Robot Social Engineering

Attacking Human Factors with Non-Human Actors

Brittany Postnikoff
University of Waterloo
Waterloo, Ontario, Canada
bpostnik@uwaterloo.ca

Ian Goldberg
University of Waterloo
Waterloo, Ontario, Canada
iang@uwaterloo.ca

ABSTRACT

Social robots may make use of social abilities such as persuasion, commanding obedience, and lying. Meanwhile, the field of computer security and privacy has shown that these interpersonal skills can be applied by humans to perform *social engineering attacks*. Social engineering attacks are the deliberate application of manipulative social skills by an individual in an attempt to achieve a goal by convincing others to do or say things that may or may not be in their best interests. In our work we argue that robot social engineering attacks are already possible and that defenses should be developed to protect against these attacks. We do this by defining what a *robot social engineer* is, outlining how previous research has demonstrated robot social engineering, and discussing the risks that can accompany robot social engineering attacks.

CCS CONCEPTS

• **Security and privacy → Social aspects of security and privacy**; • **Human-centered computing → Social engineering (social sciences)**; • **Computer systems organization → Robotics**;

KEYWORDS

human-robot interaction, social engineering, computer security, privacy

ACM Reference Format:
Brittany Postnikoff and Ian Goldberg. 2018. Robot Social Engineering: Attacking Human Factors with Non-Human Actors. In *HRI '18 Companion: 2018 ACM/IEEE International Conference on Human-Robot Interaction Companion, March 5–8, 2018, Chicago, IL, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3173386.3176908

1 INTRODUCTION AND BACKGROUND

Previous research involving human-robot interaction (HRI) and computer security and privacy has focused on finding hardware and software vulnerabilities in robots, and determining how those vulnerabilities can be exploited [2, 7]. These works have demonstrated that malicious entities can spy on and otherwise endanger people in the vicinity of robots that they have compromised. Additionally, researchers have mentioned robot use of social abilities as a vector for privacy or security risks [7], but no explicit connection had been made between HRI and social engineering — the part of computer security and privacy that targets humans and uses social interaction to circumvent security systems. In this work, we make this explicit connection.

Social engineering has many different definitions, often with extraneous or competing components [10, 12, 13]. For our purposes, we have combined several definitions to define social engineering more generally as "the process of using interpersonal skills and social constructs to manipulate an individual, or set of individuals, into doing or saying something that may or may not be in their best interests". An example of social engineering is a salesperson that uses their authority as an expert to persuade a family into buying an unnecessary extended warranty on a product or service. When a social engineer deliberately uses their manipulative social skills to achieve a set of specific goals, such as acquiring specific digital, physical, or informational resources, they are said to be performing a social engineering attack.

Social engineering attacks make use of social constructs such as impersonation, authority, lying, and so on [10, 12], many of which are also available to social robots. For example, Short et al. show the impact of robot cheating [15], and Cormier et al. illustrate that robots can wield authority [5]. In our work, we use these parallels between the social abilities of robots and human social engineers to create the first definition for *robot social engineering* and detail several of the associated risks.

2 ROBOT SOCIAL ENGINEERING

We define a robot social engineer to be a social robot with physical agency and any level of autonomy that "[uses] interpersonal skills and social constructs to manipulate an individual, or set of individuals, into doing or saying something that may or may not be in their best interests", as per the definition of social engineering given above. If a robot involved in a social engineering attack does not play an active role requiring use of its social abilities, it would not be labelled as a robot social engineer. When a robot uses social engineering to achieve a goal it is performing a *robot social engineering attack.*

2.1 Autonomy and Robot Social Engineers

We argue that a robot social engineer can have any level of autonomy in order to account for all styles of human-robot interaction. To elaborate, we posit that robot social engineers could be completely human-controlled, human-programmed, self-taught, or any other possibility in between. As of now, human-controlled, human-programmed, and mixed-controlled robot social engineers are feasible as researchers have been able to use these methods of

HRI '18 Companion, March 5–8, 2018, Chicago, IL, USA

ACM ISBN 978-1-4503-5615-2/18/03…$15.00
https://doi.org/10.1145/3173386.3176908

control to equip robots with social abilities [9]. Self-taught robot social engineers are not currently possible but we include them in order to account for future potentialities.

2.2 Embodiment of Robot Social Engineers

We suggest that the embodiment of robot social engineers can vary significantly as previous research has shown that many different robots can make use of skills used in social engineering [5, 14, 15], and not all social engineering skills must be used in every social engineering attack [13]. Saying this, it is clear that some robots will be more successful at wielding certain social engineering techniques in comparison to other robots. For example, a large humanoid robot such as Baxter would likely be better able to demonstrate and use authority over humans working on a production line than a Roomba would [1], but a familiar and trusted Roomba would likely be better able to covertly map a person's home and exfiltrate that data than a Baxter in a similar situation [8, 11].

3 ROBOT SOCIAL ENGINEERING SCENARIOS AND RISKS

Booth et al. provide the clearest proof of concept for robot social engineering [3], but do not explicitly mention social engineering or connections to security and privacy research as part of their work. In this paper, a robot waits outside a locked university dorm and asks people going in and out of the building to hold the door open for it so it can enter the restricted space. This is referred to as 'piggybacking' in both the paper and social engineering literature. While the work states unauthorized entities entering a restricted space is an unwanted outcome, and participants suggest that the robot could be a bomb, we can imagine other security and privacy risks in this scenario. For example, once the robot has social engineered its way into the building, the robot could spy on students, steal items, collect private information, and so on.

Similarly, Denning et al. mention a series of "psychological" attacks that could be performed by robots [7], including a malicious entity compromising a robot and exploiting the social bonds between the robot and those that interact with it. These psychological attacks are examples of social engineering attacks as they exploit social bonds to cause individuals to experience mental and emotional breakdowns, which are not something people typically want to do. The risk with a bullying robot social engineer is that it could use its tactics to persuade someone into releasing private information about another person or other entity, or performing actions on behalf of the robot social engineer.

In addition to these examples, we can imagine scenarios where social robots turned robot social engineer could study the behavioural patterns of people from inside their own homes, masquerade as charity workers in public spaces and request personal information from passersby, and impersonate delivery robots by adopting their company markings as a way to intercept packages and goods.

4 CONCLUSIONS AND FUTURE WORK

Our work bridges the gap between social robots and social engineering by explicitly linking the two fields and then defining robot social engineering for the first time.

As future work, we will implement user studies that demonstrate both social engineering attacks and defenses against these attacks. In particular, we plan to perform experiments that mirror existing real-world social engineering attacks. For example, we will use robots of varying abilities and embodiments to perform piggybacking attacks in various contexts, impersonation attacks where the robots pretend to be from credible companies and groups in order to collect private information, and so on. As follow up, we will directly compare the effectiveness of human and robot social engineers in the same scenario by implementing between-participant user studies. Finally, we will test whether existing techniques, such as awareness training and practice drills, used in the field of security and privacy to combat human social engineering attacks are effective against robot social engineers as well.

5 ACKNOWLEDGMENTS

We gratefully acknowledge the support of NSERC for grants RGPIN-03858 and STPGP-463324-14.

REFERENCES

[1] Amy Banh, Daniel J Rea, James E Young, and Ehud Sharlin. 2015. Inspector Baxter: The Social Aspects of Integrating a Robot as a Quality Inspector in an Assembly Line. In *Proceedings of the 3rd International Conference on Human-Agent Interaction*. ACM, 19–26.

[2] Tamara Bonaci, Jeffrey Herron, Tariq Yusuf, Junjie Yan, Tadayoshi Kohno, and Howard Jay Chizeck. 2015. To make a robot secure: An experimental analysis of cyber security threats against teleoperated surgical robots. *arXiv preprint arXiv:1504.04339* (2015).

[3] 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.

[4] Sonja Caraian, Nathan Kirchner, and Peter Colborne-Veel. 2015. Moderating a Robot's Ability to Influence People Through its Level of Sociocontextual Interactivity. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 149–156.

[5] Derek Cormier, Gem Newman, Masayuki Nakane, James E Young, and Stephane Durocher. 2013. Would you do as a robot commands? an obedience study for human-robot interaction. In *International Conference on Human-Agent Interaction*.

[6] Kate Darling. 2015. 'Who's Johnny?'Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy. (2015).

[7] Tamara Denning, Cynthia Matuszek, Karl Koscher, Joshua R Smith, and Tadayoshi Kohno. 2009. A spotlight on security and privacy risks with future household robots: attacks and lessons. In *Proceedings of the 11th international conference on Ubiquitous computing*. ACM, 105–114.

[8] Jodi Forlizzi. 2007. How robotic products become social products: an ethnographic study of cleaning in the home. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*. ACM, 129–136.

[9] Michael A Goodrich, Jacob W Crandall, and Emilia Barakova. 2013. Teleoperation and beyond for assistive humanoid robots. *Reviews of Human factors and ergonomics* 9, 1 (2013), 175–226.

[10] Christopher Hadnagy. 2010. *Social engineering: The art of human hacking*. John Wiley & Sons.

[11] Alex Hern. 2017. Roomba maker may share maps of users' homes with Google, Amazon or Apple. https://www.theguardian.com/technology/2017/jul/25/roomba-maker-could-share-maps-users-homes-google-amazon-apple-irobot-robot-vacuum. (2017). Accessed: 2017-10-25.

[12] Kevin D Mitnick and William L Simon. 2011. *The art of deception: Controlling the human element of security*. John Wiley & Sons.

[13] Francois Mouton, Louise Leenen, and Hein S Venter. 2016. Social engineering attack examples, templates and scenarios. *Computers & Security* 59 (2016), 186–209.

[14] Eduardo Benítez Sandoval, Jürgen Brandstetter, and Christoph Bartneck. 2016. Can a robot bribe a human? The measurement of the negative side of reciprocity in human robot interaction. In *Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on*. IEEE, 117–124.

[15] Elaine Short, Justin Hart, Michelle Vu, and Brian Scassellati. 2010. No fair!! an interaction with a cheating robot. In *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*. IEEE, 219–226.