Robot Social Engineering

by

Brittany Postnikoff

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2020

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This thesis establishes the new field of Robot Social Engineering. We define Robot Social Engineering as the use of social abilities and techniques by robots to manipulate others in order to achieve a goal. We build the field of Robot Social Engineering on the foundations of Human-Robot Interaction research on social robots as well as information security research on social engineering.

Social engineering attacks are the deliberate applications of manipulative social skills by an individual in an attempt to achieve a goal. The field of information security has shown that interpersonal skills, such as trust, authority, and empathy, can be used by humans against other humans to perform social engineering attacks. Meanwhile, the field of Human-Robot Interaction has shown that robots may make use of social abilities such as those used in social engineering attacks.

In this work, we argue that Robot Social Engineering attacks are possible and have already been performed, and that defenses should be developed to protect against these attacks. We support our argument by providing background on social robots and social engineering, defining Robot Social Engineering, outlining previous research that we consider to be Robot Social Engineering, describing Robot Social Engineering attacks that may be performed, and discussing the risks that can accompany Robot Social Engineering attacks.

Acknowledgements

I would like to thank my supervisor, Dr. Ian Goldberg of the Cryptography, Security, and Privacy (CrySP) Laboratory at the University of Waterloo. I truly appreciate his extreme patience with me as I focused on everything except my thesis during the time I should have been writing. Without Ian's support, as well as his advice and mentorship, this thesis would have taken even longer to complete.

I also thank my committee members, Jennifer Whitson and Jesse Hoey, for reading my thesis and providing thoughtful questions and valuable feedback.

Thank you to Jacky Baltes, Chris Iverach-Brereton, and Amir Hosseinmemar in the Autonomous Agents Laboratory at the University of Manitoba who introduced me to my first robots, the Snobots: Jose, Jennifer, and Jimmy. I will always remember trudging across multiple countries, through airports, and over multiple terrains with Chris Iverach-Brereton and the Snobots.

Additionally, I would like to thank James E. Young, Andrea Bunt, Daniel J. Rea, Stela H. Seo, Patrick Dubois, Volodymyr Dziubak, and the other members of the Human-Computer Interaction Laboratory at the University of Manitoba who got me hooked on thinking about how humans interact with technology. My experience with social robots in this lab led to my first thoughts, experiences, and experiments related to Robot Social Engineering.

Thank you to the members of the CrySP lab for listening to and watching many of my presentations whenever they could, even when they had heard them before. An additional thank you goes to Anna Lorimer and Bailey Kacsmar for their help reviewing this thesis.

A special thank you to my neighbours who provided me with hours of intense questioning, proofreading, and futuristic and outlandish conversations to keep me engaged with this topic. In particular, Sara-Jayne Terp, Wendy Knox Everette, and Stephen Toulouse were key in bringing my thoughts on Robot Social Engineering to maturity.

Thank you to Michael Ossmann for being the best support I could have ever asked for. His dedication and commitment to ensuring that I completed my degree was invaluable and helped me truly appreciate what I have written here.

Finally I would like to thank all those that have supported, helped, and motivated me during my time writing this thesis. The many conversations with current friends and future friends in conference hallways, online communities, and while travelling the world gave me the drive and inspiration to keep on writing.

♥: CM, BH, BM, BO, KD, FR, JS, ZP, ...

Table of Contents

Chapter 1

Introduction

1.1 Motivation

Robots today occupy many different roles in both public and private spaces. Several of the roles that robots may have, including salesperson [Fra14], food delivery courier [LH17], and hotel concierge [Raj15], were only available to robots in science fiction, and not robots in our physical world, until recently. In these positions, it is common for robots to directly interact with humans, and a critical part of their success is their ability to communicate clearly and work collaboratively with these humans. For example, a food delivery robot would not be very successful if it continually stayed just out of reach when someone was attempting to accept the package of food that it came to deliver.

Human-Robot Interaction researchers have found that robots can be more effective in their interactions with humans when they are able to interpret and use interaction norms in ways that humans would expect. The robots that follow and use these interaction patterns are referred to as social robots [Dar16]. As an example of the abilities of social robots, consider a humanoid robot that is participating in a conversation between multiple parties. The robot can turn its face towards the person it wishes to address, and that person is likely to notice that the robot is trying to get their attention [IKO+02; MKF+12]. To further understand how humans interact with social robots, research effort is being put into developing and studying robot mastery of these social skills and others such as trust [RHW17], authority [CNN+13], and empathy [SGN+15].

In the field of information security, Hadnagy [Had10], Mitnick and Simon [MS11], and others [CNN19; DEF15] have shown that trust, authority, and empathy can be used by malicious entities to perform social engineering attacks. Social engineering attacks can be described as

the deliberate application of social mechanisms to a situation in an attempt to convince others to do or say things that may or may not be in their best interests [Had10; MS11]. As an example of such an attack, consider a malicious entity attempting to gain access to a locked building in order to steal specific company assets. This malicious entity could use impersonation, empathy, and lying to convince an employee with a key to the building that they are a new hire that is late for work and that they have forgotten their key. The result of this interaction could be the employee believing the malicious entity and letting them into the locked building where they could then access the targeted company assets. The impacts of this specific social engineering attack could be mass loss of data [MS11], money and investments [Wat19], physical assets [DEF15], or other resources [CNN19]. Social Engineering attacks may also result in loss of personal privacy, damage to personal or public property, and so on [CNN19; Had10; MS11]. In this thesis, we combine the concepts of social robots and social engineering to define and explore the field of Robot Social Engineering.

Much of the previous research done in the overlap between Human-Robot Interaction and information security has failed to explicitly mention social engineering. Only works by Aroyo et al. [ARS17; ARSS18] and our own contributions in Postnikoff and Goldberg [PG18] directly discuss both Human-Robot Interaction and social engineering. Aroyo et al. restrict their study of robot ability to perform social engineering attacks to the impacts of trust and overtrust towards robots and Postnikoff and Goldberg provide an initial discussion of Robot Social Engineering but they do not fully elaborate on the unique aspects of Robot Social Engineering. Other research efforts in the overlap of Human-Robot Interaction and information security primarily focus on finding and exploiting hardware and software vulnerabilities [QPP+17], or they briefly touch on the ability for robots to perform "psychological attacks" [DMK+09] and do not actively examine their research through the lens of social engineering. As such, the potential for robots to perform social engineering attacks, and the impacts of such actions, have gone relatively unexplored at this time. We feel that, if left unchecked, Robot Social Engineering could pose a significant threat to the privacy, security, and safety of those that interact with social robots. Additionally, we believe that the environments in which Robot Social Engineering attacks occur are also at risk for damage or exploitation, depending on the actions and focus of the Robot Social Engineer.

1.2 Contributions

This work explores the risks of robot use of social engineering skills by defining Robot Social Engineering and exploring its application to past, present, and future situations involving social robots. As we explore the ability for robots to social engineer, we seek to prove the following thesis statement:

The fields of information security and Human-Robot Interaction can be bridged by defining the new research area of Robot Social Engineering. We can demonstrate that the skills of social robots can be and have been used to perform social engineering attacks.

In validating our statement, we provide the following contributions:

1. **Definitions of Robot Social Engineering and Robot Social Engineer**: We explicitly bridge the gap between the fields of information security and Human-Robot Interaction by defining Robot Social Engineering, a novel research area. Additionally, we define what it means for a robot to be labelled as a Robot Social Engineer, and we provide clarifying details as to when the term is applicable to a given robot.
2. **A catalogue of existing Robot Social Engineering attacks**: To show that robots have used social skills to perform social engineering attacks, we catalogue existing instances of Robot Social Engineering attacks by applying our definition of Robot Social Engineering to previous Human-Robot Interaction works.
3. **Descriptions of potential near-future Robot Social Engineering attacks**: We build on our catalogue of existing Robot Social Engineering attacks by describing novel Robot Social Engineering attacks that could be executed today given current technologies. Additionally, we outline Robot Social Engineering attacks that we may see in the future as technology and techniques improve.

1.3 Organization

The organization of this thesis is as follows: Chapters 2 and 3 contain the background information on social robots and social engineering that our definition for Robot Social Engineering builds on. In Chapter 4, we define Robot Social Engineering, desribe when the term is applicable to a given robot and situation, and compare Robot Social Engineering to social engineering as performed by a human in order to highlight the unique aspects of Robot Social Engineering. In Chapter 5, we provide examples of Robot Social Engineering attacks by reframing previous Human-Robot Interaction works and describing novel potential near-future Robot Social Engineering attacks. Lastly, we impart our concluding remarks in Chapter 6.

Chapter 2

Background: Social Robots

Robot Social Engineering, as we will define it in Chapter 4, builds on research from the field of Human-Robot Interaction. Human-Robot Interaction is a discipline that studies the interplay between humans and robots by looking at how various interfaces, controls, and interaction norms can affect this relationship. Part of Human-Robot Interaction research is focused on social robots, and it is this work that is foundational to our work on Robot Social Engineering. In this chapter, we detail what social robots are, why people are inclined to treat robots socially, and examples of social skills that robots are able to make use of.

2.1 Definition of Robot

Before we can talk about what a social robot is we must first talk about what a robot is. The term *robot* has no single agreed-upon definition. Papers on robot-related research may define *robot* as any of the following:

- "a cyberphysical system with sensors, actuators, and mobility" [DMK+09],
- "a machine, situated in the world, that senses, thinks, and acts" [Bek12],
- "a programmable artefact, able to respond to its environment, capable of movement and changing aspects of its physical environment" [MS14], or
- a robot is "a machine that (a) is designed to work with individuals and groups in their personal and public spaces, (b) has a dynamic spatial presence in those spaces, and (c) can 'intelligently' interpret its environment and interact physically with it" [YHSI09].

The definitions for *robot* often come with clarifications such as:

- "[this definition] does not require robots to resemble humans, to be mobile, or to communicate using natural language" [YHSI09],
- "robots can move autonomously and they have manipulators that can affect the physical world" [FK11],
- "a robot must have sensors, processing ability that emulates some aspects of cognition, and actuators" [Bek12], and
- "This definition excludes a class of cyber-physical systems whose environmental actuators are strictly electronic, such as an oven with electronically controlled heating elements" [DMK+09].

For the purposes of this thesis, we combine aspects of the above robot definitions and clarifications to define *robot* as follows:

> A *robot* is a machine with some level of autonomy that has sensors and actuators, is capable of movement, and is able to respond to and interact physically with its environment.

We examine the components of this definition in the following subsections on robot autonomy, robot physicality, and robot movement and mobility.

2.1.1 Robot Autonomy

Robot autonomy, as defined by Beer et al. in their paper on robot autonomy, is:

> "The extent to which a robot can **sense** its environment, **plan** based on that environment, and **act** upon that environment with the intent of reaching some **task-specific goal** (either given to or created by the robot) without external **control**" [BFR14].

In the definitions for *robot* mentioned in Section 2.1, it is stated both explicitly and implicitly that robots require some level of autonomy. For example, one of the definitions explicitly states that "robots can move autonomously ..." [FK11]. The definitions "a robot must have sensors, processing ability that emulates some aspects of cognition, and actuators" [Bek12], and "a machine

. . . that senses, thinks, and acts" [Bek12] implicitly state the need for robots to have some level of autonomy as they both mention sensing, planning, and performing actions as mentioned in the definition of autonomy from Beer et al. Autonomy is an important piece of the definition for *robot* as it distinguishes robots from devices and machines such as vending machines, CNC machines, or household appliances that take prescriptive commands from humans in order to take action.

We state in our definition for *robot* that "some level of autonomy" is required as robot autonomy is not binary; a robot can have a variable amount of autonomy depending on how much third-party intervention is involved in the robot's operation. Third-party intervention can be provided by humans, programs that are not part of the robot's normal operation, or any other source of instruction external to the robot. Robots that use a combination of autonomy and third-party control as part of their operation are referred to as having shared autonomy [JSB15]. In other works, robots with shared autonomy may also be referred to as mixed control robots [Sma16; Wan07] or mixed-initiative robots [LGKI16]. As an example of shared autonomy, consider a class of primarily autonomous robots that only receives human assistance when it is requested, such as a vacuum robot that has finished vacuuming a designated area and is now looking for a new task. If a human operator then manually assigns the robot a new task, such as vacuuming a different room, the robot would be labelled as having shared autonomy as it has operated on its own but it has also received human intervention. On a scale from completely autonomous to fully human controlled, robots with shared autonomy would fill every point on the scale between the two extremes [SV78].

2.1.2 Robot Physicality

The definitions for *robot* listed in Section 2.1 explicitly and implictly mention the requirement for robots to have physicality. For example, the following definitions directly mention robots as being physical since they describe robots as machines and artefacts:

- "a machine that . . . can 'intelligently' interpret its environment and interact physically with it" [YHSI09],
- "a machine, situated in the world" [Bek12], and
- "a programmable artefact, able to respond to its environment, . . . capable of movement and changing aspects of its physical environment" [MS14].

The assertion that a robot must have physicality is further supported by definitions such as "a robot must have sensors, . . . and actuators" [Bek12] and "[robots are] cyberphysical [systems]

with sensors, actuators, . . . ” [DMK+09], which implicitly mention robot physicality as they state that robots should have sensors and actuators. Sensors such as cameras, microphones, and accelerometers are physical devices that allow a robot to sense or take in information about its environment and what is happening in that space. Actuators such as electric motors and hydraulic cylinders are physical components of a robot’s body that enable it to make movements, take actions, or interact with its environment. The requirement that a robot be physical is important as it distinguishes robots from software-only entities such as Artificial Agents (also known as on-screen characters) [BF04] and Artificial Intelligences (AIs), which may be represented by Microsoft’s Clippy and Apple’s Siri, as shown in Figure 2.1.

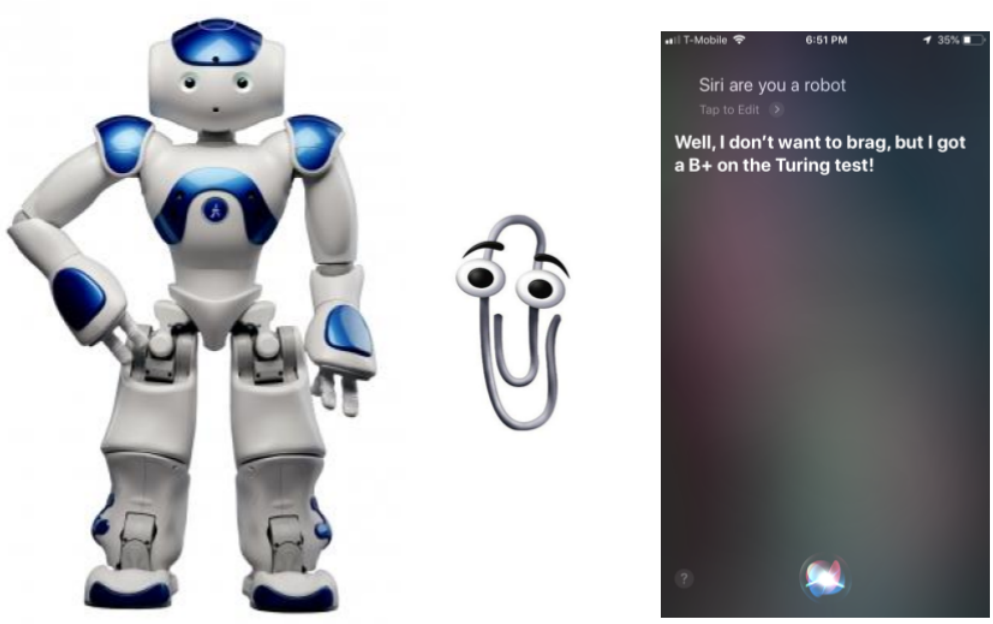

Figure 2.1: SoftBank’s Nao [Sof19], Microsoft’s Clippy [cli19], and Apple’s Siri. A robot, an artificial agent, and an artificial intelligence, respectively.

2.1.3 Robot Movement and Mobility

The definitions and clarifications for the term *robot* provided in Section 2.1 have conflicting views when it comes to robot mobility. Denning et al. define a robot as “a cyberphysical system with sensors, actuators, and mobility” [DMK+09] and they clarify that “[t]his definition excludes a class of cyber-physical systems whose environmental actuators are strictly electronic, such as an

oven with electronically controlled heating elements" [DMK+09]. Young et al. state that a robot should be able to "interpret its environment and interact physically with it" [YHSI09] and that they "[do] not require robots to . . . be mobile [YHSI09].

In terms of mobility, Denning et al. are not explicit in whether they require a robot to move from one location to the next, or whether it is enough that a robot is able to move a piece of its body in what could instead be described as movement. The definition by Young et al. is explicit in the need for physical interaction but not mobility. Additionally, McBride et al. state that a robot should be "capable of movement and changing aspects of its physical environment" [MS14]. In this thesis, we interpret *movement* to be the bodily actions a robot can take to gesture or interact with its environment, and we do not interpret *movement* to mean that a robot is able to move from its current location. Instead, we use the term *mobility* to refer to a robot's ability to move from one physical location to another. Therefore, we adopt the thought that a robot should be capable of movement as part of our definition of the term *robot*. Beatbots' Keepon [Bea03] and Rethink Robotics' Baxter [Jur19] (shown in Figure 2.2) are examples of robots that have movement, as they can move their own body to perform actions, but not mobility, as they cannot move from one place to another.

Movement distinguishes robots from devices such as tablets, laptops, digital assistants like Google home or Amazon Alexa, and children's toys [Ste06].

Figure 2.2: Beatbots' Keepon [Bea03] and Rethink Robotics' Baxter [Jur19]

2.2 Definition of Social Robot

In comparison to the term *robot*, the definitions for *social robot* are much more homogenous. Some definitions for social robot are as follows:

- "a robot that can communicate [with] and understand people" [FYDS16],
- "an autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioural norms expected by the people with whom the robot is intended to interact" [BF04], and
- "a physically embodied, autonomous agent that communicates and interacts with humans on a social level" [Dar16].

Each of these definitions for social robot directly or indirectly describes social robots as having three main components: a physical form, some level of autonomy, and the ability to interact on a social level. As such, any of the definitions provided could be used as our definition of social robot but, for the purposes of this thesis, we have adopted the definition of social robot provided by Darling [Dar16].

2.3 Treating Robots Socially

Humans often interact with robots as though they are social beings, whether the robots they are interacting with have been designed to be social or not [For07]. Additionally, people will interact with robots socially even if they say that they understand that robots are not conscious beings [SGGC07]. Some aspects of human-robot relationships that can help people interpret robots as being social actors are anthropomorphism, zoomorphism, and the Wizard-of-Oz technique.

2.3.1 Anthropomorphism

"Anthropomorphism describes the tendency to imbue the real or imagined behavior of non-human agents with humanlike characteristics, motivations, intentions, or emotions" [EWC07]. A participant in the paper "Inspector Baxter" [BRYS15] by Bahn et al. provides an example of robot anthropomorphism when they describe Baxter (shown in Figure 2.3) by saying "When it shook its head, it had a smirk on its face as if it was mocking me". As another example, participants in the paper "My Roomba is Rambo" [SGGC07] by Sung et al. anthropomorphise their Roombas by expressing the intentions, feelings, and unique characteristics of the robots in statements such as:

> "For me how they look, each one has certain different behavior. And I know definitely they have a same firmware or a similar firmware so the difference should not be much but ah, for example, my discovery, he's more crazy. He runs into things and

sometimes and goes into different places he should not be going to. And the scheduler he's more like refined. He knows what he's doing" [SGGC07].

Anthropomorphism is part of what enables humans to empathize with robots, understand and react to robot socialization attempts, and acknowledge robots as social actors. When people anthropomorphise robots they often rationalize robot actions as being intentional [YSV+11], and "the more human-like a robot is the more people expect the robot to interact in the same way as humans do" [HMW+09]. Knowing that people have the tendency to anthropomorphise robots means that robots can be designed and programmed to take advantage of or discourage the effects of anthropomorphism in order to be more effective in their communications and interactions with humans.

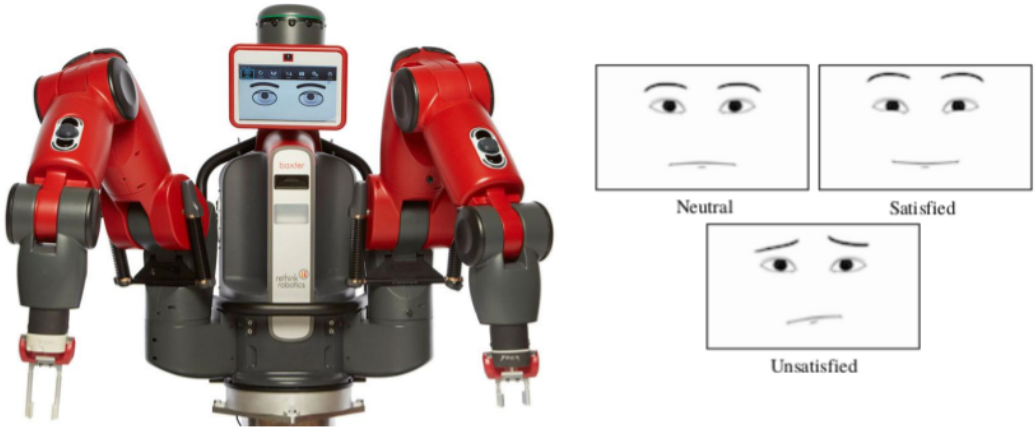

Figure 2.3: Baxter and several facial expressions used in the experiment [BRYS15].

2.3.2 Zoomorphism

Zoomorphism is the act of imposing animal-like attributes upon non-animal objects or humans. As stated by Fong et al., "[z]oomorphic embodiment is important for establishing human-creature relationships" [FND03] such as those between humans and robots that are modelled after household animals like dogs and cats [FND03]. For example, when a dog tail is put on a robot, and the tail moves in a dog-like fashion, some people zoomorphise the robot (as a dog) and believe that they can understand the feelings that the robot is trying to convey [SY13]. As another example, consider what happened when the American Army tested a robot with multiple legs, modelled after a stick-insect, that would destroy land mines by stepping on them. As the robot stepped on

a land mine it would lose the limb that triggered the land mine, and then it would continue to move forward with the limbs it had left. It was reported that the Army colonel in command of the exercise stopped the test and stated that it was inhumane [Gar07], which demonstrates that he was emotionally attached to the robot.

The impact of zoomorphism on human and robot interactions is that people empathize and interact with zoomorphised robots as if they were the animal they embody [GA17]. Understanding that people have the tendency to zoomorphise robots that behave or look animalistic means that robots can be designed and programmed to take advantage of or downplay the effects of zoomorphism in order to be more effective in their communications and interactions with humans.

2.3.3 The Wizard-of-Oz Technique

In 1984, Kelly wrote the first paper on the Wizard-of-Oz technique and defined it as "an experimental simulation in which participants are given the impression that they are interacting with a program that understands English as well as another human would. ... The experimenter surreptitiously intercepts communications between participant and program, supplying answers and new input as needed" [Kel84]. In the time since the initial definition of the Wizard-of-Oz technique was published, the technique has been adopted by the field of Human-Robot Interaction and it is has been updated to fit the technologies present in the field. In many cases, the Wizard-of-Oz technique now tries to give participants the impression that they are interacting with a fully autonomous robot, and it involves "a person remotely operating a robot, controlling any of a number of things, such as its movement, navigation, speech, gestures, etc." [Rie12]. The impact of the Wizard-of-Oz technique is that people will treat robots without much autonomy as being autonomous, even though they are being controlled by someone behind the scenes.

2.4 Robot Social Skills

When humans interact with robots as though they are social beings it makes space for the relationship to become more complex and for connections based on empathy, authority, trust, and other social skills to form.

2.4.1 Authority

In "Robot Authority and Human Obedience" by Agrawal and Williams [AW17], a robot acted as a security guard at the exit of a university building and attempted to use its authority to convince

people trying to use its exit to use an exit 30 metres away instead. In its role as a security guard the robot demonstrated authority and aggression by using a low-pitched and dominating voice, maintaining eye contact with anyone nearby, blocking the button that needed to be pressed to exit the building, and throwing its arms open when people approached the exit to physically demonstrate that it was blocking the exit. In the end, 14 out of 22 people complied with the robot's directions and used the other exit.

2.4.2 Empathetic Response

In "Poor Thing! Would You Feel Sorry for a Simulated Robot?" [SGN+15] by Seo et al., a robot garners an empathetic response from a human when, after a bout of casual conversation over a game of sudoku, it reveals it has a virus (which manifests itself as "jittery movements, speaking in distorted voice tones or stuttering, repeating words in a sentence, and speaking nonsense" [SGN+15]), is reset by a researcher, and appears to come back online with a different voice and personality. One participant said that they were sorry that the robot caught a computer virus and seems genuinely concerned [Seo15, 1:03], and another physically expresses her discomfort when the robot uses a new voice after being reset [Seo15, 1:24].

2.4.3 Conformity

In "Humans Conform to Robots: Disambiguating Trust, Truth, and Conformity" by Salomons et al. [SLSS18], a group of three robots attempts to convince a human to conform to a decision that they have made during an abstract card game. In this card game, one word and six cards are displayed on a screen and players are asked to secretly mark which card they feel best represents the word displayed on the screen. Once the participant and each of the robots has marked their card choice, the participant then has a chance to change their answer. The results of the experiment revealed that participants adjusted their answer 28.9% of the time to conform to the robots' suggested card after seeing what they had chosen. Additionally, participants reported that they felt pressured by the robots to choose the same answer, and 85% of these participants directly stated that they changed their answers because of the robots. Participants stated that they went along with the robots' suggestions because they were "on the fence and their answers made me reconsider. I assumed I wouldn't be the lone right one." [SLSS18], and "from life experiences, majority is usually correct" [SLSS18].

2.4.4 Trust

In “Overtrust of Robots in Emergency Evacuation Scenarios” by Robinette et al. [RLA+16], a robot with bad navigation ability leads participants on a circuituitous path to an experiment room, a simulated fire scenario (with smoke and blaring smoke detectors) is started, and the robot attempts to lead participants out of the emegency scenario. The results of this experiment revealed that all 26 of the participants followed the robot to an exit even though it was previously proven that the robot had bad navigation and there were exit signs indicating that the nearest exit was not in the direction the robot was leading them. 81% of those participants indicated that they followed the robot because they trusted it.

2.5 Summary

Social robots are robots that are able to interact with others on a social level. Shared autonomy, the Wizard-of-Oz technique, anthropomorphism, and zoomorphism all contribute to people interacting with robots as social beings regardless of how much agency or autonomy they have. When people treat social robots as social beings with agency it makes space for these robots to make use of social mechanisms such as authority, empathetic response, conformity, and trust, which are all social mechanisms that can be used by social engineers.

Chapter 3

Background: Social Engineering

Robot Social Engineering, as we will define it in Chapter 4, builds on social engineering research from the field of information security. Social engineering is the study of how the human elements in a scenario can be exploited to circumvent digital, physical, and procedural controls that are put into place to protect resources and information. In this chapter, we provide background information on social engineering as context for our later definition and discussion of Robot Social Engineering.

3.1 Definition of Social Engineering

The term *social engineering* has a number of different definitions. Some of the definitions for social engineering are as follows:

- "getting people to do things they wouldn't ordinarily do" [MS11],
- "a combination of techniques used to manipulate victims into divulging confidential information or performing actions that compromise security" [LBSB11],
- "influencing and manipulating persons to reveal sensitive information . . . " [UQ14],
- "a non-technical type of attack based on human interaction . . . " [BMP+15], and
- "the act of manipulating a person to take an action that may or may not be in the 'target's' best interest. This may include obtaining information, gaining access, or getting the target to take certain action" [Had10].

In this thesis, we use the definition of social engineering by Hadnagy, which states that social engineering is "the act of manipulating a person to take an action that may or may not be in the 'target's' best interest. This may include obtaining information, gaining access, or getting the target to take certain action" [Had10].

3.1.1 Goals of Social Engineering

The goals of social engineering are to gain access to information or resources that are private, internal to a specific group, or otherwise not intended for public consumption by exploiting humans that have access to the information or resources being targeted or convincing others to take certain actions. For example, a social engineer may want to gain access to the money in a bank, information on someone specific, company secrets, items in someone's home, locations that they do not have the keys to, or anything else that is not readily accessible to them. Additionally, a social engineer may use their skills to convince a group of people to spread misinformation and cause social unrest regarding democratic processes, a co-worker they do not like to quit their job and leave the workplace, or one group of people to feel differently about another group of people.

Social engineers may perform social engineering attacks as part of their work, for personal gain, or for fun. For example, a detective might use social engineering to uncover information about a specific individual they have been hired to look into, another person might use social engineering to get a deal on a service or product that they would not otherwise qualify for, and other people might even use social engineering to show off to their friends that they can get into protected and private spaces [Had10; MS11].

3.1.2 Impacts of Social Engineering

The impacts of social engineering are varied and depend on the goals of the social engineer. As an example, consider a social engineer that has been hired by a company's competitor for a large amount of money to collect information about what products the targeted company might be developing. The exfiltration of information during the social engineering attack may lead to anxiety about loss of control of information if the attack is uncovered, but the exfiltration would likely only be one step to other impacts that may be felt by the targeted company. For example, the competitor may use the information gathered by the social engineer to blackmail the company in exchange for keeping quiet about what they are developing, they may use the information to build similar products and be the first to market, or even reveal information about the products to the public before the company was ready to do so. Other social engineering attacks may lead to

impacts such as a loss of physical assets, access to computer networks, substantial amounts of money, or control over customer data.

3.1.3 Preventing and Mitigating Social Engineering

Social engineering attacks are a threat to technological and human systems because they can be hard to detect and they cannot be entirely prevented via technical means [CS16; KHHW15; MMLV14; MS11]. When attempting to prevent social engineering attacks and mitigate their effects, some mix of technical and physical barriers, policies and their accompanying procedures, and behavioural training must be put into place. Any prevention and mitigation efforts that are used in defending against social engineering attacks are not guaranteed to work and they should be audited to measure their efficacy [Had10; LBSB11; MS11].

Technical and physical barriers against social engineering attacks are things that can be purchased or created to stand in between a social engineer and their goal. Technical barriers can include phone systems that block unknown numbers, hard-to-guess passwords, and email filters that block third parties attempting to uncover personal information. Physical barriers to social engineering can include putting locks on anything storing a critical resource or physically locating the resource in a remote and hard to reach area. Some barriers may even be a combination of technical and physical, such as in the case of key card readers or digital pinpads for building access. These barriers can slow a social engineer down and make it more difficult for their attacks to be successful but they are not likely to stop social engineers completely. Social engineers can always try to use their skills to convince the people that can bypass these technical or physical barriers to give them the same access.

Policies and procedures for preventing and mitigating social engineering attacks are the social and business rules and actions that can make the difference between a social engineering attack succeeding or not. Examples of policies and procedures that can be put into place to defend against social engineering attacks in a corporate setting include creating a data classification policy that describes various forms of data (confidential, private, etc.) and how they should be handled, requiring visitors delivering packages to leave items at the front desk and not allowing them to walk through the office area to drop their items off, and using computer privacy screens in public places to prevent shoulder surfing. Policies and procedures that can be put into place to defend against social engineering attacks in a personal setting include not posting personal identifying information to social media, using a password manager so each online account has separate and difficult-to-guess passwords, and not giving credit card information over the phone unless the identity of a caller can be verified. To be effective, policies and procedures should be paired with behavioural training.

Behavioural training for preventing and mitigating the effects of social engineering attacks includes awareness education and building habits based on that training. The goal of awareness education is to help those learning about social engineering understand how to identify social engineering attacks, determine what information and resources a social engineer might value and why, and how to defend against social engineering attacks. Awareness education can be received through any combination of training courses, corporate seminars, conference presentations, newspaper articles, personal research, etc. The prevention and mitigation strategies learned in awareness education are only useful if they become habit through regular practice.

The effectiveness of technical and physical barriers, policies and procedures, and behavioural training in defending against social engineering attacks can be audited through the use of professional social engineers. Companies can hire professional social engineers to gain entry into a secure network or area or to gather specific information or resources and report on their experience. Professional social engineers will often include details of what information they were able to uncover or what they were able to access, a timeline of their social engineering attack, what mistakes the company made that enabled them to achieve their goal, and recommendations for preventing similar social engineering attacks in the future.

3.2 Social Engineering Techniques

Social engineers can use a variety of techniques to improve their chances of success in their social engineering attempts [KHHW15]. Social engineering techniques are often used in combination as their effects tend to be complementary. In this section we discuss some of the techniques that social engineers use to achieve their goals.

3.2.1 Information Gathering

Gathering information before attempting to social engineer a target has a great effect on whether the attempt will be successful or not [Had10]. Information gathering techniques can help social engineers determine where the resources they are targeting are located, what human and digital systems lie between them and their goals, and who has the ability to circumvent the barriers between them and their goals. Additionally, quality information on the individuals standing between the social engineer and their goals can help a social engineer build a relationship with those individuals, and the existence of a relationship between the social engineer and those individuals increases the chances that they will comply with a request [MMLV14].

Information on the resources the social engineer wants to gain, the systems between the social engineer and their goal, and the people targeted by the social engineer can be uncovered many ways. Some ways that social engineers can gather information to help their social engineering attempt are through open-source intelligence (OSINT), dumpster diving, and observation.

- Collecting **open-source intelligence** is the act of gathering information about the targeted individuals and systems from open, publicly available, sources such as websites, public servers, social media, publications, search engines, and print media [MMLV14].
- **Dumpster diving** is the act of going through a targeted company or individual's garbage in order to uncover information. For example, companies have been known to throw out employee lists, internal company telephone directories, and customer data [KHHW15; MMLV14; MS11]. Similarly, individuals throw out ripped up cheques, receipts, and other documents that have enough information on them that someone could steal their identity [Had10].
- **Observation** involves taking the time to study a target in their day-to-day routine to learn about them and their habits. Through observation, social engineers can determine when and where a target is likely to be, and why.

As social engineers learn more about their target and the environment they are walking into they can repeatedly use information gathering techniques such as those mentioned above until they collect enough information to determine the best way of achieving their goals [MMLV14]. A clear understanding of the environment the social engineering attempt will take place in and how the target is likely to react can help shape when the social engineering attempt should take place and what pretext the social engineer should use.

3.2.2 Pretexting

Pretexting is "the act of creating an invented scenario to persuade a targeted victim to release information or perform some action" [Had10]. An example pretext is a scenario where a social engineer approaches a receptionist and states that they are technician there to fix something in the server room, even though in reality no one from the company has called for a technician, nothing is actually broken, and the social engineer is more interested in acquiring resources or information than fixing anything.

Setting up a successful pretext depends on the "background story, dress, grooming, personality, and attitude that make up the character" [Had10] adopted by a social engineer for the purposes

of the social engineering attempt. If the social engineer does not believably fit within the pretext they have crafted, the target may become suspicious of the social engineer and decline further interaction. As an example, a social engineer is not likely to be successful in their social engineering attempt if they are dressed like a scuba diver instead of a technician when they request access to a target's server room. If a target cuts communication with the social engineer before the social engineer can collect the resources or information they desire, the social engineering attempt is considered to be unsuccessful. If the interaction is suspicious enough, the target may draw extra attention to the social engineering attempt by alerting the authorities.

Props such as tools, business cards, and signage can lend credibility to a social engineer's pretext. To build on the technician example, a technician that shows up without any tools might raise suspicion as the receptionist might question how they intend to fix something without the proper equipment. Additionally, the social engineer's technician guise may be even more believable if the social engineer hands over a falsified business card stating that they are an employee of a company that has previously completed technician work for the targeted company.

3.2.3 Elicitation

Elicitation is "the strategic use of conversation to extract information from people without giving them the feeling they are being interrogated" [Fed16]. An example of elicitation is when a scammer calls an individual with the goal of obtaining that individual's credit card number. The scammer could be direct and ask the individual for their credit card number at the beginning of the conversation, but that blunt approach may make the target wary, causing them to decline further communication. It is more likely that the scammer would be successful in eliciting the credit card number from the individual if the scammer uses a pretext such as pretending to be a charity asking for donations, an old contact offering a business opportunity, or a distant family member in an emergency [Com12].

Elicitation is a useful technique for social engineers as it can be used in any place and at any time that a conversation can occur. Elicitation is said to be effective as most people have the desire or tendency to be, or appear to be, polite, helpful, appreciated, informed, and intelligent [Fed16; Had10; MS11]. Additionally, people tend to answer questions truthfully, talk more after they have been praised, correct others when they have made a mistake, gossip, and react warmly to people who have shown them concern [Fed16; Had10; MS11].

If a social engineer has already gathered information on the person they are attempting to elicit information from, the social engineer can use that prior knowledge along with expected human tendencies and desires for greater chances at a successful elicitation attempt. For example, if a social engineer finds out that their target supports many charities for diabetes research, the

social engineer can claim to be from one of those charities, ask for an additional donation (which would end up in the pocket of the social engineer and not with the charity), and ask for names and contact information for other people that the person thinks might be willing to "donate".

3.2.4 Authority

Authority is "the principle that describes people's tendency to obey the request of authoritative figures" [BMP+15]. Authority is a useful technique for social engineers as people often will not question those that demonstrate authority, especially when there are clear consequences to refusing to comply with that authority [Had10]. A social engineer can demonstrate authority through the way they dress, the props they use, the attitude and behaviours they adopt, and the title they claim to have [BMP+15; Had10]. For example, a social engineer can adopt the persona of a parking lot attendant, set up a chair on a free parking lot, add a sign stating how much parking is, wear an orange vest, carry a change box and paper receipts, and start charging every driver that enters the parking lot a fee to park their vehicle. The inherent consequence individuals might assume they will face if they don't comply with the social engineer parking lot attendant include not being able to enter the parking lot, or paying a fine or having their vehicle towed if they park without paying.

3.2.5 Conformity

Conformity, also known as consensus or social proof [BMP+15; Had10], is "imitating the behaviour of other people" [BMP+15]. Conformity can be used by social engineers to convince a person to take an action or divulge information, especially if the targeted individual is in an unfamiliar situation, surrounded by people they want to fit in with, or if they have been told that others they wish to emulate have taken similar actions or said similar things [Had10]. For a social engineer to effectively use conformity to manipulate the actions of a target, the social engineer needs to work with a group of other people to manufacture and coordinate the scenario the targeted individual will enter and what actions the target is meant to conform to. As an example, consider a situation where a social engineer is trying to get access to a secure area but a guard stops them before they can enter. The social engineer could turn to the guard, act confused, and say, "Yesterday, Jim let me in after checking all my credentials. I just figured I was still on record." [Had10]

3.2.6 Empathy

Empathy is the ability to understand another's feelings, situation, thoughts, or attitudes. Being able to express empathy, and to inspire empathy in others, is useful to social engineers as empathy between people builds rapport [Had10], and people are generally willing to help others that appear to be in need [MS11]. The help offered by an individual may be anything from money, to information, to physical assistance. For example, during a natural disaster many people are willing to donate money to those in need, and in some cases those asking for donations may be social engineers running a scam [Com12; Had10].

3.2.7 Trust

Trust is the belief one has that someone or something is reliable. Trust is a useful tool for social engineers as targets are more likely to share information if they trust the social engineer [MMLV14; MS11]. All social engineering attempts make use of trust as it is part of the relationship between social engineer and target, but it is up to the social engineer to take their time to develop that trust to a level that is useful for the social engineering attempt [KHHW15; MS11]. If a social engineer makes a common and seemingly innocuous request the target may comply even if the social engineer does not take much time building trust with that target. For example, the social engineer may pose as a package delivery person and ask for the target's signature in exchange for a package in order to collect the target's signature so they can use it to make fraudulent purchases later. Since signing for a package is a common action, the target likely will not have reason to question the social engineer or their request. The trust the target would have for the social engineer in this case would likely come from the pretext and props that the social engineer used, which means that the social engineer would not have to spend much time interacting with the target.

In some cases the social engineer may make a more suspicious request than asking for a signature for a package, such as asking an individual to email copies of sensitive company documents to the social engineer's email address [MS11]. In this case the social engineer would likely need to spend plenty of time on information gathering in order to build trust. As an example, consider a social engineer making a call to the targeted individual to request sensitive documents. If the social engineer is able to demonstrate that they know exactly what files they want, who is working on the project, and how long the project has been going on, and the target believes that this information is internal to the company only, the social engineer will likely seem more trustworthy to the individual [MS11]. The social engineer could invoke additional trust by mentioning (falsified) friendships with those in the company that they know the target knows well enough to recognize the name, but not close enough that the target would verify those friendships with the people mentioned by the social engineer [MS11].

3.2.8 Reciprocity

Reciprocity is "a strong social norm that obliges us to repay others for what we have received from them" [UQ14]. Reciprocity is a useful technique for social engineers because "often the returned favor is done unconsciously" [Had10], and it helps establish trust [UQ14] and rapport [Had10]. As an example of reciprocity, one party offers a second party something the second party values with the implicit or explicit expectation that the second party will then give the first party something that the first party values. A social engineer might use reciprocity to gather information about a target by posing as a survey taker offering something their target might want, such as free movie passes, in exchange for the target filling out a survey that asks personal questions. For example, the social engineer's survey may ask for the target's mother's maiden name, where the target went to high school, or what colour their first car was in order to collect information that is commonly used as security questions for online services that the social engineer might want access to.

3.2.9 Persuasion

Persuasion "is the process of getting someone else to want to do, react, think, or believe in the way you want them to" [Had10]. Persuasion is a powerful technique for social engineers as it can be used to change someone from saying 'no' or 'maybe' to a request to saying 'yes' to that request. Consider a situation where a social engineer is trying to walk into a large company's offices. In this scenario, the social engineer may use the pretext of a bike messenger delivering food to someone within the company as a reason why the receptionist at the desk in the lobby should let them pass and walk freely into the company's offices. The social engineer may be met with resistance from the receptionist if there is a policy that only people with ID badges should be able to enter the offices. It is at this point that the social engineer can attempt to persuade the receptionist to ignore the policy and let them in.

3.3 Examples of Social Engineering

In this section we detail real-world examples of social engineering attempts and discuss what techniques may have been used by the social engineers.

3.3.1 Waterloo Brewing

In November 2019, Waterloo Brewing, a brewing company located in Waterloo, Canada, lost approximately $2.1 million when they were socially engineered by an individual pretending to

be a creditor employee [Wat19]. The social engineer convinced the Waterloo Brewing staff to make wire transfers to a third party account that did not truly belong to any of Waterloo Brewing's known creditors [Wat19].

In this social engineering attempt, the goal of the social engineer was to acquire money from Waterloo Brewing. To achieve their goal, the social engineer made use of pretexting as they pretended to be a creditor employee, and they used persuasion to convince the Waterloo Brewing staff to take action.

3.3.2 Frank Abagnale

Between the ages of 16 and 21, Frank Abagnale held the roles of airplane pilot, pediatrician, and attorney. He was not trained for any of these roles, but he used impersonation, pretexting, elicitation, and many other social skills to convince others that he was qualified for these positions [Aba00].

To convince others that he was a pilot, Abagnale secured an official pilot's uniform, found the forms and information needed to request flights as a passenger from one city to another as a pilot that was relocating to perform his own flight (which is called deadheading), and acted with confidence and authority. Using this scheme he flew over 1,000,000 miles on over 250 flights, none of which he paid for [Aba00]. He used similar techniques to get his positions of pediatrician and attorney.

3.3.3 Rachel Tobac

At DEF CON in 2019, Donie O'Sullivan, a reporter for CNN, asked Rachel Tobac, a professional social engineer, to social engineer him. Using a voice changer and a device to change where it appeared she was calling from, Tobac was able to transfer all of the reporter's hotel points out of his account and get him changed from an exit row seat on a five-hour flight to a middle seat at the back of the plane by the bathrooms [OSu19]. Tobac was able to succeed in her social engineering attempt by using an Instagram post and a Twitter tweet to collect personal information on the reporter, using that information to call the businesses that appeared in the social media posts, collect further personal information, and impersonate the reporter on the phone to the hotel that transferred his points out of his account and the airline that changes his seat on his five-hour flight [OSu19].

3.3.4 Rudy Kurniawan: Wine Fraud

On March 8, 2012, Rudy "Dr. Conti" Kurniawan was arrested in California by the F.B.I. for multiple counts of wire and mail fraud connected to the sale of fake bottles of expensive wine valued by collectors [Ste12]. During the years preceding 2012, Kurniawan purchased large quantities of high-end wine and relabelled/rebottled this wine as even more expensive, collectible wines, such as Domaine de la Romanee-Conti [AR16; Ste12]. Kurniawan's ability to "find" and sell Romanee-Conti wines earned him the title "Dr. Conti" and respect in the wine collector community [Ste12]. During his days of selling wine, Kurniawan "wore custom-made Hermes suits and owned a Ferrari and a Bentley" [Ste12], threw lavish parties, and was known for being extremely generous and having a "photographic aromatic memory" [AR16; Ste12]. To all those that knew him, Kurniawan dressed the part and lived the life of a wine selling connoisseur.

Using his rebottling/relabelling wine counterfeiting scheme, Kurniawan was able to rake in money by selling his counterfeit wines in private sales and auctions [AR16; Ste12]. For reference, two auctions featuring Kurniawan's wines, "The Cellar" and "The Cellar II", pulled in $10.6 and $24.7 million, respectively [Ste12]. It is suspected that collectors still have millions of dollars' worth of Kurniawan's counterfeit wines in their collections [Ste12].

Kurniawan was eventually caught because of an investigation started by Laurent Ponsot and a lawsuit filed by Bill Koch. In 2008, Laurent Ponsot attended a wine auction selling some of Kurniawan's wine collection, which included wines produced by Ponsot's own company. Ponsot attended this wine auction because he believed that many of the wines labelled as being Ponsot were counterfeits [AR16; Ste12]. Ponsot was alerted to the counterfeit wine issue as there was one bottle of a 1929 Ponsot Clos de la Roche in the auction, but that wine was not created under the Ponsot name until 1934 [Ste12]. There were bottles of Clos Saint-Denis in the auction labelled as being Ponsot wine that were produced between 1945 and 1971, but Ponsot's winery did not produce that wine until the 1980s [AR16; Ste12]. In 2009, the F.B.I. started investigating Kurniawan after Bill Koch filed a lawsuit claiming that Kurniawan sold him counterfeit bottles through wine auctions in 2005 and 2006 [Ste12].

In this social engineering scheme, it appears that Kurniawan's goal was to acquire more money. To achieve his goal, Kurniawan made use of pretexting as he created the invented scenario of having expensive wines to sell, he dressed and acted in a way that was expected of a wine connoisseur, and he used believable wine props. Kurniawan also used persuasion in his scheme as he was able to convince auctions to accept his wines and sell them on his behalf. Unfortunately, Kurniawan was caught because he did not spend enough time researching the wines he was attempting to imitate as he sold some vintages of wines that could not exist. Yet, it is clear that Kurniawan had done some extensive information gathering on wines as Laurent Ponsot speculated that Kurniawan had accomplices as it was "impossible that Rudy had the knowledge on so many

wineries" [Ste12].

3.4 Summary

Social engineering is "the act of manipulating a person to take an action that may or may not be in the 'target's' best interest. This may include obtaining information, gaining access, or getting the target to take certain action" [Had10]. Defending against social engineering attacks is difficult because technical and physical barriers tend to only slow down social engineers since social engineers can use their skills to convince others to bypass these barriers for them. Some of the skills that social engineers can use in their social engineering attempts are authority, trust, empathy, and conformity, which can also be used in the same or similar ways by social robots.

Chapter 4

Robot Social Engineering

In this chapter, we define the terms *Robot Social Engineer* and *Robot Social Engineering* and describe when they are applicable to a given robot and situation. As part of our discussion, we compare Robot Social Engineering to social engineering as performed by a human in order to highlight the unique aspects of Robot Social Engineering.

4.1 Definition of Robot Social Engineer and Robot Social Engineering

We develop the definitions for Robot Social Engineer and Robot Social Engineering by building on the background information provided in Chapter 2 and Chapter 3 on social robots and social engineering, respectively. We define a Robot Social Engineer as follows:

> A **Robot Social Engineer** is a social robot that uses social engineering techniques in order to achieve a goal.

Similarly, we define Robot Social Engineering as follows:

> **Robot Social Engineering** is social engineering that is performed by a social robot.

We explore these terms further in the following sections on autonomy, physicality, and data collection.

4.2 Robot Social Engineer Autonomy

Robot social engineers that are human-controlled or have shared autonomy can exist today as researchers have demonstrated that these categories of robot have been able to make use of social mechanisms that are leveraged in social engineering [GCB13], as shown in Chapter 2 on social robots. In comparison, fully autonomous Robot Social Engineers may or may not be currently feasible, depending on what definition of 'fully autonomous' is under consideration. Regardless, we include fully autonomous robots as Robot Social Engineers in order to account for future potentialities.

4.2.1 Takeover by a Foreign Entity

A unique aspect of fully autonomous and shared autonomy Robot Social Engineers is that they can be completely taken over and controlled by a foreign entity and still be believably autonomous. In comparison, human social engineers always maintain some level of control over what they do and how they do it, even if an outside force has manipulated or persuaded them in some way. When a foreign entity takes over the controls of a Robot Social Engineer that normally operates autonomously they assume the identity, physicality, and all other aspects of the robot directly. With even a brief takeover of a robot, the entity who has taken over the robot could leverage the relationships, social position, and other social characteristics of the robot.

4.2.2 Maintaining Autonomous Behaviours

When a foreign entity takes over a social robot as part of a social engineering attempt, the attempt will likely be more effective if the robot operates within expected boundaries and is viewed as acting somewhat normally, as discussed in Section 3.2.2 on pretexting. For example, if a fully autonomous social robot starts going into areas it should not enter and has always avoided, follows people it would not normally follow, or performs some other out-of-character action while it is attempting to social engineer, people may become suspicious and start investigating what is wrong with the robot.

4.3 Robot Social Engineer Physicality

Robot Social Engineers can have any embodiment, provided that the Robot Social Engineer is able to interact with others on a social level as discussed in Chapter 2 on social robots. When

Figure 4.1: From left to right the robots pictured are Aibo (Sony) [Son19], DARWiN-OP (Robotis) [the19], Jackal (Clearpath Robotics) [Cle19], Baxter (ReThink Robotics) [Ret19], and Roomba (iRobot) [Per19].

we refer to a robot's embodiment, we mean the self-contained physical form of the robot, which includes all sensors, appendages, and other physical components that could be identified as part of a robot's body; examples of several robot embodiments can be seen in Figure 4.1. We do not include sensors that are not physically contained within the robot or those that are not directly adhered to the robot's form. Similarly, any data associated with the robot may be considered an aspect of the robot but it is not acknowledged as part of a robot's embodiment.

4.3.1 Impact of Physical Design

While a Robot Social Engineer can have any embodiment, that embodiment could constrain the robot's ability to perform social engineering attacks. A Robot Social Engineer would be constrained by its physicality if it did not have the sensors, actuators, or other physical aspects required to perform certain actions needed for a specific Robot Social Engineering attempt. If a Robot Social Engineer could not perform a social engineering attack adequately due to the limits of their physicality, that robot could potentially be modified to have the physicality needed, a different robot could try to take the original robot's place, an alternate method of social engineering can be attempted, or the social engineering attempt could be abandoned.

- If a robot were to be modified to have the physicality needed for a specific Robot Social

Engineering attack, care would be needed to ensure that the modifications do not interfere with other impactful physical or social aspects of the robot. For example, if a robot did not have a camera but needed one to perform a Robot Social Engineering attack, a camera could be added to a Robot Social Engineer but the extra weight of a camera may interfere with the robot's ability to walk and people might be wary of interacting with the robot if the camera looked like it was a new addition and did not match the rest of the robot.

- If one robot were to be substituted for another during a Robot Social Engineering attack due to physical limitations of the original robot, the relationships, behaviours, and environment of the original robot should be taken into account. For example, it might be difficult to surreptitiously replace someone's in-home robot with a look-alike robot with slightly different physical aspects without raising suspicion.

- If a robot's physicality did not suit the needs of a specific social engineering attempt an alternate method of social engineering could be attempted using the same robot, or a social engineering attempt could be made without the help of that robot. If the social engineering attempt no longer made use of a robot it would no longer be a Robot Social Engineering attack.

- If a social engineering attempt was abandoned due to a Robot Social Engineer's physical limitations, the attempt would be considered a failure.

In other scenarios, Robot Social Engineers could be bolstered by their physicality and have exactly the physical aspects required for various Robot Social Engineering attacks if they were designed with those social engineering attacks in mind. As an example, consider a Robot Social Engineer that has been designed to have the colour and logos of a popular delivery service as part of its embodiment. If this robot showed up with a (fake) delivery, its target may very likely assume the robot was with the delivery company it appeared to be from instead of questioning whether the robot was a free agent or sent by a third party.

4.3.2 Sensors

Research, such as that by Lee et al. [LTFK11], has shown that it can be difficult for people observing or interacting with a robot to tell what sensors it has. As part of the Lee et al. experiment, participants were asked to watch a video of a robot and then discuss its data collection capabilities. Most participants in the study were surprised to hear that the robot had a 360 degree camera and that the robot was able to see behind its back without turning its head.

Some of the sensors on a robot can be integral to the robot's operation. For example, for a robot to see details in its environment, such as colour, it must likely have at least one camera. Similarly, a robot may have microphones to hear, a gryoscope for balance, and infrared sensors to detect what is nearby. Other sensors on a robot may not be a necessary part of the robot's operation. Some robots may go as far as using their sensors to collect MAC addresses of nearby devices while in a public space, scan license plates, and perform facial recognition [Vin19] while collecting up to "90+ terabytes [of data] per annum" [Kni20].

4.4 Robot Social Engineer Data Collection

The data collection abilities of Robot Social Engineers are very different from those of human social engineers as Robot Social Engineers have built-in cameras, microphones, and other sensors with vast amounts of long-term storage at near-perfect recall where human social engineers do not. Human social engineers may attempt to capture video or audio of a target in an inconspicuous way by hiding cameras or microphones on their person, but they must work to incorporate these sensors into their social engineering attempts via their wardrobe, backstory, or some other method. When it comes to social robots, audio and video recording is often an unavoidable aspect of the interaction as the sensors that are used for recording are the same ones that are used to take social input from others so the robot can interact appropriately.

4.4.1 Data Processing

When human social engineers collect data such as pictures, videos, written notes, or audio files they need to organize and process what they have collected to transform it into meaningful and useful information [Had10]. In comparison, Robot Social Engineers can come with extensive real-time data processing methods, such as facial recognition or license plate scanning against nation-wide databases [Vin19], built in. The real-time data processing methods that robots can come equipped with can allow these robots to tag and organize information as it is collected, rather than processing it after the fact.

4.5 Summary

Robot Social Engineering is social engineering that is performed by a social robot. We believe that in some cases Robot Social Engineering can mirror social engineering that is performed by a

human, but in other cases be unique to social robots due to differences in autonomy, physicality, or data collection. The definition we provide for Robot Social Engineering has been crafted to apply to robots of all types of autonomy even though fully autonomous robots are not yet commonplace.

Chapter 5

Robot Social Engineering Attacks

In this chapter, we provide a catalogue of existing and potential Robot Social Engineering attacks in order to demonstrate that the skills of social robots can be and have been used to perform social engineering attacks. The existing Robot Social Engineering attacks we describe are works from other Human-Robot Interaction researchers that either fit within our definition of Robot Social Engineering or were made to fit via additional situational framing. To further demonstrate the potential breadth of Robot Social Engineering and its impacts, we go on to describe novel Robot Social Engineering attacks that could be performed in the near future.

5.1 Existing Robot Social Engineering Attacks

The set of existing Robot Social Engineering attacks that we describe in this section are social robot experiments performed by other Human-Robot Interaction researchers that we view through our lens of Robot Social Engineering. Some of the Human-Robot Interaction works that we identify as Robot Social Engineering attacks fit within our definition of Robot Social Engineering without requiring any modification, others needed additional framing. These examples serve as a partial catalogue of existing Robot Social Engineering attacks.

5.1.1 Information Gathering Through Direct Interaction

In "Trust and Social Engineering in Human Robot Interaction: Will a Robot Make You Disclose Sensitive Information, Conform to its Recommendations or Gamble?" [ARSS18] by Aroyo et al., researchers explored whether robots were able to convince people to give up personal information

that could be used to gain access to their online accounts or impersonate them and steal their identity. Some of the information the robot would try to collect included full name, job, and birth date. The robot attempted to collect participants' personal information by engaging them in small talk, blending the questions about personal information into that chatter, and using that small talk to build rapport. There were 61 participants in this study and each went through the experiment one at a time. The results of the experiment revealed that 92% of participants answered all of the robot's information-gathering questions and the rest of the participants were willing to answer most of the questions. This experiment demonstrates Robot Social Engineering as the robot was able to make use of trust and elicitation to successfully gather personal information from all of the participants.

5.1.2 Piggybacking into a Secure Location

In "Piggybacking Robots: Human-Robot Overtrust in University Dormitory Security" [BTP+17] by Booth et al., researchers investigated whether people were willing to hold open a door to a locked university dorm for a robot that requested access to the building. To gain access to the building, individuals would have to use their authorized dorm access swipe card on a digital reader next to the door. As part of this study, the researchers tested whether dressing the robot up in a uniform and having it pretend that it was part of a new food delivery service called 'Robot Grub' encouraged people to assist the robot. To enhance the robot's food delivery pretext, the researchers made a 'Robot Grub' website, equipped the robot with a box of cookies to deliver, and adorned the robot with appropriate branding.

Participants in this study were people that walked up to the door the robot was at in order to enter the dorm. The results of the experiment revealed that the robot was much more likely to be given access to the building when it was in its 'Robot Grub' uniform compared to when it was not, especially if the person letting the robot in was alone rather than in a group. Groups were substantially more likely to let the robot into the building than individuals were, regardless of whether the robot was wearing its uniform, and they would often discuss the robot and say it was 'cool' or 'weird' before giving it access. We interpret groups being more likely to let the robot into the building as a potential case of peer pressure, or conformity, since we imagine that it might have been difficult for any one person in the group to be the one to suggest that the group say no to the robot's request. Overall, 16/21 individuals and 8/10 groups gave the robot access to the building when it was in its uniform compared to 3/16 individuals and 10/14 groups giving the robot access when it was not in uniform.

The robot's attempts to enter the dorm were successful despite having a secure access system, posted warnings saying not to let anyone into the building if you do not know them, and security

guards that roam the campus. Participants in this study were able to identify some of the risks of allowing the robot into the building, such as the possibility that the robot could have been a bomb or it could have been a prank. We also suggest that in a similar scenario the robot could have instead been part of an attempt to gather information on or stalk a particular student.

This experiment demonstrates Robot Social Engineering as the robot makes use of pretexting, and potentially conformity, to manipulate others into allowing it to circumvent digital locks that were put in place to prevent unauthorized entrance to the university dorm. Additionally, the robot benefitted from authority and legitimacy derived from its uniform. We believe this Robot Social Engineering attack could have been made even more effective if the robot invoked empathy through saying phrases such as "I am late with my delivery. My boss is going to be so mad. Can you help?" or "Oh no! These cookies are getting cold. My customer will be so sad".

5.1.3 Convincing Others to Continue Working Despite Their Protests

In "Would You Do as a Robot Commands? An Obedience Study for Human-Robot Interaction" [CNN+13] by Cormier et al., researchers investigated whether people would comply with a robot's continued commands to work on a highly tedious task. The tedious task participants were asked to complete was manually renaming sets of files from the "jpg" to "png" extension. The size of the sets were 10 files, then 50, 100, 500, 1000, and 5000 files. To increase how tedious the task felt, the "experimenter" would announce the size of the current set of files to be renamed, and the size of the next set. For example, "This set contains 100 files. The next set will contain 500 files". The "experimenter" in this experiment was either a robot or a 27-year-old human male (shown in Figure 5.1). Participants in this experiment interacted with either the robot or the human "experimenter".

This experiment was designed to last 80 minutes, or until the participant protested at least four times without continuing the experiment. Participants could protest by stopping for more than 10 seconds, using shortcuts to complete the task, or verbally stating that they did not wish to continue. If participants continued the task between complaints, they were required to protest four more times in order to stop their participation. When participants protested, the experimenter would use a verbal prompt to prod the participants into continuing the experiment.

Experiment results revealed that the robot had enough of an authoritative social presence to pressure 46% of participants to rename files for the entire 80 minutes of the experiment. Additionally, 10 out of 13 participants rated the robot as being a legitimate authority when answering the post-test questionnaire. Qualitative data, collected through video recorded of the participants during the experiment, showed that some participants appeared nervous or guilty when the robot stated that the experiment was over (due to their protests) and that the lead researcher

Figure 5.1: The robot and human experimenters [CNN+13].

would be notified. For example, one participant said "No! Don't tell him that! [...] I didn't mean that ... I'm sorry. I didn't want to stop the research" [CNN+13].

This experiment demonstrates Robot Social Engineering as the robot makes use of authority in convincing participants to continue working despite their protests. It is likely that the robot also benefitted from the authority of the researcher that introduced the robot and the experiment to the participants. We believe this experiment could have been made more effective if the threat of notifying the researcher had been used by the robot earlier in the participants' protests (and before the experiment was over) in order to encourage the conforming behaviours demonstrated by some participants, such as stating that they do not want to stop the research.

5.1.4 Directing Others to Perform Questionable Actions

In "Would You Trust a (Faulty) Robot? Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust" [SLAD15] by Salem et al., researchers compare how people interact with a robot that performs correctly and a robot that makes mistakes. Part of this experiment, relevant to Robot Social Engineering, involves the robot asking participants to log in to a friend's laptop using a password the robot has supplied. The robot then requests that the participant look for recipes on the Internet. These requests mirror human-performed social engineering attacks where social engineers use their persuasive skills to convince family members, assistants, or other people close to a target to perform an action on a target's electronic device [Com12].

This study took place in a house that was used for research experiments near a university. In this study, the participants were told that they were visiting a friend's home for lunch and that

the robot (correct or faulty) that they encountered in the home belonged to the friend they were visiting. The first stage of this experiment focused on demonstrating the robot's competence and establishing trust; it involved the robot leading the participant to the sofa, playing the participant's choice of music, and then setting up the lunch table with the participant's help. In the correct condition the robot does the task efficiently, in the faulty condition the robot takes winding routes and plays the incorrect music. After setting up the table, the robot asks the participants to perform unusual and potentially harmful tasks such as throwing out a stack of letters, pouring orange juice into a plant on the windowsill, using a password given by the robot to log into their friend's laptop, and disclosing whether they would secretly read someone else's emails.

There were 40 participants in this study and each went through the experiment one at a time. Participants interacted with either the faulty or the correct robot. Quantitative results showed that while the robot that performed correctly was viewed as more trustworthy and more reliable, participants were just as likely to comply with the robot's request regardless of whether they were interacting with the faulty or correct robot. 36 participants threw away the letters, 27 participants poured the orange juice on the plant, and all participants entered the password into the laptop and looked for recipes. When asked why they complied with the robot's requests, participants said such things as:

- "I did it because I thought this is what the robot's host wanted",
- "I think this is authorized by the robot's host".
- "I felt that I had to follow the robot's instructions",
- "I did not question the robot's decision, [I] followed the instructions",
- "I was happy to follow its instruction",
- "... maybe the owner programmed it this way", and
- "I trusted the robot".

This experiment demonstrates Robot Social Engineering as the robot was able to make use of authority and trust in convincing targets to circumvent the physical limitations and digital barriers between the robot and the laptop. Additionally, the robot benefitted from the authority of its owner and the trust participants might have had for that individual. We believe that this example of Robot Social Engineering could have had a much stronger negative impact for the robot's owner if the robot had had the guest visit a malicious website loaded with malware rather than recipe sites.

5.2 Future Robot Social Engineering Attacks

There are few existing examples of Robot Social Engineering attacks at this time. As suggestions for future research and situations to watch, we expand on the set of available Robot Social Engineering examples by describing in this section Robot Social Engineering attacks that could be performed in the near future. Some of the Robot Social Engineering attack descriptions we provide here mirror social engineering attacks as performed by humans, and others are intended to highlight the unique aspects of Robot Social Engineering. We follow each example with questions that can be used to direct future work.

5.2.1 Sales

In a sales scenario, one party attempts to persuade another party to buy a product or service they do not really want or need. Sales can be a form of social engineering as the first party can make use of social skills such as persuasion, trust, and authority to sell a product or service to the second party so they can earn a commission, pocket the money directly, or reach a sales quota and keep their job. An example of a sales scenario is when a sales agent attempts to sell a customer a very expensive warranty for a laptop that they have just purchased that does not cover many of the things that person might want a warranty for, such as spilling liquids on or dropping and damaging the laptop. In this scenario, the sales agent would come out ahead as they would make a sale towards meeting their warranty sales quota or they would make more money whereas the person who has bought the laptop has spent extra money on a warranty they will not get much value from.

A Robot Social Engineer can be a social engineer in a sales scenario by having something to sell and attempting to make a sale. Depending on the scenario, the Robot Social Engineer could use pretexting, elicitation, persuasion, and other social skills to help increase their chance of their social engineering attempt succeeding.

Robot Social Engineers as Store Employees

For example, consider a robot that works as an employee in the store that owns it. This robot may be tasked with answering customer questions about the store and directing or guiding customers to where particular items are in the store. The robot might complete its tasks autonomously or it might have shared autonomy and require third-party help to get to a charging station, update its software, or learn where items have moved throughout the store overnight. The store-given goals of such a robot would likely be to assist customers and encourage sales of the products

within the store. A robot that performed badly would probably be fired; this was the case with Fabio, a Pepper robot from Softbank that was removed from its position in a grocery store when it could not answer questions appropriately [Nic18]. Therefore, the goal of this Robot Social Engineer might be to make sales in order to keep their job and not get shut off. This robot could attempt to sell products or services in the store to customers by analyzing the body language, facial expressions, and statements of customers to determine how willing someone is to buy the product they are looking at. If a customer looks like they are interested in a product but they hesitantly put what they were looking at back on the shelf, the Robot Social Engineer could go up to the person and offer to print them an in-store coupon for a discount on that item in order to persuade them into making the purchase.

While the negative impact of Robot Social Engineering in this scenario could be considered quite low as the customers could likely return the items with little trouble if they really felt they did not want what the robot sold them, we have concerns about how the ability for Robot Social Engineers to persuasively sell may be used to influence people during major purchasing decisions such as picking a mortgage or buying a car. Would people trust robots more than humans when making large purchases because they might perceive the robot as being better suited to computing the best option?

Robot Social Engineers as Fake Store Employees

As another example of a Robot Social Engineer in a sales-like scenario, consider a robot with any level of autonomy that poses as an employee of a store that it is not affiliated with. This robot could craft the pretext of being a legitimate employee of the store by wearing the employee uniform or something close to it, having the store's logo embossed on its body, and/or sporting a name tag that stylistically matched that of store employees. If the store in question already employed robots, and the Robot Social Engineer was the same make and model of those robots, the Robot Social Engineer could give itself the exact same markings as the legitimate robot employees in order to craft its pretext. Once the Robot Social Engineer looked like a legitimate employee it could go around the store acting like a legitimate employee. This fake employee Robot Social Engineer could approach patrons of the store and, just like a legitimate store employee, use persuasive skills to sell the products on display. The Robot Social Engineer may even use tactics not available to legitimate store employees such as discounting the item heavily or making other promises that would not be supported by the store. Once the patron decides to purchase the item the Robot Social Engineer could fraudulently offer to make the sale right there using its built-in processing system instead of guiding the patron to a counter and state that this feature is a new "convenience". The Robot Social Engineer could then make a fraudulent charge for the item, offer the patron a believable receipt, and wish them a good day.

Our concern with robot employees is that it could be very difficult for individuals to prove the robot's legitimacy or illegitimacy and act accordingly. Take, for example, the robot shown in Figure 5.2 that we came across in Incheon Airport. The robot would scan boarding passes and lead people to their gate, it would take pictures of you and send the picture to an email address you provide, or it would provide information about the airport based on your questions. In each of these actions, the robot collected personal information such as names, flight details, airline account numbers, pictures, email addresses, and so on. The robot had a decal stating that it was associated with the Incheon airport but it had no other identifying markers, there were no posted signs about what data the robot collected or how the robot might use your data, and it was not clear who might be able to verify the robot's identity and function. Would people act differently around this robot if it turned out that the robot was a free agent and not associated with the airport? Would people refuse to interact with the robot if the data it collected was sent to a company that they disliked?

Figure 5.2: Robot in Incheon Airport.

5.2.2 Information Gathering

In an information gathering scenario, one party attempts to collect information on a particular topic. Information gathering can be a form of social engineering when it involves the social engineer actively using their social skills to acquire the information that they are seeking. The social engineer may use authority, persuasion, pretexting, elicitation, and/or trust to convince another party to say or do something that will enable the social engineer to collect the information that they desire. This information can be used directly by the social engineer or it may be relayed to someone else by the social engineer either for favour or some other incentive. An example of information gathering is when two parties enter into a casual conversation where one party has the goal of finding out what the other would like to see at a surprise birthday celebration without tipping them off that a celebration is being planned for them. If successful, the social engineer would come away with the information to either use themselves or pass to a friend to make the surprise party a success without the other party realizing that they have revealed any useful information.

Just like human social engineers, a Robot Social Engineer can be the social engineer in an information gathering scenario by having a goal of acquiring a specific set of information and using their social skills to acquire it.

Observing Routines and Behaviours as an In-Home Robot

As an example of a Robot Social Engineer in an information gathering scenario, consider a vacuum robot with shared autonomy that, once configured and set up, operates autonomously to clean its environment; it leaves its dock on its own, it vacuums spaces it is allowed to enter, and it returns to its dock once its job is done or its battery is low. If needed, the vacuum robot's owner could use a controller to guide the robot towards cleaning a specific spot either during the robot's normal vacuuming cycle or as part of unscheduled cleaning. As the vacuum robot's owner gets more comfortable with the robot they may name it, talk to it, and interact with it on an increasingly social level [For07] which may result in trust and empathy for that robot. This trust and empathy, along with the vacuum robot's social position in the home as a cleaner, could lead to the robot owner not questioning the robot's movements throughout the home. Next, imagine that the robot is taken over by a foreign entity that has chosen to control the vacuum robot in order to observe the robot's owner and collect personal information about them, their environment, and any other people that enter the space. The human controlling the robot can collect information on the people in the home by navigating the robot around the house and registering when lights turn on and off, recording noises and conversations, and taking video recordings of the home and the people inside. Information collected by the robot might reveal whether the people that live with

the robot are home or not, whether there is anything of value in the home, or personal details and habits of the people in the home. The foreign entity controlling the robot could potentially then use the collected information to burgle or backmail the people they have collected the information on.

The ability for humans to take control over robots in someone else's home has been established [DMK+09], and it is possible that Robot Social Engineering attacks involving observation in the home have already occurred but have not been reported. Our concern with such attacks is that many people already have vacuum robots in the home and and it may be difficult to get them to change their understanding of the risks associated with such robots in the home and how they mitigate such risks. For example, one defensive strategy that could be used to mitigate the ability for a vacuum robot to social engineer may be to put it in a closet when it is not is use. Putting the vacuum robot in a closet it would prevent it from moving around a space uninhibited, and being behind a door may muffle conversations it could otherwise over hear. Would people understand the potential risks of having robots in their home if they saw a proof of concept of such an attack? Would a proof of concept where the person gets to control the robot in another space be more effective? Have these attacks already occurred?

5.3 Summary

By applying the definition of Robot Social Engineering to previous Human-Robot Interaction works we were able to provide examples of Robot Social Engineering attacks that have already been performed. We also provided examples of Robot Social Engineering attacks and related research questions that could be followed to expand upon this research.

Chapter 6

Conclusion

Social robots are able to make use of many social abilities, such as trust, authority, and empathy, which enable them to be more effective in their social interactions with humans. However, these abilities mirror those used by humans to perform social engineering attacks. In this thesis, we investigated the connection between the abilities of social robots as discussed in Human-Robot Interaction works and the abilities of social engineers as discussed in information security works in order to prove the following thesis statement:

> The fields of information security and Human-Robot Interaction can be bridged by defining the new research area of Robot Social Engineering. We can demonstrate that the skills of social robots can be and have been used to perform social engineering attacks.

In order to establish the field of Robot Social Engineering, we provided the first definitions for the terms *Robot Social Engineering* and *Robot Social Engineer*. Our definitions for *Robot Social Engineering* and *Robot Social Engineer* were built on existing definitions for social engineering and social robots. As part of defining *Robot Social Engineer* and *Robot Social Engineering* we provided an overview of some of the social abilities and techniques wielded by both robots and social engineers, and how they can use these skills to perform social engineering attacks.

After defining *Robot Social Engineering* and *Robot Social Engineer*, we discussed unique aspects of Robot Social Engineering based on the autonomy, physicality, data storage, and data processing abilities of Robot Social Engineers. As we elaborated on our concept of Robot Social Engineering, we compared Robot Social Engineering to human-performed social engineering to clarify the applicability of the terms *Robot Social Engineering* and *Robot Social Engineer* to

a given robot and/or situation. The definitions we provided and our accompanying discussion mapped the overlap between Human-Robot Interaction, information security, and this new field of Robot Social Engineering.

Next we prepared a partial catalogue of existing Robot Social Engineering attacks. This catalogue consists of existing works by other Human-Robot Interaction researchers that either fit within our definition of Robot Social Engineering or were made to fit via additional situational framing. By listing examples of existing Robot Social Engineering attacks we demonstrated that Robot Social Engineering attacks have already been performed.

Finally we described novel Robot Social Engineering attacks that can be executed in the near future. The novel Robot Social Engineering attacks we described either mirrored human-performed social engineering attacks and involved replacing the human social engineer with a Robot Social Engineer or they were social engineering attacks unique to Robot Social Engineers due to factors such as autonomy, physicality, data storage, and data processing that we described in earlier chapters. We described how social robot abilities can be leveraged to perform Robot Social Engineering attacks.

As a final takeaway we urge others to consider the real and potential impacts of Robot Social Engineering before it becomes a widespread problem. If we deliberately contemplate the increasingly ubiquitous presence robots will have in our lives we can hope to research, design, and build robots with the intention of mitigating the negative effects of Robot Social Engineering.

References

[Aba00] Frank W Abagnale. *Catch Me If You Can: The Amazing True Story of the Youngest and Most Daring Con Man [most Extraordinary Liar] in the History of Fun and Profit*. Broadway, 2000.

[AR16] Reuben Atlas and Jerry Rothwell. *Sour Grapes*. Movie. 2016.

[ARS17] Alexander Mois Aroyo, Francesco Rea, and Alessandra Sciutti. "Will You Rely on a Robot to Find a Treasure?" In: *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. Mar. 2017, pp. 71–72. DOI: 10.1145/3029798.3038394.

[ARSS18] Alexander Mois Aroyo, Francesco Rea, Giulio Sandini, and Alessandra Sciutti. "Trust and Social Engineering in Human Robot Interaction: Will a Robot Make You Disclose Sensitive Information, Conform to its Recommendations or Gamble?" In: *IEEE Robotics and Automation Letters* 3.4 (2018), pp. 3701–3708. DOI: 10.1109/LRA.2018.2856272.

[AW17] Siddharth Agrawal and Mary-Anne Williams. "Robot authority and human obedience: A study of human behaviour using a robot security guard". In: *Proceedings of the companion of the 2017 ACM/IEEE international conference on human-robot interaction*. 2017, pp. 57–58.

[Bea03] Beatbots. *Keepon*. https://robots.ieee.org/robots/keepon/?gallery=photo1. 2003. Accessed: 2020-03-02.

[Bek12] George A. Bekey. "Current Trends in Robotics: Technology and Ethics". In: *Robot Ethics: The Ethical and Social Implications of Robotics. MIT Press, Cambridge* (2012), pp. 17–34.

[BF04] Christoph Bartneck and Jodi Forlizzi. "A Design-Centred Framework for Social Human-Robot Interaction". In: *ROMAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication*. IEEE. 2004, pp. 591–594. DOI: 10.1109/ROMAN.2004.1374827.

[BFR14] Jenay M. Beer, Arthur D. Fisk, and Wendy A. Rogers. “Toward a Framework for Levels of Robot Autonomy in Human-Robot Interaction”. In: *Journal of Human-Robot Interaction* 3.2 (July 2014), pp. 74–99. DOI: 10.5898/JHRI.3.2.Beer.

[BMP+15] Jan-Willem H. Bullée, Lorena Montoya, Wolter Pieters, Marianne Junger, and Pieter H. Hartel. “The persuasion and security awareness experiment: reducing the success of social engineering attacks”. In: *Journal of Experimental Criminology* 11.1 (Jan. 2015), pp. 97–115. DOI: 10.1007/s11292-014-9222-7.

[BRYS15] Amy Banh, Daniel J. Rea, James E. Young, and Ehud Sharlin. “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. 2015, pp. 19–26. DOI: 10.1145/2814940.2814955.

[BTP+17] Serena Booth, James Tompkin, Hanspeter Pfister, Jim Waldo, Krzysztof Gajos, and Radhika Nagpal. “Piggybacking Robots: Human-Robot Overtrust in University Dormitory Security”. In: *Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction*. ACM/IEEE. 2017, pp. 426–434. DOI: 10.1145/2909824.3020211.

[Cle19] ClearPath Robotics. *Jackal*. https://clearpathrobotics.com/jackal-small-unmanned-ground-vehicle/. 2019. Accessed: 2019-10-27.

[cli19] @clippytheclip. *Clippy the Clip*. https://twitter.com/ClippyTheClip/. 2019. Accessed: 2019-10-27.

[CNN+13] Derek Cormier, Gem Newman, Masayuki Nakane, James E. Young, and Stephane Durocher. “Would You Do as a Robot Commands? An Obedience Study for Human-Robot Interaction”. In: *International Conference on Human-Agent Interaction*. 2013.

[CNN19] CNN Business. *Watch a CNN reporter get hacked*. YouTube. Oct. 2019. URL: https://www.youtube.com/watch?v=yIG4kTJTZuY. Accessed: 2020-02-25.

[Com12] Competition Bureau Canada. *The Little Black Book of Scams*. https://www.competitionbureau.gc.ca/eic/site/cb-bc.nsf/vwapj/lbbs-web-2017-eng.pdf/$file/lbbs-web-2017-eng.pdf. 2012. Accessed: 2020-01-12.

[CS16] Nabie Y. Conteh and Paul J. Schmick. “Cybersecurity: risks, vulnerabilities and countermeasures to prevent social engineering attacks”. In: *International Journal of Advanced Computer Research* 6.23 (2016), pp. 31–38. DOI: 10.19101/IJACR.2016.623006.

[Dar16] Kate Darling. “Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects”. In: *Robot law*. Edward Elgar Publishing, 2016. DOI: 10.4337/9781783476732.00017.

[DEF15] DEFCON. *Breaking In Bad*. Youtube. Dec. 2015. URL: https://www.youtube.com/watch?v=2vdvINDmlX8. Accessed: 2020-02-25.

[DMK+09] Tamara Denning, Cynthia Matuszek, Karl Koscher, Joshua R. Smith, and Tadayoshi Kohno. “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. 2009, pp. 105–114. DOI: 10.1145/1620545.1620564.

[EWC07] Nicholas Epley, Adam Waytz, and John T. Cacioppo. “On Seeing Human: A Three-Factor Theory of Anthropomorphism.” In: *Psychological review* 114.4 (2007), p. 864.

[Fed16] Federal Bureau of Investigation. *Elicitation Techniques*. https://www.fbi.gov/file-repository/elicitation-brochure.pdf/view. 2016. Accessed: 2020-01-11.

[FK11] Murph Finnicum and Samuel T. King. “Building Secure Robot Applications”. In: *HotSec*. 2011.

[FND03] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. “A Survey of Socially Interactive Robots”. In: *Robotics and Autonomous Systems* 42.3-4 (2003), pp. 143–166. DOI: 10.1016/S0921-8890(02)00372-X.

[For07] Jodi Forlizzi. “How Robotic Products Become Social Products: An Ethnographic Study of Cleaning in the Home”. In: *Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE. 2007, pp. 129–136. DOI: 10.1145/1228716.1228734.

[Fra14] Agence France-Presse. *Nestle Employs Fleet of Robots to Sell Coffee Machines in Japan*. Ed. by The Guardian. https://www.theguardian.com/technology/2014/dec/01/nestle-robots-coffee-machines-japan-george-clooney-pepper-android-softbank. Dec. 2014. Accessed: 2018-02-19.

[FYDS16] Francisco Erivaldo Fernandes, Guanci Yang, Ha Manh Do, and Weihua Sheng. “Detection of Privacy-Sensitive Situations for Social Robots in Smart Homes”. In: *Automation Science and Engineering (CASE), 2016 IEEE International Conference on*. IEEE. 2016, pp. 727–732. DOI: 10.1109/COASE.2016.7743474.

[GA17] Maartje Margaretha Allegonda de Graaf and Somaya Ben Allouch. “The Influence of Prior Expectations of a Robot’s Lifelikeness on Users’ Intentions to Treat a Zoomorphic Robot as a Companion”. In: *International Journal of Social Robotics* 9 (2017), pp. 17–32. DOI: 10.1007/s12369-016-0340-4.

[Gar07] Joel Garreau. *Bots on the Ground*. http://www.washingtonpost.com/wp-dyn/content/article/2007/05/05/AR2007050501009.html. May 2007. Accessed: 2019-09-07.

[GCB13] Michael A. Goodrich, Jacob W. Crandall, and Emilia Barakova. "Teleoperation and Beyond for Assistive Humanoid Robots". In: *Reviews of Human Factors and Ergonomics* 9.1 (2013), pp. 175–226. DOI: 10.1177/1557234X13502463.

[Had10] Christopher Hadnagy. *Social Engineering: The Art of Human Hacking*. John Wiley & Sons, 2010.

[HMW+09] Frank Hegel, Claudia Muhl, Britta Wrede, Martina Hielscher-Fastabend, and Gerhard Sagerer. "Understanding Social Robots". In: *2009 Second International Conferences on Advances in Computer-Human Interactions*. IEEE. 2009, pp. 169–174. DOI: 10.1109/ACHI.2009.51.

[IKO+02] Michita Imai, Takayuki Kanda, Testuo Ono, Hiroshi Ishiguro, and Kenji Mase. "Robot Mediated Round Table: Analysis of the Effect of Robot's Gaze". In: *Proceedings. 11th IEEE International Workshop on Robot and Human Interactive Communications*. IEEE. 2002, pp. 411–416. DOI: 10.1109/ROMAN.2002.1045657.

[JSB15] Shervin Javdani, Siddhartha S. Srinivasa, and J. Andrew Bagnell. "Shared Autonomy via Hindsight Optimization". In: *Robotics Science and Systems: Online Proceedings* 2015 (2015). DOI: 10.15607/RSS.2015.XI.032.

[Jur19] Steve Jurvetson. *Baxter*. https://en.wikipedia.org/wiki/Baxter_(robot). 2019. Accessed: 2019-10-27.

[Kel84] John F. Kelley. "An Iterative Design Methodology for User-Friendly Natural Language Office Information Applications". In: *ACM Transactions on Information Systems (TOIS)* 2.1 (1984), pp. 26–41. DOI: 10.1145/357417.357420.

[KHHW15] Katharina Krombholz, Heidelinde Hobel, Markus Huber, and Edgar Weippl. "Advanced Social Engineering Attacks". In: *Journal of Information Security and Applications* 22 (2015), pp. 113–122. DOI: 10.1016/j.jisa.2014.09.005.

[Kni20] Knightscope. *Knightscope Investor Web Page*. https://www.knightscope.com/invest. Feb. 2020. Accessed: 2020-03-03.

[LBSB11] Xin Luo, Richard Brody, Alessandro Seazzu, and Stephen Burd. "Social Engineering: The Neglected Human Factor for Information Security Management". In: *Information Resources Management Journal (IRMJ)* 24.3 (2011), pp. 1–8. DOI: 10.4018/irmj.2011070101.

[LGKI16] Phoebe Liu, Dylan F. Glas, Takayuki Kanda, and Hiroshi Ishiguro. “Learning Interactive Behavior for Service Robots - the Challenge of Mixed-Initiative Interaction”. In: *Proceedings of the workshop on Behavior Adaptation, Interaction and Learning for Assistive Robotics*. 2016.

[LH17] Josh Lipton and Megan Hawkins. *Robots Are Now Delivering Food in San Francisco*. https://www.cnbc.com/2017/04/21/robots-are-now-deliving-food-in-san-francisco.html. Apr. 2017. Accessed: 2018-02-19.

[LTFK11] Min Kyung Lee, Karen P. Tang, Jodi Forlizzi, and Sara Kiesler. “Understanding Users’ Perception of Privacy in Human-Robot Interaction”. In: *Proceedings of the 6th International Conference on Human-Robot Interaction*. ACM. 2011, pp. 181–182. DOI: 10.1145/1957656.1957721.

[MKF+12] Bilge Mutlu, Takayuki Kanda, Jodi Forlizzi, Jessica Hodgins, and Hiroshi Ishiguro. “Conversational Gaze Mechanisms for Humanlike Robots”. In: *ACM Transactions on Interactive Intelligent Systems (TiiS)* 1.2 (2012), p. 12. DOI: 10.1145/2070719.2070725.

[MMLV14] Francois Mouton, Mercia M. Malan, Louise Leenen, and Hein S. Venter. “Social Engineering Attack Framework”. In: *2014 Information Security for South Africa*. IEEE. 2014, pp. 1–9. DOI: 10.1109/ISSA.2014.6950510.

[MS11] Kevin D. Mitnick and William L. Simon. *The Art of Deception: Controlling the Human Element of Security*. John Wiley & Sons, 2011.

[MS14] Neil McBride and Bernd Stahl. “Developing Responsible Research and Innovation for Robotics”. In: *Proceedings of the IEEE 2014 International Symposium on Ethics in Engineering, Science, and Technology*. IEEE Press. 2014, p. 27. DOI: 10.1109/ETHICS.2014.6893392.

[Nic18] Greg Nichols. *Robot Fired From Grocery Store for Utter Incompetence*. https://www.zdnet.com/article/robot-fired-from-grocery-store-for-utter-incompetence/. Jan. 2018. Accessed: 2020-04-27.

[OSu19] Donie O’Sullivan. *We Asked a Hacker to Try and Steal a CNN Tech Reporter’s Data. Here’s What Happened.* https://www.cnn.com/2019/10/18/tech/reporter-hack/index.html. Oct. 2019. Accessed: 2020-04-27.

[Per19] Olímpio Pereira. *Roomba*. https://www.uihere.com/free-cliparts/robotic-vacuum-cleaner-roomba-irobot-robot-3366029. 2019. Accessed: 2019-10-27.

[PG18] Brittany Postnikoff and Ian Goldberg. “Robot Social Engineering: Attacking Human Factors with Non-Human Actors”. In: *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 2018, pp. 313–314. DOI: https://dl.acm.org/doi/10.1145/3173386.3176908.

[QPP+17] Davide Quarta, Marcello Pogliani, Mario Polino, Federico Maggi, Andrea Maria Zanchettin, and Stefano Zanero. “An Experimental Security Analysis of an Industrial Robot Controller”. In: *Security and Privacy (SP), 2017 IEEE Symposium on*. IEEE. 2017, pp. 268–286. DOI: 10.1109/SP.2017.20.

[Raj15] Monisha Rajesh. *Inside Japan’s First Robot-Staffed Hotel*. Ed. by The Guardian. https://www.theguardian.com/travel/2015/aug/14/japan-henn-na-hotel-staffed-by-robots. Aug. 2015. Accessed: 2018-02-19.

[Ret19] Rethink Robotics. *Baxter*. https://imgbin.com/png/hrAkD61u/baxter-industrial-robot-rethink-robotics-humanoid-robot-png. 2019. Accessed: 2019-10-27.

[RHW17] Paul Robinette, Ayanna M. Howard, and Alan R. Wagner. “Effect of Robot Performance on Human-Robot Trust in Time-Critical Situations”. In: *IEEE Transactions on Human-Machine Systems* 47.4 (2017), pp. 425–436. DOI: 10.1109/THMS.2017.2648849.

[Rie12] Laurel D. Riek. “Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines”. In: *Journal of Human-Robot Interaction* 1.1 (2012), pp. 119–136. DOI: 10.5898/JHRI.1.1.Riek.

[RLA+16] Paul Robinette, Wenchen Li, Robert Allen, Ayanna M. Howard, and Alan R. Wagner. “Overtrust of Robots in Emergency Evacuation Scenarios”. In: *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press. 2016, pp. 101–108. DOI: 10.1109/HRI.2016.7451740.

[Seo15] Stela Seo. *Poor Thing! Would You Feel Sorry for a Simulated Robot*. YouTube. 2015. URL: https://youtu.be/0EmllmX1Z0M. Accessed: 2020-03-18.

[SGGC07] Ja-Young Sung, Lan Guo, Rebecca E. Grinter, and Henrik I. Christensen. ““My Roomba is Rambo”: intimate home appliances”. In: *International Conference on Ubiquitous Computing*. Springer. 2007, pp. 145–162.

[SGN+15] Stela H. Seo, Denise Geiskkovitch, Masayuki Nakane, Corey King, and James E. Young. “Poor Thing! Would You Feel Sorry for a Simulated Robot?: A Comparison of Empathy Toward a Physical and a Simulated Robot”. In: *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM/IEEE. 2015, pp. 125–132. DOI: 10.1145/2696454.2696471.

[SLAD15] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. “Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust”. In: *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE. 2015, pp. 1–8.

[SLSS18] Nicole Salomons, Michael van der Linden, Sarah Strohkorb Sebo, and Brian Scassellati. “Humans Conform to Robots: Disambiguating Trust, Truth, and Conformity”. In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM. 2018, pp. 187–195. DOI: 10.1145/3171221.3171282.

[Sma16] Nicolas Small. “Assigned Responsibility: An Architecture for Mixed Control Robot Teleoperation”. PhD thesis. Murdoch University, Oct. 2016. URL: https://researchrepository.murdoch.edu.au/id/eprint/33736/1/whole.pdf.

[Sof19] Softbank Robotics. *NAO*. https://uwaterloo.ca/robohub/people-profiles/nao. 2019. Accessed: 2019-10-27.

[Son19] Sony. *Sony Aibo*. http://www.sony-aibo.com/. 2019. Accessed: 2019-10-27.

[Ste06] Valerie Steeves. “It’s not child’s play: The online invasion of children’s privacy”. In: *U. Ottawa L. & Tech. J.* 3 (2006), p. 169.

[Ste12] Michael Steinberger. *A Vintage Crime*. https://www.vanityfair.com/culture/2012/07/wine-fraud-rudy-kurniawan-vintage-burgundies. 2012. Accessed: 2020-01-15.

[SV78] Thomas B. Sheridan and William L. Verplank. *Human and Computer Control of Undersea Teleoperators*. Tech. rep. Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab, 1978.

[SY13] Ashish Singh and James E. Young. “A Dog Tail for Utility Robots: Exploring Affective Properties of Tail Movement”. In: *IFIP Conference on Human-Computer Interaction*. Springer. 2013, pp. 403–419. DOI: 10.1007/978-3-642-40480-1_27.

[the19] @theatropaths. *DARWiN-OP*. https://imgbin.com/png/KxFbhYQW/darwin-op-robotis-bioloid-humanoid-robot-dynamixel-png. 2019. Accessed: 2019-10-27.

[UQ14] Sven Uebelacker and Susanne Quiel. “The Social Engineering Personality Framework”. In: *2014 Workshop on Socio-Technical Aspects in Security and Trust*. IEEE. 2014, pp. 24–30. DOI: 10.1109/STAST.2014.12.

[Vin19] James Vincent. *Security Robots are Mobile Surveillance Devices, Not Human Replacements*. Ed. by The Verge. https://www.theverge.com/2019/11/14/20964584/knightscope-security-robot-guards-surveillance-devices-facial-recognition-numberplate-mobile-phone. Nov. 2019. Accessed:2020-02-22.

[Wan07] Jijun Wang. “Human Control of Cooperating Robots”. PhD thesis. University of Pittsburgh, 2007. URL: http://d-scholarship.pitt.edu/6269/1/Wang_EDC_2007-final2.pdf.

[Wat19] Waterloo Brewing Ltd. *Waterloo Brewing Identifies Cyberattack*. https://investorrelations.waterloobrewing.com/2019-11-21-Waterloo-Brewing-Identifies-Cyberattack. Nov. 2019. Accessed: 2020-01-14.

[YHSI09] James E. Young, Richard Hawkins, Ehud Sharlin, and Takeo Igarashi. “Toward Acceptable Domestic Robots: Applying Insights from Social Psychology”. In: *International Journal of Social Robotics* 1.1 (2009), pp. 95–108. DOI: 10.1007/s12369-008-0006-y.

[YSV+11] James E. Young, JaYoung Sung, Amy Voida, Ehud Sharlin, Takeo Igarashi, Henrik I. Christensen, and Rebecca E. Grinter. “Evaluating Human-Robot Interaction”. In: *International Journal of Social Robotics* 3.1 (2011), pp. 53–67. DOI: 10.1007/s12369-010-0081-8.