and 11 remained, respectively.

Given a video of each study and EDA data from the Empatica watch sampling at a rate of 4Hz, we first calibrated the start of the data with a timestamp in the video. Calibrations were set by observing a spike in the accelerometer of the Empatica watch,
which could be matched with a sharp movement of the watch in a video. Observing the start of the Empatica watch purely from viewing the video was too inaccurate, as obstacles often obstructed the watch from sight of the camera. The motion spikes were either performed in the beginning of the video as instructed by the moderator, or observed during a shift in arm position, or from the removal of the watch at the end of the video. Thus the video timestamp of the start of the Empatica watch has a margin of error of 1 second because timestamps were given to the nearest second.

Once we obtained the timestamp of the Empatica data in relation to the video, we recorded the timestamps of the two cheat conditions by visual inspection. The timestamp given coincided with the gesture change of the robot, thus measuring the participant’s response to the gesture change. The analysis of the EDA data following these two cheat occurrences for each video, therefore, establishes the degree of reaction from each participant.

Additionally, to establish a crude baseline estimate for the EDA response of the entire video, we measured EDA responses for six different “random samplings” of data points, which consisted of timestamps 30, 20, and 10 seconds before the first cheat and 10, 20, and 30 seconds after the second cheat for each video. Numbers of EDA spikes at these points not accounting for other aspects basically formed a baseline for spikes caused by movement throughout the videos lying in each cheat condition.

Eight other points in the video were recorded and their EDA responses analyzed. Four of these points were timestamps of the final two rounds before the first cheat in which the participant experienced the same result as after the robot had performed its cheat as well as the first two rounds after the second cheat also with the same result.
Basically, analysis for these four points accounted for an EDA response to a general win, loss, or draw to the robot. The other four samples consisted of the first two rounds and the last two rounds, in other words, rounds 1, 2, 29, and 30. Analysis of these rounds accounted for an EDA response relative to other round, controlling for spikes closer happening more likely in specific rounds.

We used Ledalab, a Matlab-based software specializing in analysis of skin conductance data to isolate spikes in EDA in response to the events described above with the obtained timestamps. Using a series of python scripts, we converted the CSV files downloaded from the Empatica website into text files readable by Ledalab, and created event files for each participant that coincided with the timestamps of the videos. Given EDA data and events, Ledalab outputs a text file with information for each event, among which are spikes (if they exist) in response to the event. For the purposes of this paper, spikes were counted if they measured above 0.1 microsiemens (µS) and if they occurred from 0 to 5 seconds after the initial response. According to the Electrodermal Activity Textbook 2012, a SCR response usually happens between 1 and 4 seconds after the event, but given our margin of error we use the 0 to 5 second range [3]. The textbook also confirms the 0.1 microsiemens spike height. Thus, for the two cheat events, as well as the other 14 scenarios described above, events were created for each participant, fed into the Ledalab along with the Empatica data, and spikes in response to the events were recorded.

**Results:**

We aggregated the Ledalab output across the four conditions, first calculating the percentage of cheats in each condition that had a spike immediately following, then
calculating the percentage of videos in each condition that had a spike following at least one of the two cheats. We found that 38% of the cheats have EDA spikes in condition 1, while 50%, 20%, and 36% had spikes in conditions 2, 3, and 4, respectively. In terms of percent of videos that had an EDA spike, 58%, 64%, 30%, and 45% of videos had spikes in conditions 1, 2, 3, and 4.

We saw, at first glance, that condition 2 had the highest percentage of EDA spikes, followed by condition 1, 4, and 3, in that order.

Remember, however, that movement of the Empatica watch or another source of noise could overestimate the number of spikes following a particular cheat. Thus, as described in the methods section, we attempted to correct for the overestimation by finding the spikes following random samplings of the videos. The following graph shows the percentage of spikes following samples 30, 20, and 10 seconds before the first cheat and 10, 20, and 30 seconds after the second cheat, as well as the percentage of videos containing at least one spike for the same samples.
20% of the samples have spikes and 33% of the videos have at least one spike under condition 1, as opposed to 48% and 82% for condition 2, 31% and 36% for condition 3, and 26% and 55% for condition 4. We saw that percentages for this baseline are very low in condition 1, very high in condition 2, and moderate in conditions 3 and 4, suggesting that the spikes in condition 1 are affected little by the underlying noise in the video, and suggesting the opposite for the EDA spikes in condition 2.

We quantified this deviation by providing the difference between the percentage of spikes following cheat occurrences and the percentage of spikes following random samples.
In the first condition, we saw a positive difference of 18% in spikes following the cheat minus our derived baseline along with a 25% video difference. On the other hand, conditions 2, 3, and 4 had differences of 2% and -18%, -11% and -30%, and 10% and 9%. These numbers suggested that condition 1 had significantly more spikes over the baseline following the cheats than in the other conditions. However, we failed to account for some factors in this analysis. We moved on towards the next analyses, those of accounting for round placement and round result.

To account for round placement, we sampled the first two rounds and last two rounds of each video, and we again checked for EDA spikes after these samples.
The results, while slightly lower on the percentage of spikes for condition 2, seemed similar to the previous results found in the random samples. The percentage of videos that had at least one spike in condition 2 remains the highest. Once again, the percentages for the first condition were the lowest across the board, suggesting low levels of noise in the EDA data of those participants.

We then accounted for rounds of the same result as after the robot switches gestures.
Again, the results seemed similar, and we moved on to inspect the difference between the cheat spikes and these same result rounds.

Even more so in this graph we saw the significant difference between percentages in the first condition and the others. There are 16% more spikes in the cheats than in the same
result rounds and 17% more videos with spikes, as opposed to 6% and -9% for condition 2, -1% and -10% for condition 3, and 6% and -9% for condition 4.

Overall, the data, while admittedly sparsely populated, suggested a significant trend towards spikes in EDA data after cheats in condition 1, implying that participants had emotional responses to the robot cheating.

**Discussion:**

In adjusting for the number of videos with at least one spike for a group of samples, we often received a negative number. This happened largely as a result of having either 3 or 4 samples instead of only 2 cheats, which would make it more likely to have a video with at least one spike across those samples. We kept this because more samples allow for more accurate numbers in determining the percentage of samples with spikes. Additionally, the number of videos with spikes in condition 1 adjusted for the baseline noise was still positive and significant, so while we may have overcorrected, a clear deviation between condition 1 and the others has still been highlighted.

Furthermore, because of the eliminations due to overwhelming levels of noise in videos, only around half of the participants in the previous study were able to be analyzed in this study. Admittedly, the number of participants across conditions were slim, but the data and results have been suggested repeatedly by different samplings throughout the videos.

**Conclusion:**

Litoiu et al. concluded that greater associations of intelligence with robots during a rock-paper-scissors game resulted not from the motion of changing a gesture, but from triggering a cheat detector within humans [1]. This study, using the physiological data
given by the Empatica watch worn by participants, expanded upon the previous study to confirm the trigger through analysis of participants’ EDA responses, ensuring a level of accuracy unattainable by the subjectivity of a questionnaire.

After accounting for spikes caused by excess motion or other potential sources of noise, we realize that, just as in the previous study, the participants in condition 1 had stronger emotional reactions towards the cheats. These were reflected in percent of cheats with spikes minus the baseline EDA spike numbers of the condition. Participants in condition 1 experienced spikes in EDA data for around 16-18% of samples and 17-25% of videos, after being adjusted for baseline noise levels. In contrast, participants in condition 2, while having higher percentages in initial cheat spikes than in condition 1, experienced less total spikes in EDA data as a result of having much more noise, averaging 8% of samples having accompanying spikes and negative numbers for spikes in videos. Condition 3 had negative numbers for all its adjusted percentages, and seemed not to deviate from the baseline number of spikes. Participants in condition 4 experienced spikes in EDA data for around 6-10% of the samples, but also had a negative number in terms of percentage of videos.

Overall, the data suggests that like in the previous study, participants paid close attention to when the robot cheated from a losing position to a winning position. The data indicates that participants had particularly strong emotional reactions to the robot cheating in this way, triggering the cheat detector suggested by Litoiu et al. in not only a cognitive manner but initiating a visceral reaction as well.
References:

