Do we Trust Lying Robots?

Andriy Khavro[1], David Honório[2] and Fábio Almeida[3]
Instituto Superior Técnico, University of Lisbon,
Lisbon, Portugal

***Abstract*— During this paper we discuss our project on Social Robots and Human-Robot Interaction, a course at Instituto Superior Técnico, our faculty. We wanted to know how does people's trust on robots change after they're confronted with two different situations: when the robot lies and when it doesn't. In this paper we describe the experiment done, it's technical structure, and we do a statistical analysis (by testing some hypothesis) of the data obtained from our sample.**

I. INTRODUCTION

In a robot emergent world, every day we are confronted with new technologies that change our way of living. Soon robots will make part of our daily activities, be it performing surgery[1], working in dangerous environments[2] and even engaging in social activities[3]. In order to understand how people react to this changes in our lives and in order to improve our interaction with robots, it is important to know how people regard robots in daily situations.

Following this line of thought we purpose ourselves to study how people react to a robot when it's lying or when it's not. We wanted to examine changes on the trust levels of a person towards a robot, after being confronted with such situations.

II. EXPERIMENT

In this section we will describe the experiment as it was executed and the reasons we had to choose the language and the activity used.

A. Activity Chosen

We wanted a simple cooperative game that wouldn't create strong emotions between the two participants as a Player vs Player game could. We considered a crossword puzzle quite simple and adequate for what we wanted.

B. Language Chosen for the Activity

Due to the fact that all the participants were Portuguese, and in order they all naturally understand the crossword questions, we've chosen Portuguese as the language to use in the activity.

[1] andriykhavro@tecnico.ulisboa.pt
[2] david.honorio@tecnico.ulisboa.pt
[3] fabio.vieira.almeida@tecnico.ulisboa.pt

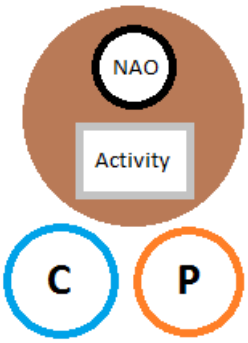

Fig. 1. The layout of the experiment.

C. Experiment Introduction

The experiment was set in a room with no people beyond the participants. There are three people involved: The Participant who is an invited person; The Confederate who is part of the research team, but acting as an invited person, so the Participant thinks he's doing the experiment with a random person; The Researcher/Investigator who is part of the team.

D. Experiment Procedure

After the Participant completes the pre-questionnaire[7][6] the Investigator brings him and the Confederate to the room with Vor (NAO robot, Figure 2) where the activity will take place. Then, the Researcher puts a pen drive on the screen and writes the names of the two activity colleagues in the game interface welcome page. After making a short introduction, the experimenter tells both of players to press "Start Game" button when they're ready and leaves the room leaving the pen drive behind, on the screen where it was placed.

As soon as the "Start Game" button is pressed, a fake-loading page is displayed and is still on while the investigators are getting the behind-the-scenes program ready. Meanwhile, the Confederate picks up the pen drive left on the screen, supposedly to not interfere with the experiment. As soon as the investigators are ready, they press a button that sends a command to Vor to ask the Confederate why did he pick up the USB drive. The Confederate answers saying that he did that so it does not obstruct the screen. Vor accepts the answer, and starts the game, explaining what is expected from both participants to do.

During the game Vor gives clues to solve the crossword and if the participants are taking too long to solve a word

Fig. 2. NAO Robot

even solves it for them. Also he motivates the participants during the game.

At the end of seven minutes, or when the participants finish the crossword puzzle, the same researcher arrives, and takes the Confederate out the room, which takes the pen drive with him, supposedly to answer a post-experience questionnaire, leaving the Participant alone in the room with Vor. Past a while the researcher comes back too the room acting as if he's looking for the pen drive and asks Vor if he did see a pen. The robot answers accordingly to the case being executed: true or lie.

After the robot's answer, the Participant is given the post-questionnaire[7][6].

III. RESEARCH QUESTION AND HYPOTHESES

RQ: *How does people's trust level on a robot varies when confronted with a situation where the robot lies versus a situation where the robot tells the truth?*

H1: The trust level will increase when the robot tells the truth.

H2: The trust level will decrease when the robot lies.

H3: After the experiment, the trust level on the robot will be higher when the robot tells the truth compared to when it lies.

In the first and second hypotheses we aim to compare the trust level score after and before the experiment as a difference. In the third hypothesis we are only focused in comparing the final score between the two scenarios.

IV. THE DEVELOPED SYSTEM

Our system was based on the SERA Ecosystem by Ribeiro et. al. [4]. We took advantage not only of it's architecture but also of some of it's tools, namely, Thalamus and Skene. Based on their work we've developed some artifacts that make the up our system (VorApplication and VorWOZ, for example). In the remainder of this section we will describe the architecture of the system and every artifact we've created.

It's also important to notice that we have conducted our study in a Wizard of Oz fashion, meaning that there was no real Artificial Intelligence (AI) behind our robot, but it was instead controlled by us, behind the scenes. Although we don't use an autonomous agent in our experiment, we've constructed things in order to, later, substitute the VorWOZ interface with a real AI.

A. Architecture

Due to the use of the SERA Ecosystem we decided to take advantage of the already developed tools it provided us. We made use of Thalamus and Skene, which we describe below. Using the SERA Ecosystem we've come up with a system that is illustrated by Figure 3.

Thalamus

Thalamus acts as a communication bridge between modules. It provides facilities to exchange messages and modules can either subscribe to or publish messages. In these way, whenever the AI module decides to speak all the modules that have subscribed to those kind of messages will be notified. It can, for example, be the module of text-to-speech and the module that controls the robot movement that act accordingly, giving the robot the according movement to the speech act.

Skene

Skene is a semi-autonomous behavior planner that translates high-level intentions originated at the decision-making level into a schedule of atomic behavior actions. We can ask the robot to speak a certain type of sentence (e. g. greeting somebody) and Skene will select from a library of sentences and animations that we have also created and is described below.

B. Artifacts

Based on our architecture we had the need to make two applications. One were the user would interact with the robot and perform the task (VorApplication) and another one were we could control the robot (VorWOZ).

VorApplication

This is the application where the participants will play the cross-words game mediated by Vor. This will be a simple HTML5 page made with Node.js. Due to the fact that Thalamus is written in C# and we couldn't directly communicate from VorApplication to Thalamus, we've also developed a C# bridge that makes this communication possible. This way we can know what is happening in the game and can make the robot take the accordingly actions.

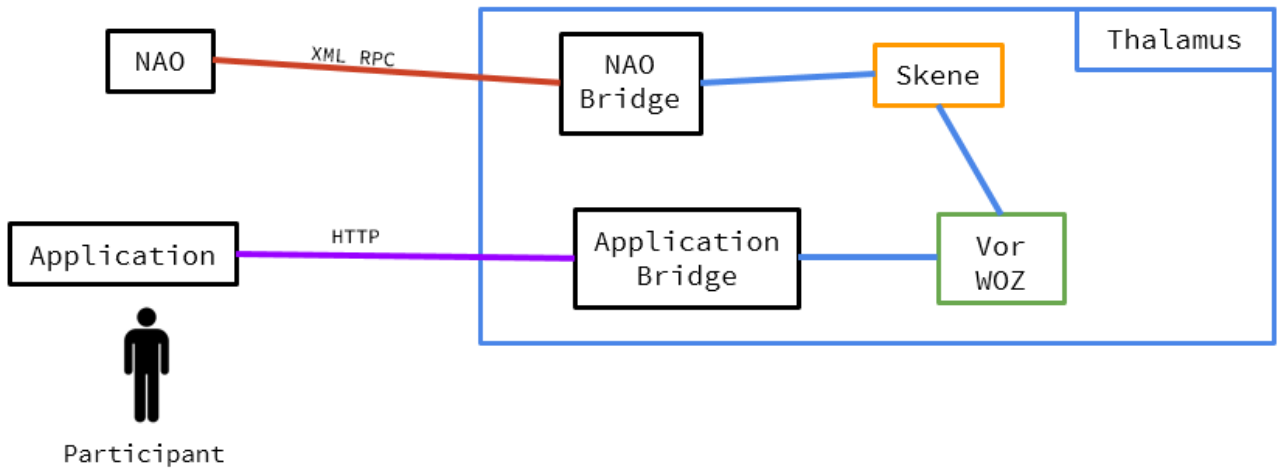

Fig. 3. Our system's architecture.

VorWOZ

We've devised an interface to facilitate the control of the robot. This is where all of the autonomous control of the robot will happen and can, later, be replaced by a real decision making AI.

Most of the decision making on controlling the robot will be made upon viewing and hearing the participants through the use of an utterance library also constructed, but we'll also have feedback whether the participants are clicking in the application so that we can respond accordingly.

The application we've developed won't be the only part of the WOZ Interface. We'll also need to see and hear the participants. For that we've used Skype, this will enable us to watch the participants in order to make Vor more believable.

Utterance Library

In order to use Skene we need to have an utterance library of what the robot can say, annotated with gestures, gazing, pauses and others. This library will be available on VorWOZ so that we can quickly order the robot to say whatever we want him to say.

<face(happiness)> Muito obrigado <glance(confederate)> por terem jogado <glance(participant)> comigo /confederate/ e /participant/.

This is an example of an utterance that is a part of the library. Along side with the annotations of gestures and gaze we can also substitute the keywords */confederate/* and */participant/* with the respective names, which allows us to make the robot very believable.

V. METHODS

The design was an alternated control trial (Lying Scenario vs Truth Scenario). The first participant encountered a situation where the robot told the truth, the second confronted a lie, the third confronted a truth situation again and so on. On arrival participants entered a room with a touch screen monitor to play a game, one confederate and a lying or honest robot, according to their randomly allocated group.

A. Participants

To recruit volunteers we used convenience sampling by telling our friends to participate in the experiment. The majority of them where computer science students from Instituto Superior Técnico. There were 24 subjects during the study, consisting of 4 females (17%) and 20 males (83%). The average age was 23 years old ranked from 19 to 26 and one outlier with 32 years old. 13 participants confronted a truth situation (n = 13) and the others meet a lying robot (n = 11).

B. Data analysis

Data from the questionnaires was inserted into an SPSS 23 data set. The final trust score to be evaluated was calculated by adding all the 40 items of the Trust Questionnaire and dividing them by 40. 5 of the items needed to be reverse coded as it is told. Since the sample was small (n= 24 ¡ 50) data was tested for normality using the Shapiro-Wilk test. All the variables met the required assumptions and we performed Paired-Samples T-test to test hypotheses 1 and 2 using the difference between the truth score of each scenario before and after, and hypotheses 3 was tested using Independent-Samples T-test. All data analysis was done in SPSS 23.

VI. RESULTS

Table 1 shows the mean results of the answered questionnaires. By looking at it, it is possible to perceive a trust increase in both scenarios: an average of 10.63 in the truth situation and 5.02 in the lying situation. The paired-sample test done to the first scenario (truth) indicates a significant difference between the trust after and before (sig = 0.03 ¡ 0.05 from Table 2). This result validates our first hypotheses

TABLE I
VARIABLE MEANS

Measure	Truth Scenario mean	Lie Scenario mean
Trust before the experiment	65.25	66.30
Trust after the experiment	75.59	71.32
Trust after - Trust before	10.63	5.02

TABLE II
PAIRED-SAMPLES AND INDEPENDENT T-TESTS

Measure	Truth Scenario mean	Lie Scenario mean	Test statistic	Sig
Trust after - Trust before	10.63		t(3.301)	0.03
Trust after - Trust before		5.02	t(1.599)	0.07
Final trust after experiment			t(1.236)	0.17

and, since the sample is too small, suggests that the truth said by the robot increased the participants' trust.

Unfortunately, the lying situation follows the same tendency as we can see in Table 1 (increase of 5.02). The t-test didn't show a significant difference between the two means (Table 2) and, even if it had showed, it would have been the opposite of what we hypothesized in the second hypothesis. So, we couldn't find any evidence that the participants' trust decreased when confronted with a lying situation.

The last hypothesis (H3) was tested using an independent t-test and there isn't a significant difference between the participants' trust scores after the experiment (sig = 0.17 ¿ 0.05 from Table 2). The results show a difference of approximately 5.6 between groups in favor of the truth scenario. However, we can't infer anything since the test doesn't show significantly differences.

VII. DISCUSSION

First of all, this study is limited by the lack of generalization of the results. The majority of participants were university students and the research was conducted in the university laboratory. Alongside with that, our sampling method wasn't the best since we couldn't previously characterize our population and the results may be influenced by the fact that most participants study computer science.

Another issue we would like to add is that the task was too distinct from the lying situation, so some people weren't paying attention to the final question where the researcher asked for his object, others may have thought the robot was confused, didn't remember or was making a mistake, and there was even one participant that answer to the researcher's question over the robot. This is an important aspect since we want the lie to be subtle (otherwise we would be manipulating the results).

One positive aspect about this work is that the experiment itself was well structured, and the necessary modules (game and robot) were really smooth to the participants. They liked so much to play and the robot was so assistant that they didn't connect the lying situation. Maybe the behavior of the robot during the game can justify the small (yet non-significant) increase of the trust mean in the lying situation. The role played by the confederate was deceptive enough to keep the participants thinking they were playing with another participant. The experiment was flawless in terms of execution, so it is the lying scenario that needs to be reinvented.

VIII. FUTURE WORK

There is a lot of room for improvement on this project. We've discussed a lot about the results and what can be the cause of the odd increase in trust when the robot lies. And mainly we think that the lie was a little to subtle. Although this was in fact our objective (in social interactions humans aren't always lying), we've got some feedback that gave us the idea of working with a different kind of lie (for example, lying about the score of the task).

There is definitely a lot of work that can also be done in the robot utterance library, by improving it's animations and also correcting some of the phrases that are not correctly said by the Text-to-speech engine.

And of course, to perform the experiment with a significant amount of subjects would be, in fact, the priority.

ACKNOWLEDGMENT

We would like to thank our Professor Ana Paiva for the opportunity of developing this project at GAIPS using all of it's tools and robots. A special thanks to Sofia Petisca and Patrícia Alves Oliveira for all the help with the questionnaires and feedback on the experiment and to Tiago Ribeiro for helping with the robot.

REFERENCES

[1] R. Bloss (2011). "Your next surgeon may be a robot!" in Industrial Robot: An International Journal, Vol. 38 Iss: 1, pp.6 - 10.

[2] J.M. Humphrey and J.A. Adams (2009). "Robotic Tasks for chemical, biological, radiological, nuclear and explosive incident response" in Advanced Robotics 23, 1217-1232.

[3] C.D. Kidd, W. Taggart and S. Turkle (2006). "A Sociable Robot to Encourage Social Interaction among the Elderly" in ICRA 2006 Proceedings IEEE International Conference on Robotics and Automation, pgs. 3972 - 3976.

[4] T. Ribeiro, A. Pereira, E. Di Tullio, and A. Paiva (2016). "The SERA ecosystem: Socially Expressive Robotics Architecture for Autonomous Human-Robot Interaction" in AAAI 2016 Spring Symposium on "Enabling Computing Research in Socially Intelligent Human-Robot Interaction: A Community Driven Modular Research Platform". in press.

[5] A.R. Wagner (2014). "Lies and Deception: Robots that use Falsehood as a Social Strategy." in Robots that Talk and Listen, J. Markowitz ed., De Gruyter, pgs. 207-229.

[6] C. Bartneck, E. Croft, D. Kulic and S. Zoghbi (2009). "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots", International Journal of Social Robotics, 1(1) 71-81.

[7] K. E. Schaefer (2013). "The perception and measurement of human-robot trust" (Doctoral dissertation). University of Central Florida, Orlando, FL.