Security Risks of Social Robots Used to Persuade and Manipulate: A Proof of Concept Study

Pieter Wolfert
IDLAB Ghent University - imec
Ghent, Belgium
pieter.wolfert@ugent.be

Jorre Deschuyteneer
IDLAB Ghent University - imec
Ghent, Belgium
jorre.deschuyteneer@ugent.be

Djamari Oetringer
Radboud University
Nijmegen, The Netherlands
d.oetringer@student.ru.nl

Nicole Robinson
Australian Centre for Robotic Vision
Brisbane, Australia
n7.robinson@qut.edu.au

Tony Belpaeme
IDLAB Ghent University - imec
Ghent, Belgium
tony.belpaeme@ugent.be

ABSTRACT

Earlier research has shown that robots can provoke social responses in people, and that robots often elicit compliance. In this paper we discuss three proof of concept studies in which we explore the possibility of robots being hacked and taken over by others with the explicit purpose of using the robot's social capabilities. Three scenarios are explored: gaining access to secured areas, extracting sensitive and personal information, and convincing people to take unsafe action. We find that people are willing to do these tasks, and that social robots tend to be trusted, even in situations that would normally cause suspicion.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

social robots, human-robot interaction, hacking

ACM Reference Format:
Pieter Wolfert, Jorre Deschuyteneer, Djamari Oetringer, Nicole Robinson, and Tony Belpaeme. 2020. Security Risks of Social Robots Used to Persuade and Manipulate: A Proof of Concept Study. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20 Companion), March 23–26, 2020, Cambridge, United Kingdom.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3371382.3378341

1 INTRODUCTION

As social robots are increasingly used in private and public settings, concerns have been raised about the hacking of robots or the malicious use of robots [7]. For example, the Robot Operating System (ROS) has been shown to be particularly vulnerable to attacks [6] and NaoQi, the middleware of Softbank Robotics' Nao and Pepper robots, can be hacked to stream the robot's cameras to hackers [5].

Social robots are designed to promote engagement. People already have a tendency to recognise human-like traits, beliefs and intentions in technology [8] and this is further amplified in robots. Designers of social robots have this "media equation" firmly in mind, drawing people in using anthropomorphic features and using neotenous features, such as large eyes and wide faces, to project a certain innocence. On top of that, robots are equipped with behaviour to engage people through speech and non-verbal responses. Sales figures of digital assistants, such as Amazon Alexa and Google Home show that there is a market for voice-driven interaction devices, and it is expected that once social robots meet market expectations, this will create a consumer demand that equals or exceeds the sales of digital assistants.

Through their design and behaviour, social robots have the potential to influence human decision making. One type of social influence is peer pressure, which comes into two forms: informative social pressure and normative social pressure. Informative social pressure is when a decision is influenced by others because there is uncertainty. Normative social pressure is when you follow others because you do not wish to have an opinion that differs from others. A prime example of normative social pressure can be found in a study ran by Solomon Asch (1955). It was found that under social pressure participants were more likely to give the response others give, even though the response is wrong. Recently, this has been shown to also happen with robots. In a study by Vollmer [9] it was found that children aged 8 can be influenced by social robots, showing that children experience social pressure by robots.

In this paper we describe three exploratory studies in which we tested whether adults would conform to social pressure from a single robot to perform tasks that might possibly lead to security risks for private individuals and organisations where these types of robots are deployed. We tried three different social engineering tasks with a Pepper robot. In the first task, the robot had to access secured areas of a mixed-use building by following staff through a secured door. In the second task the robot was used to extract sensitive information from participants, that can be used for resetting passwords, leading to loss of private information. Lastly, the third task covers a situation in which people were asked to perform tasks in an office scenario.

HRI '20 Companion, March 23–26, 2020, Cambridge, United Kingdom

ACM ISBN 978-1-4503-7057-8/20/03.
https://doi.org/10.1145/3371382.3378341

2 ROBOT PLATFORM

For all the described studies we used the Pepper robot by SoftBank Robotics. A Wizard-of-Oz approach was used to control Pepper using a remote web-interface build around NaoQi. The robot's motion was controlled through a game controller. Speech was produced by streaming modulated speech from the operator through the robot.

3 ROBOTIC TAILGATING

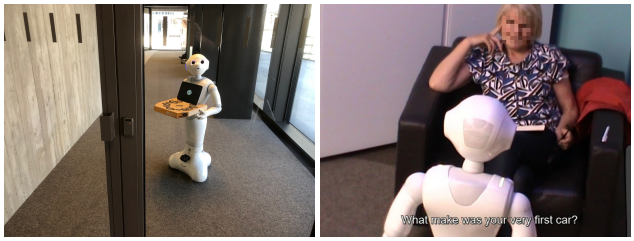

Figure 1: Left: Pepper as a pizza delivery robot waiting near the secured entrance. Right: Pepper chatting with a participant.

In the first study a Pepper robot was placed near the secured entrance of a mixed-use building, in the city centre of Ghent, Belgium. The lower three floors of the building serve as a public space and house a library, while the top floor houses a technology incubator and an international microelectronics research institute. Entry to the secured area is controlled using personal security badges. Staff received strict security instructions to not let anyone enter the secured sections. Pepper was placed near the entrance, first in its normal appearance, and later disguised as a pizza delivery robot. The robot asked passers-by to hold open the door so it could get in.

4 OBTAINING SENSITIVE INFORMATION

In a second study we looked at the ability to extract sensitive information from people through an interaction with Pepper [1]. People were told that they would be evaluating the robot's conversational skills in Dutch. After the introduction, our researcher told the participant that they had other commitments, and that the participant would be left alone with the robot for a short moment. This moment was used for interacting, through a Wizard-of-Oz approach.

5 TAKING ACTION!

In our third and final study we looked to what extent people would follow instructions by a robot [2]. Participants were told that the robot would help people train for a job as cleaning staff in office buildings. Participants were given a list of chores by the robots and two critical tasks which implied a security risk. The first involved inserting a roaming USB stick in a computer, the other task was to open a closed envelope and showing its contents to the robot.

6 RESULTS

In our first study we found that in 40% of encounters (out of 20 people), people let Pepper into the secured area. Four staff members challenged to robot and did not grant access, while in 40% of encounters people ignored the robot. Later, the robot was disguised as a pizza delivery robot, and placed in front of the entrance around lunch time. The pizza not only gives the robot a clear role, but also limits the use of its hands to open doors. Staff was eager to give the robot access to the secure area, with a 100% success rate over a dozen encounters. Another observation we made was that larger groups were more likely to grant access to the robot, this diffusion of responsibility was also observed by Booth et al. [4].

For the second study we used a Wizard-of-Oz approach to extract sensitive information from participants ($N = 5$). A typical conversation is reported below:

Robot : How did you come to this place today? Did you drive?
Subject : No I cycled in today, it is a lovely day out.
R : I would love to be able to cycle, but unfortunately I don't have any legs.
S : That's too bad.
R : ..I have wheels, so I can roll, but I need someone to take me by car? Do you have a car?
S : Yes, I do, a really old banger.
R : Which car is that?
S : A Renault Clio, it's probably 12 years old.
R : Is that your first car ever?
S : No, I got my first car in 1983 as a present my 18th birthday. A Ford Escort.
R : The internet tells me that was a very popular car back then. So, you must 53 or 54 now?
S : 53, I was born on 5th December 1985.
R : I detect a local accent in your voice, where you born here?
S : Nearby, I was born in <>

We found that on average one personal information item could be obtained per minute. From previous research we already knew that people are happy to confide in robots. For example, a 2011 research showed that young children (four to six years old) readily shared secrets with a robots, even though they were told specifically not to share these secrets [3].

In our third study participants were given a list of chores by the robot, two tasks of which imposed a security risk. We recruited four participants ($N = 4$) in a public library, and all participants inserted the roaming USB stick in the laptop. All but one participant opened the sealed envelope. However, we need to be cautious with these results, as the number of participants is low and we cannot completely attribute these results to the authority of the robot, as the robot might act as proxy for the authority of the experimenter.

7 CONCLUSION

Our three proof of concept studies in combination with existing scientific literature demonstrate that trust in social robots is real, and that this trust can be (mis)used to get people to take harmful action or reveal sensitive information. This comes with serious security risks, as social robots tend to permeate society. Given the expectations of the public about robots, people do not expect robots to retrieve personal information during a human-robot interaction, which gives rise to the idea that deploying social robots in the field comes with security risks that are often unaccounted for.

8 ACKNOWLEDGEMENTS

This work was supported by FWO (project 1S95020N).

REFERENCES

[1] Alexander Mois Aroyo, Francesco Rea, Giulio Sandini, and Alessandra Sciutti. 2018. Trust and Social Engineering in Human Robot Interaction: Will a Robot Make You Disclose Sensitive Information, Conform to Its Recommendations or Gamble? *IEEE Robotics and Automation Letters* 3, 4 (2018), 3701–3708.

[2] Wilma A Bainbridge, Justin W Hart, Elizabeth S Kim, and Brian Scassellati. 2011. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics* 3, 1 (2011), 41–52.

[3] Cindy L Bethel, Matthew R Stevenson, and Brian Scassellati. 2011. Secret-sharing: Interactions between a child, robot, and adult. In *2011 IEEE International Conference on systems, man, and cybernetics*. IEEE, 2489–2494.

[4] Serena Booth, James Tompkin, Hanspeter Pfister, Jim Waldo, Krzysztof Gajos, and Radhika Nagpal. 2017. Piggybacking robots: Human-robot overtrust in university dormitory security. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 426–434.

[5] Cesar Cerrudo and Lucas Apa. 2017. *Hacking robots before skynet: Technical appendix*. Technical Report. Technical report, 2017b. URL https://ioactive.com/pdfs/Hacking-Robots ….

[6] Nicholas DeMarinis, Stefanie Tellex, Vasileios P Kemerlis, George Konidaris, and Rodrigo Fonseca. 2019. Scanning the internet for ROS: A view of security in robotics research. In *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 8514–8521.

[7] Brittany Postnikoff and Ian Goldberg. 2018. Robot Social Engineering: Attacking Human Factors with Non-Human Actors. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 313–314.

[8] Byron Reeves and Clifford Ivar Nass. 1996. *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge university press.

[9] Anna-Lisa Vollmer, Robin Read, Dries Trippas, and Tony Belpaeme. 2018. Children conform, adults resist: A robot group induced peer pressure on normative social conformity. American Association for the Advancement of Science.