Experimental Social Engineering

Investigation & Prevention

Jan-Willem Bullee

EXPERIMENTAL SOCIAL ENGINEERING
INVESTIGATION AND PREVENTION

Jan-Willem Bullee

Graduation Committee

CHAIRMAN AND SECRETARY:

prof. dr.	P.M.G. Apers	University of Twente, The Netherlands

SUPERVISORS:

prof. dr.	P.H. Hartel	University of Twente, The Netherlands
prof. dr.	M. Junger	University of Twente, The Netherlands

CO-SUPERVISOR:

dr.	L. Montoya	University of Twente, The Netherlands

MEMBERS:

prof. dr.	R.J. Wieringa	University of Twente, The Netherlands
prof. dr.	E. Giebels	University of Twente, The Netherlands
prof. dr.	M.J.G. van Eeten	Delft University of Technology, The Netherlands
prof. dr.	W. Stol	Open University, The Netherlands
prof.	P. Ekblom	University College London, United Kingdom

predict
prioritise
prevent
TREsPASS

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 318003 (TREsPASS).

CTIT

CTIT Ph.D. Thesis Series 1381-3617, No. 17-443
Centre for Telematics and Information Technology
P.O. Box 217, 7500 AE Enschede, The Netherlands

ISBN: 978-90-365-4397-2
ISSN: 1381-3617
DOI: 10.3990/1.9789036543972

Typeset with LaTeX.
Printed by NetzoDruk Enschede BV.

EXPERIMENTAL SOCIAL ENGINEERING
INVESTIGATION AND PREVENTION

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. T.T.M. Palstra,
on account of the decision of the graduation committee,
to be publicly defended
on Friday, 6th of October 2017 at 14:45

by

JAN-WILLEM HENDRIK BULLEE

born on the 18th of March 1985
in Gorinchem, The Netherlands.

This dissertation is approved by:

Supervisors:
prof. dr. P.H. Hartel
prof. dr. M. Junger

Co-Supervisor:
dr. L. Montoya

Summary

Social engineering is the usage of social manipulation and psychological tricks to make the targets assist offenders in their attack. This practice manifests itself in e.g. phishing emails or cold call telephone scams. The aim of the thesis was to investigate the understanding of social engineering attacks in an organisational setting. In particular, the effectiveness both of the threat and the countermeasures were investigated. Three kinds of social engineering experiments were performed, each using a different modality (i.e. Face-to-Face (F2F), email and telephone). In each experiment, the targets (i.e. participants) were persuaded to perform actions that contribute to their victimisation. The subjects ($N = 162$) in the F2F experiment were visited by an offender in their offices and asked them to hand over their office keys. The subjects ($N = 593$) in the email experiment received a phishing email with the request to provide Personally Identifiable Information (PII). The subjects ($N = 92$) in the telephone experiment were persuaded to download and execute software from an untrustworthy website. A portion of the participants in both the F2F and telephone experiment received an intervention to reduce victimisation. The result was that 58.62% of those in the F2F experiment complied with the offender, compared to 36.96% who were priorly informed on how to detect and react to social engineering. In the telephone experiment, 40% complied with the offender, compared to 17.2% who received an intervention. Furthermore, 19.3% of those who received a generic phishing email complied, compared to 28.9% that received a spear phishing email. There was no effect of age, sex and using authority on victimisation found, whereas having had an intervention, receiving a spear phishing email and cultural background did have an effect. It is concluded that awareness raising about dangers, characteristics and countermeasures related to social engineering proved to have a significant positive effect on protecting the target. The research also shows that awareness-raising campaigns reduce the vulnerability only in the short term. In phishing emails, the use of a personalised opening sentence increases its success. The results of these experiments allow practitioners to focus awareness campaigns to maximise their effectiveness.

Samenvatting

Social engineering is het inzetten van sociale manipulatie en psychologische trucs om het doelwit te laten assisteren in de aanval. In het dagelijks leven zijn deze praktijken beter bekend als bijvoorbeeld email phishing of telefoon fraude. Het doel van dit proefschrift was om inzicht te krijgen in social engineering binnen een organisatie. In het specifiek is er onderzocht hoe groot het gevaar is en hoe sterk de tegenmaatregelen zijn. Er zijn drie soorten social engineering experimenten uitgevoerd, ieder met een andere modaliteit (Face-to-Face (F2F), email en telefoon). In alle experimenten zijn de deelnemers verleid om een actie uit te voeren waardoor zij slachtoffer werden. De proefpersonen ($N = 162$) in het F2F-experiment zijn door een aanvaller benaderd die hen vroeg hun kantoorsleutel te overhandigen. Verder hebben de proefpersonen ($N = 593$) in het email-experiment hebben een phishing email ontvangen met de vraag om persoonsgegevens af te staan. Tot besluit zijn er proefpersonen ($N = 92$) via de telefoon verleid tot het downloaden en uitvoeren van software afkomstig van een niet legitieme website. Een deel van de proefpersonen in zowel het F2F als het telefoonexperiment hebben vooraf een interventie ontvangen om de kans op slachtofferschap te verminderen. Het resultaat was dat 58.62% van de proefpersonen in het F2F-experiment de aanvaller gehoorzaamde, tegenover 36.96% die vooraf was geïnformeerd over hoe zij sociale engineering kunnen herkennen. In het telefoonexperiment, gehoorzaamde 40% de aanvaller, tegenover 17.2% die vooraf geïnformeerd was. Tot besluit, 19.3% van de proefpersonen die een algemene phishing email ontving gehoorzaamde de aanvaller, tegenover 28.9% die een gepersonaliseerde email kreeg. Bewustwording van gevaren, kenmerken en tegenmaatregelen gerelateerd aan social engineering bewees een significant effect te hebben op het neutraliseren van de aanvaller. Het onderzoek suggereert dat het effect van bewustwordingscampagnes alleen van korte duur is. Als het om phishing emails gaat, dan vergroot het gebruik van een gepersonaliseerde opening het succes. De resultaten uit de experimenten maakt het mogelijk om gerichte bewustwordingscampagnes uit te voeren en het effect te maximaliseren.

Acknowledgements

A wise man once said: "Jan-Willem, practising science is a group activity." The past four years taught me, among other things, that his person was correct. In the spirit of this lesson, I will dedicate this page to thank some people.

Everybody, thank you for your support throughout the years.
I could not have done it without you.

A special thanks goes to:
My promoters Marianne and Pieter. Thank you for your wisdom and guidance, it was a true pleasure working with, and learning from, you.

My daily supervisors: Lorena I have very much enjoyed our whiteboard sessions, sipping cappuccino and exchanging our views on the various topics.
Wolter, thanks for being there the first two years. It was great having you.

John, many thanks for the UK English editing of the papers this thesis is based on.

Bertine, Geert-Jan and Suse. Those morning coffees were a great start of the day.

All DIES and SCS PhD candidates and colleagues: Alexandr, Ali, Andreas, Arjan, Bence, Chong, Chris, Christoph, Dan, Dan, Didier, Dina, Eelco, Eleftheria, Elmer, Erik, Faiza, Hans, Herson, Inés, Joao Luiz, Jonathan, Klaas, Luuk, Luís, Maarten, Marco, Marten, Maya, Michael, Prince, Raymond, Riccardo, Roel, Roeland, Sandro, Soumik, Susanne, Tim, Wilbert and Yuxi.

The colleagues in the TREsPASS project.

Those who taught me psychology as well as computer science.

My friends. On top of that my two paranymphs: Robby and Tony.

My family for their unconditional love and support. Mom, Dad, Pim, you know me as a man of few words. Thanks :)

Last but not least, Jessica who was there for me the entire time <3.

–Jan-Willem Bullee, March, 2017

About the author

Figure 1: The Author

Jan-Willem Bullee studied both Computer Science and Psychology at the University of Twente. In his master project, he combined the two disciplines to find emerging leadership in small groups based on visually observable behaviour.
In January 2013 Jan-Willem joined the TREsPASS project as PhD candidate, where he is mainly involved in the work packages dealing with model specification and the data management process. His research interests include persuasion and deception in the context of cyber crime. In TREsPASS he uses experiments to gain insight into the behaviour of target and offender, and the preventive measures against fraudulent offences, in particular against social engineering.

Contents

Don't play the odds, play the man.

Harvey Specter – Suits

1

Introduction

This chapter is based on (Bullée, Montoya, Junger, & Hartel, 2017b).

The human is often the weakest element in information security systems (Orgill, Romney, Bailey, & Orgill, 2004; Happ, Melzer, & Steffgen, 2016). Offenders use deception and manipulation as an attack vector (the method that is used to execute an attack) to make their targets release information one should not have or make them perform malicious actions (Gupta, Agrawal, & Garg, 2011; Huber, Kowalski, Nohlberg, & Tjoa, 2009). This kind of dishonesty is called social engineering and is considered as the biggest threat to information systems (Rouse, 2006).

In the computer sciences, social engineering is associated with calling a target and asking for a password (Winkler & Dealy, 1995). Deception is used to make targets release information or perform a malicious action on behalf of the offender (Huber et al., 2009). Abraham and Chengalur-Smith (2010) use the following definition for social engineering: "*The use of social disguises, cultural ploys, and psychological tricks to get computer users (i.e. targets) to assist hackers (i.e. offenders) in their illegal intrusion or use of computer systems and networks*". However, social engineering could also be used more generally and described as a non-technical type of attack based on human interaction complementing technical attacks. What is considered as social engineering in the thesis is the use of deception and manipulation of the human element to assist the offender (Bosworth, Kabay, & Whyne, 2014; Dang, 2008). What does not constitute social engineering is physical violence, extortion, bribery, blackmailing and the like.

One of the dangers of social engineering attacks is their harmless and legitimate appearance so that targets (i.e. a person and not the goal of the attack) are unaware of being victimised (The Federal Bureau of Investigation, 2013; Hadnagy & Wilson, 2010). The result of a social engineering attack can be disastrous (Gupta et al., 2011). In March 2011, a group of offenders used social engineering techniques to hack into RSA Security, the company known for the RSA two-factor authentication tokens. The cost related with the breach was estimated at 66.3 million United States Dollar (USD). In the period after the breach, RSA's gross margins narrowed from 67.6% to 54.1% (King, 2011). Another aspect of the breach is the reputation damage RSA Security encountered. In the security business, one is only as good as one's reputation. The true impact of this attack is unknown, however it gave the competition an opening for closing in (Savage, 2012).

Social engineering has been used already for quite some time as an attack vector. Perhaps one of the oldest accounts is from the ancient Greek army, during the Trojan War. After being at war with the Trojans for ten years, the Greek army seemed to retreat from the battlefield and left the Trojans an enormous wooden statue of a noble horse. The Trojans welcomed their gift and celebrated their victory over the Greeks. During the night, Greek soldiers appeared from within the horse, and they conquered

the city (Graves, 1992). Myth or not, this example illustrates how the human element is abused to achieve a goal that was impossible by technical means.

A wide range of IT security practitioners annually gathers at the Black Hat conference. During the 2015 edition, 494 visitors from both Amsterdam and Las Vegas were asked about popular hacking methods. More than 80% stated that their favourite method is social engineering (Chmielewski, 2015). The ISACA and RSA Conference surveyed a global population of cyber security professionals (ISACA, 2015). The results of their survey revealed that 46.45% ($N = 327$) of the organisations was successfully exploited by social engineering and 68.32% ($N = 481$) by phishing (ISACA, 2015). Based on data gathered by Proofpoint during field research around the world, they concluded that in 2015, social engineering was the #1 attack vector. An offender's favourite way to beat cyber security is using people as substitute for exploits (Proofpoint, 2016, p. 2). This success is illustrated in the financial sector. During a period of two years, 100 banks from over 30 countries were victimised by an unknown group of offenders. Their *Modus Operandi* (MO) included the use of social engineering by persuading bank employees to open malicious email attachments that infected their PC. According to Law Enforcement Agencies, the estimated total loss was around 1 billion USD (Kaspersky, 2015). This shows that, social engineering is popular, effective and should be considered a serious threat. Therefore, the thesis aims to investigate the success of social engineering as an attack vector.

1.1 Social engineering explained with a scenario

Social engineering can have many forms and it is best researched in a specific context. The thesis focusses for practical reasons on social engineering in an organisation context. Consider Scenario 1.1 as an example of a social engineering attack (The SANS Institute, 2012, p. 2). The scenario will be dissected to explain social engineering and link it to theory. In particular, theories from crime science and social psychology will be used. First, the Routine Activity Theory (RAT) from crime science will be used to provide an insight in the occurrence of the crime. Second, persuasion principles, cognitive bias and other social psychology theories will be used to further explain the behaviour during a social engineering attack.

Scenario 1.1. Retrieve credit card details

> "You have been travelling and just checked into your hotel room. As you walk into your room and set your bag down, your phone rings. A nice girl introduces herself as Rebecca from the hotel front desk. She explains there has been an issue during check-in and she needs to re-confirm your credit card information. Assuming she is calling from the hotel front desk, you provide your credit card information. She then informs you everything has been resolved and to enjoy your stay."

Unfortunately, the person on the telephone was not Rebecca from the front desk. Instead, it was an offender ringing every hotel room in an attempt to victimise someone. What happened that made the traveller give away his credit card details to a stranger?

The Problem Analysis Triangle (PAT), an element from the RAT, considers three elements: *i*) offenders, *ii*) targets/victims and *iii*) places. A crime occurs when a likely offender and a suitable target converge in time and place plus there is a lack of capable guardians (Cohen & Felson, 1979). For each element, a protecting controller can be installed (Clarke, 1997). The controller of a likely offender (i.e. handler) is someone who knows that person well and can exert control over him/her. Common handlers are life partners, family, friends or probation and parole officers. Capable guardians are the controllers of the target or victim. Usually, these are the people protecting themselves, family members, public police or private security. Managers control places from becoming crime scenes; this usually is the owner or a representative (e.g. lifeguard or custodian) (Clarke, 1997). Each element will be discussed in more detail below.

In Scenario 1.1 the three elements of PAT can be identified. The offender is the person who called, the target is the traveller, and the location is the telephone in the hotel room that is answered by the traveller. Furthermore, from a psychological point of view four facets can be distinguished, two facets that relate to the offender and two that relate to the target: *i*) the offender using deception through impersonation, *ii*) the offender uses persuasion principles to strengthen the argument, *iii*) cognitive bias occurs at the target and *iv*) the target is unaware of appropriate defence Each facet will be discussed in detail below.

1.1.1 Deception through impersonation

Deception is essential in social engineering. Without deception, it would be a legitimate request for information. Buller and Burgoon (1996) provided the following definition for deception: *"Deception occurs when communicators control the information contained in their messages to convey a meaning that departs from the truth as they know it."*

Pretending to be someone else is one of the most popular methods used by social engineers (Bosworth et al., 2014). By using the identity of another person a particular role or authority is assumed. Common targets to be impersonated are system administrators, help desk employees and corporate executives. Offenders choose these specific roles since people trust them. Therefore, they are likely to assist the imposter in whatever they need (Bosworth et al., 2014).

The difficulty humans have in detecting lies explains both the success of deception and impersonation (Vrij, Granhag, & Porter, 2010). It is hard to detect lies, due to the absence of a single indicator that can be used reliably (e.g. Pinocchio's nose) (Vrij et al., 2010). To give an illustration on how hard lie detection is, a meta-analysis containing $k = 206$ studies with a combined $N = 24\,482$ subjects was conducted. The results showed that on average humans classify truths and lies 54% correct, i.e. almost random. No effect of the modality was found between audio, audiovisual and textual cues for detecting deception (C. F. Bond & DePaulo, 2006).

In Scenario 1.1, the offender lies about the identity of Rebecca from the front desk. By assuming this identity, the request for the credit card details is in context and therefore perceived as legitimate.

1.1.2 Persuasion principles

Once a person is a target, the offender can use social influences to change the odds of compliance in his favour. Six principles of persuasion can be used to increase the offender's probability of success: authority, conformity, reciprocity, commitment, liking and scarcity (Cialdini, 2009). For each persuasion principle, an explanation and a counter tactic is provided.

i) Authority is the principle that describes people's tendency to comply with the request of authoritative figures. If people are unable to make a thorough decision, the responsibility to do so is transferred to the group or someone they believe is in charge. Crisis and stress activate the behavioural trait of responsibility transition.

ii) Conformity or social proof, is imitating the behaviour of other people. The strength of conformity depends on group size and the majority performing a particular be-

haviour (R. Bond & Smith, 1996).

iii) Reciprocity refers to the giving of something in return. The target feels indebted to the requester for making a gesture, and even the smallest gift puts the requester in an advantageous position.

iv) Commitment refers to the likelihood of sticking to a cause or idea after making a promise or agreement. In general, when a promise is made, people have the tendency to honour it, which increases the likelihood of compliance.

v) Liking someone puts that person in a favourable position. People tend to like others who are similar regarding interests, attitudes and beliefs (Cialdini, 2009).

vi) Scarcity occurs when a product, service, or information has limited availability. People therefore perceive an increased value and attractiveness towards these products, which makes them more desired than others.

Muscanell, Guadagno, and Murphy (2014) describe the best practices to resist social influences (i.e. persuasion principles). Best practices to counteract the individual persuasion principles are the following questions:

i) Authority: When approached by an authority: "Is this person truly whom he claims to be?"

ii) Conformity: The fact that many others do something does not guarantee that it is a correct behaviour, hence: "Would I do the same if I was alone in this situation?"

iii) Reciprocity: "Why did I get this favour? Is this an act of kindness or part of a manipulation strategy?"

iv) Commitment: Evaluate all events as independently as possible: "Do I really want this?"

v) Liking: Separate the request from the person: "What would I say if the request came from a different person?"

vi) Scarcity: Once something is scarce, the perceived value increases: "Is this still an attractive offer if it wasn't scarce?"

In Scenario 1.1, the offender used the authority principle by impersonating Rebecca from the front desk. This is a tactical move from the offender since the front desk personnel is in charge of the guest registrations, room control and credit checks. Therefore it is likely that Rebecca knows this and is 'allowed' to have credit card information. An appropriate way to react is to think carefully about whom you are talking to. Section 1.1.4 discusses this in more detail.

1.1.3 Cognitive bias

The Dual-process accounts of reasoning theory, also known as System 1 and System 2 explains cognitive bias (Kahneman, 2011). System 1 is an automatic system that reasons very fast, unconsciously with low effort and is based on heuristics. System 2 on the other hand, is a controlled system that reasons consciously and is relatively slow (Kahneman, 2011). Since people do not have sufficient cognitive capacity to process all sensory input, their decision-making involves using rules of thumb (i.e. heuristics); this involves System 1 (Cialdini, 2009). Heuristics work well in most circumstances, until a heuristic fails and a cognitive bias occurs (Gigerenzer, 1991; Tversky & Kahneman, 1974).

A possible explanation of the traveller's behaviour is the availability heuristic. This is based on the first thought that comes to mind, given a particular context. The availability heuristic has the following underlying mechanism; if something is immediately recalled, it is important, or more important than the memories that were not immediately recalled. One consequence of the availability heuristic is that people will base their judgements on recent information (Esgate, Groome, & Baker, 2005). In Scenario 1.1, the traveller is in the context of a hotel and remembers that the credit card was used for payment.

Furthermore, the optimism bias also plays a role. People believe that positive events are more likely to occur to them than to other people (Weinstein, 1980). The inverse is also true: people believe that negative events are more likely to occur to other people than to themselves. The optimism bias is of persistent nature, while multiple attempts are made to reduce it, optimistic bias remains (P. Harris, Middleton, & Joiner, 2000). In Scenario 1.1, if the traveller was aware of the social engineering threat, a possible thought could be: "Someone else is more likely to be targeted, but if I am targeted, I will be more able to resist this threat than someone else."

1.1.4 Defence

It is hard to defend against something when someone is unaware of the threat. A first step in the reduction of the social engineering threat is to make people aware, both of the existence and on how to react (Bosworth et al., 2014). In Scenario 1.1, the traveller did not identify this episode as a potential threat or scam, and therefore no appropriate actions could be taken. Embedded in this unawareness is the 'truth bias', assuming that people don't lie to one another, based on the presumption that most communication is honest (McCornack & Parks, 1986). A more suitable approach would be to use the call-back approach. The rationale behind this approach is to ver-

ify the identity via trusted and verified paths (Gragg, 2003). This involves terminating the connection and calling the front desk yourself on a known number. Alternatively, one could invest more effort and physically visit the front desk in person. The preceding countermeasures are two examples of the target being its own guardian.

1.2 Urgency and impact

During the 2015 edition of the Cybersecurity & Privacy (CSP) forum, 44 cyber security professionals were surveyed regarding their understanding of Socio-Technical Cyber Threats (The TREsPASS Project, D1.3.3, 2015). In particular, it was asked what they perceived as the biggest threat of the past 15 years (i.e. since the year 2000) and what their thoughts are on the evolution of the threat in the future five years (i.e. 2015 - 2020). A Thematic Analysis was used to analyse the data. Regarding the future outlook, 100 (92 unique) responses were given by the participants. To get an impression of the responses, refer to Figure 1.1 for a word cloud.

Figure 1.1: Word cloud of survey responses regarding the socio-technical cyber threat themes of the next 5 years (i.e. 2015 - 2020). Adopted from (The TREsPASS Project, D1.3.3, 2015).

There were three main themes identified by the researchers, each with two or three sub-themes: *i*) Who is involved in an attack, describes who are thought to be involved in future attacks. *ii*) What is beneficial for the offender, describes the aspects an offender will use that contribute to a successful attack. *iii*) What is used in an attack, describes methods used by offenders in an attack. For an overview of themes and sub-themes, relate to Figure 1.2.

The gist of these findings is that the attacks will become more targeted. The focus shifts towards members of specific (e.g. controversial or minority) groups. It is ex-

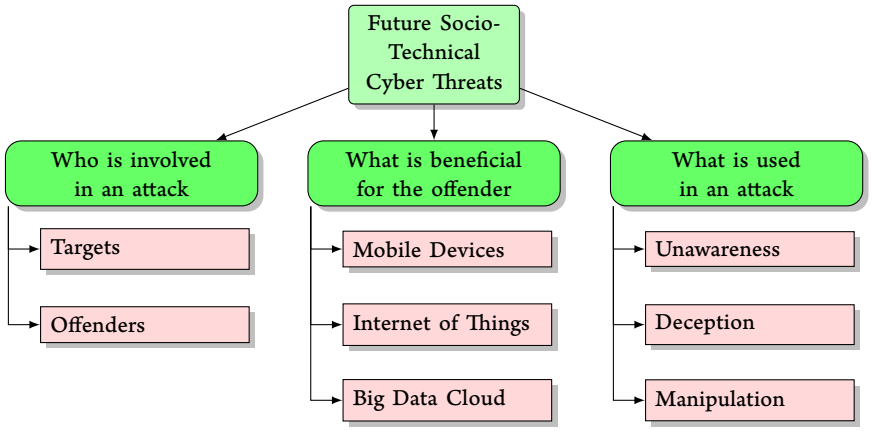


Figure 1.2: Overview of socio-technical cyber threat themes expected to emerge within the next 5 years (2015 - 2020). Adopted from (The TREsPASS Project, D1.3.3, 2015).

pected that offenders (e.g. organised crime) will adopt their MO to current methodology/technology. This means that mobile devices will become the modality for the attack. One way of doing this is by focussing on mobile applications (e.g. WhatsApp). Offenders will use information harvested from social media to make their attack to appear legitimate. The common prediction is that unawareness will still be a problem and offenders will misuse this. Finally, the attacks will result more often in identity thefts in both the physical and digital realm. In sum, attacks will become more targeted, using mobile communication and the outcome will be more severe.

Four out of the eight sub-themes that are expected to be relevant in the future, appear in the thesis. 'Big Data Cloud' illustrates the data sources an offender can utilise in an attack. All information regarding targets, in the experiments in the thesis, originates from public sources such as the corporate website. 'Unawareness' is addressed in the interventions in the experiments. 'Deception' relates to the trickery that offenders use, in the thesis both persuasion and impersonation are used. 'Manipulation' describes the type of attacks, the thesis discusses three types of attack (i.e. Face-to-Face (F2F), email and telephone).

These predictions were made in April 2015, two years later we see that some of these predictions became a reality. Two recent events will be discussed and reflected on the predictions. First, offenders use the mobile messenger application Whatsapp to distribute their scam. The victim receives a message from one of its contacts that encourage to click and receive a £100 worth of vouchers for the supermarket. Clicking the link allows the offender to install malicious software on the phone (McGoogan, 2016).

Second, in May 2017, a ransomware worm took hostage computers in 74 countries (Thomson, 2017). A spear phishing email containing a malicious attachment caused the initial infection. Three of the predictions made in 2015 can explain the infection of patient zero: *i*) Unawareness towards phishing; The initial infection was caused by a spear phishing email. *ii*) Offenders adapt their MO; The latest exploits were used to spread the ransomware across networks. *iii*) Specific targets; the ransomware targeted a specific group of organisations that used older software versions.

1.3 The success of social engineering in general

Social engineering involves interaction between people and can therefore happen using various communication channels (i.e. modalities). Known, but not limited to, channels for social engineering are email, telephone and F2F.

Social engineering via email (i.e. phishing), is systematically field tested in a multiplicity of contexts. The reported success rates are between 2.4% and 92.5% (Holm, Flores, Nohlberg, & Ekstedt, 2014; K. Hong, Kelley, Tembe, Murphy-Hill, & Mayhorn, 2013; Jagatic, Johnson, Jakobsson, & Menczer, 2007; Kearney & Kruger, 2014; Sheng, Holbrook, Kumaraguru, Cranor, & Downs, 2010; Tembe, Hong, Murphy-Hill, Mayhorn, & Kelley, 2013; Wright, Jensen, Thatcher, Dinger, & Marett, 2014). The broad range of successes suggests that the victimization by phishing is dependent on the content and context of the message. Furthermore, there are limited explanatory variables reported that could further explain the success, refer to Table 4.1. This will be discussed in more detail in Chapter 4.

In the computer sciences, social engineering is often associated with calling a target and asking for a password (Winkler & Dealy, 1995). Related to this is a telephone scam that has been carried out by offenders claiming to belong to Microsoft's technical support department (Arthur, 2010). By convincing the targets that there is something wrong with there PC, offenders manage to get access to bank accounts and defraud people. Regarding F2F social engineering, an example is the rogue interviewer who poses as an auditor and visits employees in the office. During the visit the auditor could asks questions regarding password usage where one of the questions half way could be 'what is your user name' and 'what is your password' (Orgill et al., 2004). Social engineering via email (i.e. phishing) relates to persuading someone to provide login credentials which can be used illicitly by the offender (Greening, 1996).

Literature shows that social engineering works (Mann, 2008) and that it is effective (Schellevis, 2011). Regarding telephone and F2F social engineering, there are many case studies (i.e. a scenario that is performed once) that describe it as easy and

successful, refer to e.g. (Mitnick & Simon, 2002; Mann, 2008). It is noted in a police investigation of a Nigerian 419 scam (a form of advance-fee fraud) that one of the limitations is that the only available data is in the form of incident reports, i.e. successful attempts (Schoenmakers, de Vries Robbé, & van Wijk, 2009). Although this information is interesting, this does not say anything about how successful the attack is, since the number of attempts is unknown. However, the number of incidents provides the lower bound. There is (to the best of our knowledge) no information on success rates available.

In sum, the success rate of email social engineering is known, however that of telephone and F2F social engineering is unknown. Furthermore, four psychological facets were discussed; deception through impersonation, persuasion principles, cognitive bias and defences. It is unknown to what extend these facets contribute to the success of social engineering. Therefore, based on the preceding, the research question was defined: "*What factors explain social engineering and how can these be used as a preventive measure?*"

1.4 Research objectives

The aim of the thesis was to investigate the understanding of social engineering attacks in an organisational setting. In particular, the effectiveness both of the threat and the countermeasures was investigated. The majority of the studies in the thesis are based on field tests in an organisation. The results provide a snapshot of the organisation and an insight into who can benefit most from participating in a training program.

The thesis provides:

- Systematic testing of the social engineering threat. The studies of social engineering in an office environment, using three different modalities (i.e. F2F, email and telephone). Each study mimicked a situation that could be performed by a real offender. The outcomes can be used by policy makers and related support departments to deploy training and awareness programs efficiently and provide help to those who need it most.
- Systematic testing of interventions that counter the success of social engineering. Two studies tested the use of countermeasures by providing general information and a subtle reminder to subjects in the social engineering studies. One study also tested the effect of time on the countermeasure. These outcomes are a first step in reducing the vulnerability of organisations and protecting their employees.
- An extension of attack trees to allow modelling of the social engineering threat

and prioritization on the basis of the data obtained by systematic testing. In information security, attack trees are used to assess security. An attack tree describes the security of systems in a tree structure, based on varying attacks. The goal of the attack is the root node and different ways of achieving that goal as child and leaf nodes. For an example of an attack tree, refer to Figure 6.1. In order to do a quantitative analysis, values need to be assigned to the leaf nodes of the tree. In a typical attack tree, attribute values are based on estimations using a 3-point scale (e.g. low-medium-high) (Bagnato, Kordy, Meland, & Schweitzer, 2012). Although this is an elegant method, it is prone to cognitive biases. The proposed solution allows to include data based on experiments to overcome the limitations.

1.5 Structure of the thesis

The thesis is divided into 7 chapters; each chapter is briefly introduced, and the relation with the other ones is described. The majority of the chapters contains material that has been published in scientific journals, at refereed conferences, or is in the process of being published. Therefore the chapters can also be read independently from each other. To the best of my knowledge, the content of the thesis is my own work. Where information has been derived from other sources, this has been indicated. For all chapters in the thesis, I am the main author. The co-authors provided supervision throughout the process, and this is made explicit at the beginning of each chapter.

Chapter 2 is a literature analysis that explores the extent to which persuasion principles are used in successful social engineering attacks. The scenario analysis illustrates how to exploit the human element in security. However, during the process new questions appeared: *i*) What is the success rate of social engineering attacks? This question is answered in Chapters 3, 4 and 5. *ii*) How effective are countermeasures to reduce the threat? Chapters 3 and 5 will give an insight into two different interventions. *iii*) How does combining persuasion principles influence the success rate of a social engineering attack? This is investigated in Chapter 3. *iv*) What is the influence of culture on the outcome of a social engineering attack? Chapter 4 provides an *a posteriori* inclusion on culture.

In Chapter 3, an experiment using F2F social engineering is presented. The aim of this study is to explore to what extent an intervention reduces the effects of social engineering. The conclusion is that awareness-raising about countermeasures associated with social engineering has a significant positive effect on neutralizing the offender. The question "What is the long term effect of an intervention?" appeared in the pro-

cess and is investigated in Chapter 5.

Chapter 4 describes a social engineering experiment that aims to explore how the opening sentence of a phishing email influences the outcome. Victimisation of phishing emails is well explained by four demographic variables: sex, age, Years of Service (YoS) and culture. There was one suggestion for improvement: Cultural scores were added *a posteriori*.

Chapter 5 describes how an information campaign counters the effects of social engineering via the telephone. The conclusion is that scam awareness-raising campaigns reduce vulnerability only in the short term.

Chapter 6 provides a methodology for including the results of social engineering experiments risk modelling tools. In particular an extension of attack trees is made that allows to include regression models in the analysis.

Finally, the results of the research presented in the thesis are reviewed in Chapter 7. A future outlook and suggestions for future research will be discussed subsequently.

For an overview of the questions and hypotheses in the thesis, refer to Figure 1.3. The main research question in the thesis is on the top of the tree structure. The level below is for the four chapters that contribute in answering the main question. For each chapter the questions and hypotheses are presented. Those in orange discuss the factors that explain social engineering, whereas those in red discuss the countermeasures.

For a summary of the factors and modalities investigated in the thesis, refer to Table 1.1. The effect on compliance for each factor is presented. Compliance refers to the target (i.e. the person who is in contact with the offender rather than the main goal of the attack) complying with the request of the offender. For each modality, two interpretations are discussed; from the offender and organisation perspective. Regarding the intervention in the F2F modality, there is a negative effect on compliance. This means that if employees received an intervention, the compliance rate would decrease. From an organisation's point of view, providing an intervention to employees will decrease both their vulnerability and that of the organisation. However, for an offender, it is advised to avoid those that have received an intervention, since the probability of succeeding is lower compared to those that did not get an intervention. Moreover using a spear (i.e. personalised) opening in the email modality, there is a positive effect on compliance. When using a spear opening, the compliance rate will increase. From an organisation's perspective, personalised social engineering emails are a bigger threat compared to non-personalised ones. It is advised to make this part of an awareness training within the organisation. For an offender it is bene-

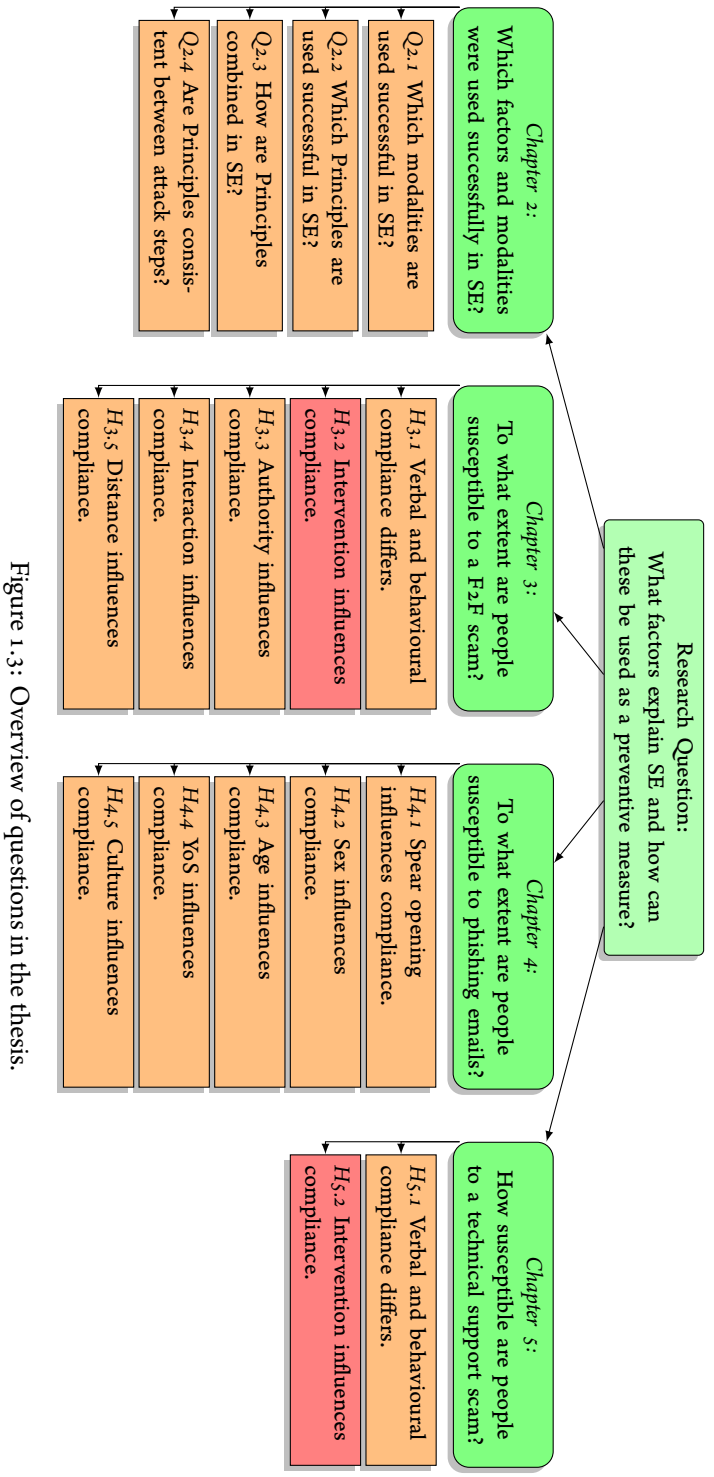


Figure 1.3: Overview of questions in the thesis.

ficial to use a personalised opening, hence a higher probability of succeeding. Finally, the intervention in the phone modality, there is a mixed effect on compliance. The time between the providing the intervention and testing matters. For those who received the intervention one week before the test, there was a negative effect on compliance. Those who received the intervention two weeks before the test, there was no effect on compliance. From the point of view of the organisation, those who received the intervention longer ago are more susceptible. Therefore it is advised to keep the preventive measure alive. From the offender, the probability of succeeding is lower when approaching someone who received the intervention only a short time ago.

Table 1.1: Factors influencing the outcome of social engineering attacks, in the thesis, divided into factors that relate to the target and offender. The number (in parentheses) refers to the chapter providing the details.

	FACTOR	MODALITY		
		F2F (3)	MAIL (4)	PHONE (5)
TARGET	SEX	None	None	None
	AGE	None	None	None
	CULTURE	–	Positive	–
	YEARS OF SERVICE	None	Negative	None
	INTERVENTION	Negative	–	Mix
	BUILDING	None	–	None
	DISTANCE	None	–	–
OFFENDER	AUTHORITY	None	–	–
	SPEAR	–	Positive	–
	SEX	None	–	None

– = Data not available;
None = no significant effect was found;
Positive = increases compliance;
Negative = decreases compliance;
Mix = compliance is moderated by third variable;

Throughout the thesis several datasets will be used. For an overview of the datasets, their names, number of observations and where these will be used, refer to Table 1.2.

Table 1.2: Summary of datasets used in the thesis.

NAME	UNIT[1]	*N*	VARIABLES[2]	SOURCE	CHAPTER[3]
`scenario`	interaction	142	11	Literature study	2
`survey`	subject	49	3	Questionnaire	3 & 5
`f2f-se`	subject	163	19	Field study	3
`mail-se`	subject	593	10	Field study	4
`phone-se`	subject	142	12	Field study	5

[1] UNIT refers to unit of analysis;
[2] VARIABLES refers to number of variables in the dataset;
[3] CHAPTER refers to the chapter where the dataset is used;

1.6 The sample size in experiments

While performing the experiments, the final sample size was never close to the expected sample size. Such 'limited' sample size has two implications: *i*) More effort was needed to obtain a sufficient sample size (i.e. recruiting another population and repeat the experiment) and *ii*) In the analysis only a limited number of independent variables could be used. Several stages caused the pool of subjects to shrink. For an overview of how the number of subjects shrank, refer to Table 1.3:

- *Context criteria* relates to the environment of the subject, e.g. using a particular type of door lock or having access to a desk phone.
- *Population criteria* describes the role of the subject, e.g. being a PhD-candidate or a Post-doctoral researcher.
- *In office* specifies the number of subjects that were visited in e.g. their office and asked if they wanted to participate as a research subject.
- *Consent signed* relates to the number of subjects that actual signed to informed consent and agreed to participate.
- *Subjects available* is the number of subjects that was available and did participate in the experiment.
- *Retract consent* describes the number of subjects that felt uncomfortable while participating. Therefore they decided to retract their consent and be no longer involved in the experiment.
- *Final N* relates to the sample size and number of subjects in the dataset that can be used for data analysis.

- *% subjects* is the percentage of subjects that ended up in the dataset compared to the *population criteria*.

Table 1.3: How the sample size of four populations shrank.

	COUNTRY AND MODALITY			
Factor	NL $F2F_1$	NL $F2F_2$	NL phone	SGP phone
Context criteria	583	337	385	300
Population criteria	354	195	266	150
In office	-	-	-	75
Signed Consent	-	-	-	36
Available at experiment	118	45	92	16
Retracted consent	-	-	-	1
Final *N*	118	45	92	15
% subjects	33.3	23.1	34.6	10

Oh, it's quite simple. If you are a friend,
you speak the password, and the doors will open.

Gandalf – The Lord of the Rings

2

On the anatomy of social engineering attacks

This chapter is based on (Bullée, Montoya, Pieters, Junger, & Hartel, 2017).

2.1 Introduction

Social engineering is the art of exploiting the weakest link in information security systems (i.e. the people who use them) (Bosworth et al., 2014). The targets are deceived to release information or perform a malicious action on behalf of the offender (Huber et al., 2009). In Chapter 1 social engineering is explained using crime science and social psychology theories. One of these social psychology theories involves social influence techniques. This chapter focusses on the social influence techniques offenders use in social engineering attacks to make their targets comply. Success stories from the social engineering literature were analysed to answer the question: "*Which modalities and factors have been used successfully in social engineering attacks?*"

Information security incidents are often caused by human failure (Chan, Woon, & Kankanhalli, 2005), rather than technical failure (Schneier, 2000). Since humans handle information systems, information security needs to take not only the technical but also the human element into account. The attack on the human element of security is called 'social engineering.' This technique consists of using social influences to convince people that the offender (e.g. social engineer) is whom he claims or pretends to be. The offender takes advantage of people to obtain information or knowledge he should not have (Gupta et al., 2011). Social engineering constitutes a security risk since it can be used to bypass Intrusion Detection Systems (IDS), firewalls and access control systems. One of the dangers of social engineering attacks is their harmless and legitimate appearance, so that targets are unaware of being victimized (The Federal Bureau of Investigation, 2013; Hadnagy & Wilson, 2010). The result of a social engineering attack can be disastrous (e.g. crippled corporate networks, identity theft or monetary loss) (Gupta et al., 2011).

Security experts have stated that as our culture becomes more dependent on information technology and technical prevention improves, social engineering will be the greatest threat to any security system (Rouse, 2006). Information security has traditionally been treated as a technical problem, resulting in information security teams being staffed solely by engineers (Rhee, Kim, & Ryu, 2009). This biased perspective of information security has led to the human factor being underestimated.

The field of human error, in a group or organizational context, has been widely researched (e.g. (Chikudate, 2009; Edmondson, 1996; Zhao & Olivera, 2006)). The type of human error that relates to unintentional errors is called a 'slip'. Slips refer to failures in executing the behaviour as planned and are mainly caused by attentional failure (Whittingham, 2004; Zhao & Olivera, 2006). Examples of slips are: *i*) a cashier forgetting to apply the discount to one product, *ii*) not properly closing a fire escape

or *iii*) leaving a USB key containing confidential data on the train. The individual knows how to perform the task, has the intention to do it, but doesn't do it properly (Zhao & Olivera, 2006). Another type of error is that induced by external parties with the intention to make one fail a procedure (i.e. offenders using influencing techniques to manipulate their targets into compliance). Examples are: *i*) sharing credit card details with a stranger, *ii*) providing physical access to someone without appropriate credentials or *iii*) supplying information that is needed for a spy to access the corporate network. The latter three examples relate to social engineering attacks.

Social engineering attacks can be explained using the analogy of the stack of Swiss cheese slices, known as the Swiss Cheese Model or the Cumulative Act Effect (Reason, 1990). This model depicts the various elements of an organisation as layers. In an ideal world these layers are intact, however in the real world, they resemble slices of cheese with holes. A single layer that has a hole poses no problem. However, if there are holes in multiple layers and these align, an attack can take place, i.e. the system is penetrable and therefore an employee could become a victim of social engineering. Offenders can use influence techniques to create holes in the cheese or to align them.

This study investigates the errors induced by external parties, in the context of information security. Following the analogy of the Swiss Cheese Model, this study aims to identify the number, characteristics and alignment of cheese slices (i.e. attack steps) which led to successful social engineering attacks (i.e. system failure).

From a psychological point of view, social engineering is part of *decision making*. Since people do not have the cognitive capacity to process all information, decision-making involves using rules of thumb (i.e. heuristics) (Cialdini, 2009). These mental shortcuts (resulting from experience and genetics) work well in most circumstances and increase the likelihood of solution on a task or problem. However, heuristics might not always lead to the correct solution and result in a cognitive bias instead (Gigerenzer, 1991; L. Harris, 2007; Lewis, 2012; Tversky & Kahneman, 1974). Social engineers (i.e. the offenders) are well aware of the flaws in human logic and nudge the heuristics of their targets into systematic errors (i.e. cognitive biases) to make them comply (Bosworth et al., 2014; Dang, 2008; Kennedy, 2011; Luo, Brody, Seazzu, & Burd, 2011; Twitchell, 2009).

Studies on traditional crime have analysed both offenders and victims. However, in computer-related crime, it is difficult to find offenders for research purposes. In criminology, it is common to analyse the crime from the perspective of the targets themselves. Information extraction techniques such as interviews or questionnaires can be used to get details about the crime. On the other hand, it is harder to get the the offender's point of view, which is especially the case when the target is unaware of

being victimized because the offender is unknown to him. This study aimed to gain an insight into social engineering attacks from an offender's perspective and the sources used are accounts by social engineers describing social engineering scenarios. The central question in this chapter is: *"How are social influences successfully used by social engineers to persuade their targets to comply with their requests?"*

Refer to Scenario 2.1 for a typical example of a social engineering attack from *The art of deception: Controlling the human element of security* (Mitnick & Simon, 2002).

Scenario 2.1. Free trip to San Francisco

> "Hi," says the voice at the other end of the line. "This is Tom at Parkhurst Travel. Your tickets to San Francisco are ready. Would you like us to deliver them, or would you like to pick them up?" "San Francisco?" says Peter. "I'm not going to San Francisco." "Is this Peter Abels?" "Yes, but I don't have any trips coming up." "Well," the caller says with a friendly laugh, "Are you sure you don't want to go to San Francisco?" "If you think you can talk my boss into it…" says Peter, playing along with the friendly conversation. "Sounds like a mix-up," the caller says. "On our system, we book travel arrangements under the employee number. Maybe somebody used the wrong number. What's your employee number?" Peter obligingly recites his number.

From this social engineering example it is clear that the caller tricks the employee into disclosing his employee number. By having this small piece of specific information, one has knowledge that can be used to mimic an insider. By using a small lie, the caller made the employee fail to execute the 'correct' behaviour (which in this case would be to callback the travel agency on a known number).

The first step in unravelling this complex topic is to find out how an individual can be persuaded to comply with malicious requests from offenders. By analysing offender accounts we aim to find commonly-used methods in social engineering attacks.

2.1.1 Principles of persuasion

Once a person is a target, the offender can use social influences to change the odds of compliance in his favour. Compared to gullibility (i.e. the willingness to believe someone or something in the absence of reasonable proof (Greenspan, 2008)), persuasion is a property of the offender, rather than of the target. A subset of six social influences (referred to as persuasion principles) were investigated by Cialdini (2009):

Authority, Conformity, Reciprocity, Commitment, Liking and Scarcity. For details regarding the persuasion principles, refer to Chapter 1.1.2.

Effectiveness of persuasion principles

Persuasion principles have been researched for several decades and there is empirical evidence that shows their effectiveness. Below we present a short overview of meta-analyses related to each persuasion principle.

The most famous study that illustrated *Authority* is the classical shock experiment performed by Stanley Milgram (1963). 66% of the participants did not hesitate to deliver a deadly dose of 450V to a human test subject, which they were instructed to perform by a man they believed to have legitimate Authority. An evaluation of 23 replication studies, conducted over the past 35 years, studied the obedience to Authority paradigm. Almost half of the studies (i.e. 11 studies) showed a lower rate of obedience compared to the initial study by Milgram, ranging from 28% to 65%. The remaining 12 studies showed an equal or higher rate of obedience, ranging from 66% to 91% (Blass, 1999). Although the studies found different rates of obedience, it can be concluded that Authority is persuasive.

In 1951 Salomon Asch conducted an experiment on the effects of *Conformity*. One of the outcomes was that 75% of the participants followed the majority, even when the majority stated an incorrect answer. This study was replicated many times in various contexts. A meta-analysis on Conformity, containing 133 studies showed it to be effective (R. Bond & Smith, 1996). The level of Conformity depends on the number of people (i.e. size of the majority) displaying a certain behaviour and can be estimated as: $I = \frac{1}{e^{4n}}$ where $n = \frac{1}{e^{N^{1.75}}}$. In this function I is the level of Conformity and N the majority size (Tanford & Penrod, 1984).

Reciprocity can be illustrated by means of reciprocal concessions, also known as the Door-in-the-Face (DitF) technique. Compliance is the result of first rejecting the extreme initial request and then asking for a moderate alternative. The requester changes the initial request, whereas the receiver feels inclined to change as well, resulting in the change from a "no" to a "yes" (Cialdini et al., 1975). A meta-analysis ($k = 22$ studies, $N = 3\,164$ subjects) showed that the DitF technique is effective. Those in the DitF group have higher odds of complying (with the requester) than those in the control group (Pascual & Guéguen, 2005).

Commitment is sometimes referred to as the Foot-in-the-Door (FitD) technique. Once a person has been induced to comply with a small request, he/she is more likely to comply with a larger demand (Freedman & Fraser, 1966). A meta-analysis ($k = 22$,

$N = 3\,124$) showed that the FitD technique is effective. Those in the FitD group have higher odds of complying than those in the control group (Pascual & Guéguen, 2005). The mechanisms behind DitF and FitD are diametrically opposed. A comparison ($k = 22$, $N = 3\,192$) to find out which technique is more effective showed that both are equally effective (Pascual & Guéguen, 2005).

The effect of *Liking* on self-disclosure has been investigated extensively, in particularly the question: "Do we disclose more to people we like?" A meta-analysis ($k = 31$) showed that people indeed have the tendency to disclose more personal information to people they like. Furthermore, there is little evidence that females and males differ in this respect (Collins & Miller, 1994).

Unlike the other principles, no meta-analysis was found for the Scarcity principle, therefore an individual study is discussed. *Scarce* products have an increased perceived value. The idea of not having something or a potential loss plays an important role in human decision making. Health researchers demonstrated this effect in a study involving breast cancer awareness pamphlets. The pamphlets stating the potential losses of not screening had significant more effect than pamphlets stating what the potential gain was (Meyerowitz & Chaiken, 1987).

The effect of Scarcity was compared to Conformity and a control condition, involving 153 university students in two emotional states: *i*) fear and *ii*) romance. Results showed that: *i*) the fear state Conformity appears to be more persuasive than the control condition (non-Conformity), *ii*) the fear state Scarcity appears to be less persuasive than the control condition (non-Scarcity), *iii*) the romance state Scarcity appears to be more persuasive than the control condition (non-Scarcity) and *iv*) the romance state Conformity appears to be less persuasive than the control condition (non-Conformity). Thus, fear improves the effect of Conformity and romance improves the effect of Scarcity (Griskevicius et al., 2009).

For a summary of meta-analytical findings, refer to Table 2.1. The *p*-value indicates whether there is sufficient statistical evidence that the principle was effective compared to the controls. This was the case for the Conformity, Reciprocity, Commitment and Scarcity principles. For Authority and Liking no such statistic was reported in the literature.

Based on Table 2.1, the persuasion principles were ranked based on the compliance rate (i.e. effectiveness): *i*) Authority, *ii*) Commitment, *iii*) Reciprocity, *iv*) Conformity. Authority is therefore the persuasion principle with the strongest effect.

The persuasion principles seem to be effective, although none of the studies found were focussed on the IT context. This research aims to answer the question: *"How are social influences successfully used by social engineers to persuade their targets to comply with their requests?"* One expectation was that in social engineering attacks, some persuasion principles would be used more often than others.

Modality as moderator of persuasion

Social engineering is has several modalities (Winkler & Dealy, 1995). Research on persuasion principles in the psychology literature shows that they are sometimes used over the telephone, whilst sometimes they are used Face-to-Face (F2F). An overview of differences in modality is presented below.

The influence of modality is illustrated in a meta-analysis on the DitF persuasion strategy. This study compared ($k = 56$) studies using the F2F modality with ($k = 31$) studies using the telephone modality. The combined result of $N = 7\ 641$ subjects, showed no statistical difference between the two (O'Keefe & Hale, 2001).

A meta-analysis on modality found no effect on compliance for both verbal ($k = 78$, $N = 7\ 138$) and behavioural ($k = 39$, $N = 3\ 125$) compliance (Feeley, Anker, & Aloe, 2012). Verbal compliance refers to saying that you will do something, whilst behavioural compliance to actual doing so. This means that modality does not 'moderate' the relationship between strategy and compliance.

In a variation of the Milgram experiment (refer to section '*Effectiveness of Persuasion Principles*'), the Authority (i.e. experimenter) left the room and gave instructions via the telephone. The results are an obedience rate of 20.5%, a significantly lower outcome compared to the 66% for the F2F condition (Milgram, 1965, 1974). The power of the Authority seems to drop severely when there is no F2F contact.

The influence of modality was tested on a combination of product judgement (i.e. recall, attitude and behavioural intentions) and type of goods (i.e. a product or a service) offered for sale (Szymanski, 2001). The product and the service were both presented via a F2F or telephone modality. The telephone modality had statistically significant higher scores (i.e. more influence) compared to the F2F modality. For goods, on the other hand, no difference was found in modality for behavioural intentions (Szymanski, 2001).

The results show that for some persuasion principles, modality 'moderates' the re-

Table 2.1: Overview of meta-analyses scores for each persuasion principle.

PRINCIPLE	*N*	*k*[1]	COMPLIANCE[2]	STATISTIC	*p*[3]	REFERENCE
Authority	1 041[4]	23	62.5 [28.0 - 91.0]	-	-	(Blass, 1999)
Conformity	4 627	133	28.8 [2.1 - 60.1]	$d = 0.92$	***	(R. Bond & Smith, 1996)
Reciprocity	3 164	22	41.1 [3.2 - 84.5]	$t(21) = 2.26$	**	(Pascual & Guéguen, 2005)
Commitment	3 124	22	45.2 [3.2 - 100]	$t(21) = 2.71$	**	(Pascual & Guéguen, 2005)
Liking	-	31	-	$d = 0.72$	*	(Collins & Miller, 1994)
Scarcity[5]	311	1	-	$F(1, 305) = 5.34$	**	(Griskevicius et al., 2009)
Reciprocity *vs* Commitment	3 192	22	-	$t(21) = 0.1$	ns	(Pascual & Guéguen, 2005)

* $p < .05$; ** $p < .01$; *** $p < .001$;

[1] k indicates the number of studies included in the meta-analysis;

[2] Average unweighed percentage of compliance [the number between brackets indicates the min and max];

[3] The p-value tests whether the principle was effective in relation to the controls;

[4] N estimated by author of the present study based on available studies;

[5] Operationalised as romance;

lation between request and compliance, refer to Table 2.2. These mixed outcomes also indicate that there are other variables of influence and that the operationalisation is important.

Table 2.2: Overview of meta-analyses scores for F2F vs phone.

PRINCIPLE	MODERATOR	N	k[1]	p[2]	REFERENCE
Authority	-	80	1	***	(Milgram, 1965)
Reciprocity DitF	-	7 641	87	ns	(O'Keefe & Hale, 2001)
Reciprocity DitF	Verbal	7 138	78	ns	(Feeley et al., 2012)
Reciprocity DitF	Behavioural	3 125	39	ns	(Feeley et al., 2012)
Sales presentation	Goods	146	1	*	(Szymanski, 2001)
Sales presentation	Service	114	1	ns	(Szymanski, 2001)

* $p < .05$; ** $p < .01$; *** $p < .001$;
[1] k indicates the number of studies included in the meta-analysis;
[2] The p-value tests whether the principle was affected by the modality;

Applying persuasion principles to bypass security

Research on persuasion principles is commonly associated with the social sciences (e.g. refer to (Blass, 1999; R. Bond & Smith, 1996; Pascual & Guéguen, 2005; Collins & Miller, 1994)). We next provide an example to illustrate how an offender could use each persuasion principle in an IT-related office environment setting:

Authority An offender claims that he works in the IT department, or that he is an executive in the company (Mitnick & Simon, 2002).

Conformity The offender calls a target, says that he is conducting a survey and names other people from the organization who already participated. The survey embeds a series of questions that draw the victim into revealing the user-name and password.

Reciprocity The offender calls a target and identifies himself as an employee from the IT department. The offender explains that there was a computer virus and that some data servers have to be restored. It would help the employee, if the target provides the password. The target is more likely to reciprocate since the caller offers him help.

Commitment An employee is called from the finance department and asked for something small. The employee is called again later and asked for something bigger; this process is continued until the goal is reached.

Liking An offender acts friendly, gives compliments and states that he studied at the same university as the target, hence emphasizing similarities.

Scarcity A call centre employee rings and gives a fake sales pitch. The target is not in-

terested and wants to hang up the phone, but the offender rapidly mentions that the first 100 people who register on a given website will win two vouchers for a particular event.

These six short scenarios are examples of how an offender can influence a target. The offender can use the six principles of persuasion to achieve a higher probability of compliance.

2.1.2 Crime scripts

Understanding how offenders use social influences to convince their targets to comply is a key element in dissecting social engineering attacks. To dissect an attack into steps, we use the concept of the 'crime script'. Crime scripts relate to *attack templates* and in technical research areas, crime scripts are referred to as *Attack life cycles* (Marconato, Kaaniche, & Nicomette, 2012). Cornish (1994) argued that crime scripts can be used to understand how a crime is executed. Crime scripts consist of sequential 'script functions' and accompanying 'script actions'. These scripts organize the knowledge and understanding of the routine behavioural process; preconditions, initiation, actualisation, doing and post-conditions (Cornish, 1994). Breaking down crimes into small steps increases the comprehension of the crime as a whole. Each step in the crime script resembles a slice in the Swiss Cheese model. However, this is not necessarily a 1-to-1 relationship since the Swiss Cheese model relates to defender mechanisms and it is possible that an offender bypasses several defence layers at once using a single attack step (or vice-versa). Moreover, this dissection is useful for designing specific interventions.

Crime scripts have already been developed many times, for example in the field of resale of stolen vehicles (Tremblay, Talon, & Hurley, 2001), organized crime (Rowe, Akman, Smith, & Tomison, 2012), illegal waste activity (Thompson & Chainey, 2011), serial sex offences (Beauregard, Proulx, Rossmo, Leclerc, & Allaire, 2007) and drug manufacturing (Chiu, Leclerc, & Townsley, 2011) and whereas templates were used to identify rapists' target selection patterns (Beauregard, Rebocho, & Rossmo, 2010). However, to the best of our knowledge, there is no crime script analysis on social engineering attacks from the viewpoint of the offender. We therefore investigate the social influences used by an offender in a social engineering attack.

2.1.3 Experimental context

It is shown in literature that social engineering works (Mann, 2008), that it is effective (Schellevis, 2011) and that targets are often unaware of being victimized (The Federal

Bureau of Investigation, 2013; Hadnagy & Wilson, 2010). However, there is limited information available on the inner working of such attacks. Penetration testing reports occasionally surface, but these represent the proverbial needle in a haystack and are 'uncontrolled' in the statistical sense. Books on social engineering (Hadnagy & Wilson, 2010; Mann, 2008; Mitnick & Simon, 2002; Mitnick, Simon, & Wozniak, 2011) provide an insight into how successful social engineering attacks are executed. To the best of our knowledge, no studies exist that identify social influences in social engineering attack scenarios. Therefore, in this chapter social engineering cases are dissected to find out what social influences are used.

2.1.4 Research question

The objective of the present research was to find out: "*Which modalities and factors have been used successfully in social engineering attacks?*" Four questions and sub-questions were formulated:

Q1) Literature indicated that social engineering appears in different modalities, the first question is: "What modality is used to execute social engineering attacks?"

Q2) Since, research showed the effectiveness of principles of persuasion, we aim to find out: "Which persuasion principles are used in the context of social engineering attacks?"

Q3) To the best of our knowledge, there is no literature describing the effect of combining persuasion principles on compliance. We ask the question: "How are the persuasion principles combined within a single step in social engineering attacks?"

- **Q3.1)** "How many persuasion principles are used in an attack step?"
- **Q3.2)** "How many attack steps are used in a crime script?"
- **Q3.3)** "How are persuasion principles used in the course of a crime script?"

Q4) To the best of our knowledge, there is no crime script analysing social engineering. To clarify how such a script looks like over time, we ask the following question: "What is the consistency among persuasion principles between two steps in the crime script?"

- **Q4.1)** "Do principles differ over the course of a crime script?"
- **Q4.2)** "How related are two consecutive attack steps within a crime script?"
- **Q4.3)** "How related are the first and last step of a crime script?"

2.2 Method

The data set that is used in this chapter is constructed by reading books and analysing their content.

2.2.1 Data selection

Goodreads is a social cataloging website where book lovers can review, rate and catalogue what they have read. It currently contains 900 million books and 34 million reviews. There are over 100 books on Goodreads' bookshelf related to 'social engineering'.[1] The following book inclusion criteria were used: *i*) social engineering involves a non-technical social attack against the operator of a computer system and, *ii*) the book contains case studies illustrating the use of social engineering. The top 4 of books most often identified as 'social engineering' by the community of Goodread members and met the two inclusion criteria were included in the analysis: *i*) *The Art of Deception: Controlling the Human Element of Security* (Mitnick & Simon, 2002), *ii*) *Ghost in the Wires: My Adventures as the World's Most Wanted Hacker* (Mitnick et al., 2011), *iii*) *Hacking the Human: Social Engineering Techniques and Security Countermeasures* (Mann, 2008) and *iv*) *Social Engineering: The Art of Human Hacking* (Hadnagy & Wilson, 2010).

According to Scopus[2] (accessed on November 2016), the four books together are cited 335 times in scientific publications; 277, 13, 7 and 38 times respectively. Compared to books written by other computer hackers, only *The Art of Deception* can be considered as highly cited (i.e. 277 times), whilst *Social Engineering: The Art of Human Hacking* can be considered as moderately cited. The story of Richard Jones is told in *Underground: Tales of Hacking, Madness and Obsession on the Electronic Frontier* (Dreyfus & Assange, 2012) and was cited 10 times. The memoires of Julian Assange, as in *Julian Assange: The Unauthorised Autobiography* (Assange, 2011) is cited 7 times. Former black hat hacker Kevin Poulsen wrote the book *Kingpin: How One Hacker Took Over the Billion-Dollar Cybercrime Underground* (Poulsen, 2011), which is cited 13 times.

In each book in the analysis, social engineering is illustrated by means of success scenarios, executed by the authors, their friends or other professionals from the field. The present analysis includes only scenarios that contain interactions between at least two humans. Scenarios that consist only of pure hacking, such as obtaining access to information via an open port on the server found using a port-scanner were excluded because these lack human interaction. In addition, we excluded forms of threatening such as blackmailing and other human forms of interactions such as bribing, as we don't consider these to constitute social engineering. What is considered as social engineering is the use of deception and manipulation of the human element to release

[1] https://www.goodreads.com/shelf/show/social-engineering
[2] http://www.scopus.com

information to the offender (Bosworth et al., 2014; Dang, 2008; Huber et al., 2009). In total there were 74 social engineering scenarios extracted for the analysis.

Quality of the books and ecological validity (i.e. realism)

An internet search was performed to assess the quality of the four books and their ecological validity (i.e. their degree of realism). The ratings of the four books found in three on-line sources (i.e. `books.google.com`, `amazon.com` and `goodread.com`) were analysed. The ratings show that more than 70% of the readers rated the books as being either good or very good. Some of the scenarios that originate from *The Art of Deception: Controlling the Human Element of Security* were described as fictionalized (Mitnick & Simon, 2002). For the other three books, there was no evidence found about either fabrication or fictionalization of stories, based on the book reviews at Goodreads and Amazon.

In an attempt to obtain more insight into the extent to which scenarios in *The Art of Deception* were fictionalized: *i*) the online legal research database Westlaw was queried and *ii*) the author was contacted. The results show that Kevin Mitnick was convicted two times: the first time in 1988 and a the second time in 1996. In 1988 he was charged with 'possession of unauthorized access devices' (in violation of 18 U.S.C. §1029(a)(3)), and computer fraud (in violation of 18 U.S.C. §1030(a)(4)). He pled guilty and was sentenced to twelve months in prison followed by a three year period of supervised release under special conditions. The company where the computer fraud was performed is mentioned in the book *The Art of Deception* (Mitnick & Simon, 2002); however, there are insufficient details in the court documents to validate the method (*USA v Mitnick*, 1998). In February 1995 Mitnick was charged with 'access device fraud' (in violation of 18 U.S.C. §1029) and computer fraud (in violation of 18 U.S.C. §1030). A twenty-five count indictment against him alleging 'computer fraud', 'wire fraud' and 'interception of wire communications' relating to additional alleged illegal conduct by him was presented in September 1996. This indictment contains a list of organisations which are mentioned in the scenarios used in *Ghost in the Wires* (Mitnick et al., 2011). Similarly, no details were found in the court documents to validate the method (*USA v Mitnick*, 1996). The court documents obtained via the Westlaw online legal research provided no additional insight into the scenarios. What was therefore found from this validation investigation was that some of the companies mentioned in the books' scenarios also appear in the indictments, therefore it seems plausible that the book *The Art of Deception* is not a work of fiction. Contacting the author was unsuccessful, since no reply on our email was received.

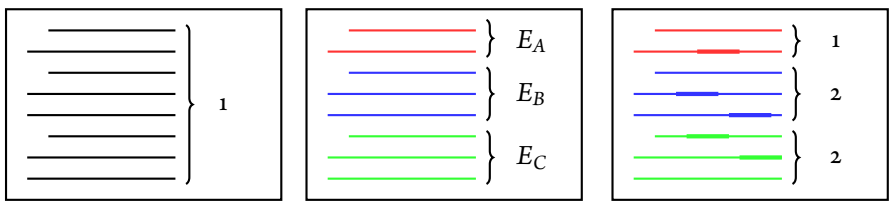
2.2.2 Readers

All 74 scenarios were independently coded by two researchers. The first researcher is a 29 year old male PhD-candidate with a background in both psychology and computer science. The second researcher is a 22 year old female student assistant with a background in information and communication sciences. An inter-rater reliability analysis using the Kappa statistic was performed to determine the consistency among researchers.

2.2.3 Procedure

To ensure agreement between the researchers, they both: *i*) processed a description of the persuasion principles, *ii*) performed coding on a test dataset of 5 scenarios and *iii*) discussed the outcome of the training results. The description of the principles was the same as that of Chapter 1.1.2. The set of training scenarios was a random selection from all scenarios.

After the training, both readers agreed that analysing a scenario as a whole (refer to Figure 2.1a) would bias the results, since multiple people can be involved and an offender can approach each individual differently. Therefore, it was decided to split the scenarios into attack steps, each containing a single interaction between two individuals. For example, if the offender first talked to *EmployeeA* and next to *EmployeeB* and finally to *EmployeeC*, the scenario is split into three attack steps (refer to Figure 2.1b). The persuasion principles used by the offender were coded for each attack step (refer to Figure 2.1c). Figure 2.1c shows 3 interactions containing 1, 2 and 2 persuasion principles respectively.



(a) One Scenario (b) Three attack steps (c) Five persuasion principles

Figure 2.1: The process of dissecting a scenario into persuasion principles. A scenario is first split in attack steps (i.e. interactions). For each interaction the persuasion principles are identified.

All the scenarios were coded twice, meaning that the work was not split and concatenated afterwards but that instead, the resulting dataset consists of consensual results. After coding all attack steps, the inter-rater agreement was calculated. The scores

of both readers were compared to generate the final dataset. If both readers identified the same principles for a given attack step, there was consensus. However, when there was a difference in the codes, the readers discussed the different views and came to a conclusion (the majority of differences related to one coder accidentally marking the wrong principle).

Figure 2.2 shows the dissection of a 'real' scenario from (Mitnick & Simon, 2002). The scenario contains 2 attack steps, where each attack step contains 1 persuasion principle. In both attack step 1 and 2 the offender uses impersonation together with the authority principle. In attack step 1, this was achieved by claiming to be an attorney, whilst in attack step 2 by claiming to be a staff member from the R&D department. In both attack steps authority was operationalised by means of titles.

2.2.4 Variables

There are 2 kinds of variables in the analysis, those related to: *i*) the crime script and *ii*) the attack step. The variables related to the crime script are: 'modality' and 'steps'. The variables related to the attack step are: 'persuasion principles' (6 variables), 'other', 'first / last' and 'former / latter'. The six dichotomous variables *persuasion principle* were dummy coded as 0 = not used, 1 = used. In case none of the 6 persuasion principles seemed appropriate, the variable *other* recorded other social influence tactics the offender used to deceive the target. The variable was a string variable, hence allowing an open-ended response. In the open-ended responses, there was one suggestion given related to 'Overloading'. Overloading can be used while conducting a questionnaire by putting the trick question between other questions. Due to the amount of information the brain has to process, there is a transition to a passive mental state, in which information is absorbed rather than evaluated (Janczewski & Colarik, 2008). The *modality* variable was a string variable recording the modality used by the offender (e.g. F2F or Telephone). The *steps* variable was an integer recording the number of attack steps in each crime script (1 = there was a single attack step in the crime script). The categorical variables *first / last* contained the persuasion principle(s) used in the first and last attack steps of a crime script. The categorical variables *former / latter* contained the persuasion principle(s) used in a chronological attack-step-pair of a crime script. All variables were stored in the `scenario` dataset, refer to Table 1.2.

2.2.5 Analysis

The first question *(i.e. What modality is used to execute social engineering attacks?)* involves comparing the frequencies of different modalities to perform social engineer-

Cracker Robert Jorday had been regularly breaking into the computer networks of a global company. The company eventually recognized that someone was hacking into their terminal server, and could connect to any computer system at the company. To safeguard the corporate network, a dial-up password was required on every terminal server.

Robert *called*[a] the Network Operations Center `posing`[b] as an attorney with the **Legal Department**[c] and said he was having trouble connecting to the network. The network administrator explained that there had been some recent security issues, so all dial-up access users would need to obtain the monthly password from their manager. Robert wondered what method was being used to communicate each month's password to the managers and how he could obtain it.

It turned out that the password for the upcoming month was sent in a memo via office, mail to each company manager.

[a]Using the phone modality
[b]Impersonation
[c]Authority

Robert *called*[a] the company after the first of the month, and reached Janet, the secretary of a manager. He said, "*Janet, hi.* `This is Randy Goldstein`[b] *in* ***Research and Development***[c]*. I know I probably got the memo with this month's password for logging into the terminal server from outside the company but I can't find it anywhere. Did you get your memo for this, month?*"

Yes, she said, she did get it.

He asked her if she would fax it to him, and she agreed. He gave her the fax number of the lobby receptionist in a different building on the company campus. He had already made arrangements for faxes to be held for him, and be forwarded to an on-line fax service. When this service receives a fax, the automated system sends it to the subscriber's email address.

The new password arrived at the email dead drop that Robert set up on a free email service in China. Best of all, he never had to show up physically at the location of the fax machine.

[a]Using the phone modality
[b]Impersonation
[c]Authority

Figure 2.2: An example of a dissected scenario. The scenario contains 2 attack steps, each step has 1 persuasion principles. Furthermore, the offender uses the telephone to contact the targets.

ing attacks. The variable 'Modality' was tested using Cross Tabulation and Chi Square analysis[3].

The second question *(i.e. Which persuasion principles are used in the context of social engineering attacks?)* involves the frequencies of persuasion principles used. The variable 'Persuasion Principles' was recoded into a single variable (i.e. an entry for each occurrence of the persuasion principle) and was tested using Fisher's Exact test[4].

The third question *(i.e. How are the persuasion principles combined within a single step in social engineering attacks?)* contains three sub-questions: *i*) "How many persuasion principles are used in an attack step?", *ii*) "How many attack steps are used in an crime script?" and *iii*) "How are persuasion principles used in the course of a crime script?" The aim of the question is to get an insight in how social engineering attacks are executed, which is relevant for developing countermeasures. The first sub-question involves the variable 'Persuasion Principles', which was recoded into a new variable containing the number of persuasion principles used for each attack step and it was tested using Fisher's Exact test[3]. The second sub-question compares path lengths using the 'Steps' variable and it was tested using Fisher's Exact test[4]. The third sub-question involves the variables 'Persuasion Principles' and 'Steps' and it was tested using Fisher's Exact test[4]. In total there were 4 analyses performed, one for each persuasion principle (except for Scarcity and Conformity due to insufficient observations).

For the fourth question *(i.e. What is the consistency among persuasion principles between two steps in the crime script?)* contains 3 sub-questions: *i*) "Do principles differ over the course of a crime script?", *ii*) "How relate two consecutive attack steps within a crime script?" and *iii*) "How relate the first and last step of a crime script?" These sub-questions enable a better understanding of the crime scripts. The first sub-question provides an insight into the use of persuasion principles at different stages of the attack and involves the variables 'Persuasion Principle' and 'Steps'. The variable Persuasion Principle was recoded as 0 = not using persuasion principle, 1 = using persuasion principle. This was tested using Cross Tabulation and Chi Square.

The second and third sub-questions are used to analyze changes in tactics at a micro attack level. The second sub-question compares two consecutive attack steps. It involves the variables 'Former / Latter' and it was tested using Fisher's Exact test[4]. The third sub-question compares the first attack step with the last using the variables 'First / Last'. This was tested using Fisher's Exact test[4]. Ultimately these analyses en-

[3]The following two data assumptions must be met for Chi Square analysis: *i*) independence and *ii*) minimum frequency of 5 observations per cell (Field, Miles, & Field, 2012). Independence relates to putting a single observation in only one cell. In case one assumption is not met, the Fisher's Exact test should be used instead.

[4]Fisher's Exact test was used because the minimum number of observations was not met.

able to find out whether the offender's principle selection is drawn at random or not, which would influence the design of awareness campaigns by security managers. For an overview of all variables used in the different tests, refer to Table 2.3.

Table 2.3: Overview of variables and statistical test.

			ASSUMPTIONS		
QUESTION[1]	VARIABLE[2]	IDEAL TEST[3]	IND[4]	OBS[5]	APPLICABLE TEST[6]
Q1	Modality	Chi Square	Y	N	Fisher's Exact
Q2	Persuasion Principles	Chi Square	Y	N	Fisher's Exact
Q3.1	Persuasion Principles	Chi Square	Y	N	Fisher's Exact
Q3.2	Steps	Chi Square	Y	N	Fisher's Exact
Q3.3	Steps(Authority)	Chi Square	Y	N	Fisher's Exact
Q3.3	Steps(Commitment)	Chi Square	Y	N	Fisher's Exact
Q3.3	Steps(Liking)	Chi Square	Y	N	Fisher's Exact
Q3.3	Steps(Reciprocity)	Chi Square	Y	N	Fisher's Exact
Q4.1	Persuasion Principles	Chi Square	Y	Y	Chi Square
Q4.2	Former / Latter	Chi Square	Y	N	Fisher's Exact
Q4.3	First / Last	Chi Square	Y	N	Fisher's Exact

[1] This column refers to the specific question and sub-question;
[2] This column refers to the variables that were used in the statistical test;
[3] Ideal test refers to which test ideally to use if all the assumptions were met;
[4] IND refers to the independence assumption that in required for some tests;
[5] OBS refers to the minimal number of observations that is required for some tests;
[6] The applicable test given the assumption results;

In this study the focus is on the 'Execution' phase of the crime script, which follows the preparation and target selection phases. All the interactions (i.e. attack steps) found in the 74 scenarios were annotated with a time-component to keep track of the chronological order.

2.3 Results

2.3.1 Descriptive statistics

A total of 74 scenarios were analysed. This work resulted in the identification of 142 attack steps containing 180 occurrences of persuasion principles. Out of the 142 attack steps, 125 used persuasion principles whilst 17 involved other social influences. The number of attack steps containing persuasion principles is statistically different from those containing other social influences $[N = 142, \chi^2 = 49.5, df = 1, p = .001]$. The category of 'other social influences' covers 12% of all social influences used by offenders

in their attack steps. This category contains: *i*) Act ignorant 1x (5.9%), *ii*) Creating curiosity 2x (11.8%), *iii*) Distracting 1x (5.9%), *iv*) Empathy/Pity 2x (11.8%), *v*) Just ask for it 9x (52.9%) and *vi*) Overloading 2x (11.8%).

All attack steps (i.e. interactions) are summarized in Table 2.4, which shows where the scenario originated from. Table 2.4 also shows that: *i*) Authority as a single principle is the most commonly used attack step; it is used in 76 (53.5%) of all attack steps, and it is identified in all 6 steps over time; *ii*) double principles (i.e. using two principles in one attack step) include Authority in all 27 (100%) attack steps; *iii*) Liking was used in 10 (91%) of the cases which use 3 persuasion principles in an attack step; *iv*) both Commitment, Conformity and Reciprocity are only executed once as single principles; *v*) the most frequently used attack path is AUTHORITY - AUTHORITY - COMPLIANCE; All crime scripts and attack steps listed on Table 2.4 are summarized in a single tree structure (refer to Figure 2.3). The attack steps containing similar strategies (i.e. persuasion principles) at the same point in time are merged into a single node in the tree. This tree gives an overview of all the social engineering scenarios from the books and shows the frequent and less frequent attack paths. There are other studies using multiple crime scripts for generalization purposes, e.g. on the topic of wildlife trafficking (Lavorgna, 2014). However, this is the first study that combined multiple crime scripts into a single visualisation.

2.3.2 Inter-rater agreement

The researchers' inter-rater reliability was: $N = 852$, $\kappa = .909$, 95% CI $[.874, .944]$, $p = .000$. The N represents 142 attack steps times 6 possible persuasion principles per step. The results indicate there is an almost perfect agreement between the two researchers (Landis & Koch, 1977).

2.3.3 *Q1:* "What modality is used to execute social engineering attacks?"

Out of the 142 attack steps, 123 (86.6%) were executed via telephone, whilst 18 (12.7%) were executed F2F and once via chat. There was a statistically significant difference in the modality used in executing persuasion principles in social engineering attack steps, $[N = 142, \chi^2 = 43.983, df = 1, p = .000]$. Regarding modality of attack steps, the telephone modality was used considerably more often whilst attack steps containing other means considerably less often.

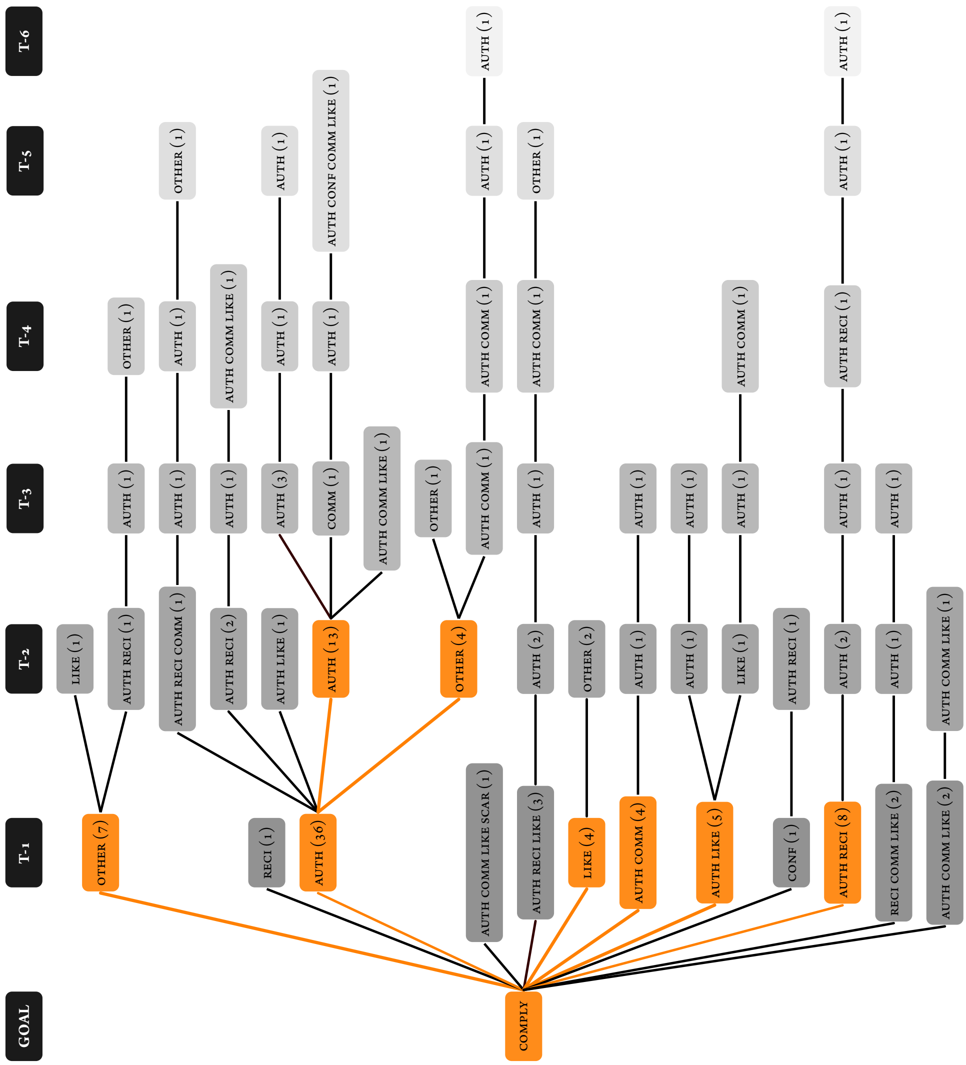


Figure 2.3: Tree structure of social engineering scenarios. The time flows from right (initial contact) to left (achieve goal). The frequency of occurrence is specified inside the brackets. The orange coloured nodes resemble the heart of social engineering attacks.

Table 2.4: Combination of principles found in books. The left-most principle constitutes the initial step in the attack.

PRINCIPLES	FREQ	WHERE TO FIND[1]
OTHER – COMPLY	5x	1:[6, 16, 36, 30, 47]
OTHER – OTHER – AUTH – COMPLY	1x	1:[2]
OTHER – AUTH – COMPLY	2x	1:[37, 11]
OTHER – AUTH – AUTH – AUTH RECI COMM – AUTH – COMPLY	1x	3:[5]
OTHER – AUTH COMM – AUTH – AUTH – AUTH RECI LIKE – COMPLY	1x	2:[4]
OTHER – AUTH – AUTH RECI – OTHER – COMPLY	1x	1:[27]
OTHER – LIKE – COMPLY	2x	2:[16], 4:[2]
AUTH – COMPLY	15x	1:[5, 10, 12, 17, 21, 22, 25, 29, 35] 1:[41, 45, 46], 2:[12, 14], 3:[3]
AUTH – AUTH – COMPLY	8x	1:[1, 26, 38, 44, 49], 2:[1, 11], 3:[1]
AUTH – AUTH – AUTH – COMPLY	2x	1:[28,40]
AUTH – AUTH – AUTH – AUTH – AUTH – COMPLY	1x	2:[18]
AUTH – AUTH – AUTH COMM – COMPLY	1x	2:[13]
AUTH – AUTH – AUTH COMM – AUTH COMM – OTHER – AUTH – COMPLY	1x	1:[34]
AUTH – AUTH – AUTH LIKE – COMPLY	1x	2:[5]
AUTH – AUTH – AUTH RECI – AUTH – AUTH – AUTH RECI – COMPLY	1x	2:[8]
AUTH – AUTH – RECI COMM LIKE – COMPLY	1x	4:[1]
AUTH – AUTH RECI – COMPLY	1x	2:[10]
AUTH – AUTH RECI LIKE – COMPLY	1x	1:[14]
AUTH COMM – COMPLY	3x	1:[8, 23] 2:[6]
AUTH COMM – AUTH – LIKE – AUTH LIKE – COMPLY	1x	1:[3]
AUTH COMM LIKE – COMPLY	1x	3:[2]
AUTH COMM LIKE – AUTH – AUTH – COMPLY	1x	2:[7]

Continued on next page

Table 2.4 – continued from previous page

PRINCIPLES	FREQ	WHERE TO FIND[1]
AUTH COMM LIKE – AUTH – AUTH RECI – AUTH – COMPLY	1x	2:[9]
AUTH COMM LIKE – AUTH COMM LIKE – COMPLY	1x	1:[9]
AUTH COMM LIKE SCAR – COMPLY	1x	2:[17]
AUTH CONF COMM LIKE – AUTH – COMM – AUTH – AUTH – COMPLY	1x	4:[3]
AUTH LIKE – COMPLY	3x	1:[48], 2:[15], 3:[6]
AUTH LIKE – AUTH – COMPLY	1x	2:[3]
AUTH RECI – COMPLY	6x	1:[7, 15, 24, 39], 2:[2], 3:[4]
AUTH RECI – AUTH – COMPLY	1x	1:[13]
AUTH RECI – CONF – COMPLY	1x	1:[18]
AUTH RECI LIKE – COMPLY	1x	1:[32]
LIKE – COMPLY	2x	1:[4, 20]
LIKE – OTHER – COMPLY	1x	1:[43]
RECI – COMPLY	1x	1:[31]
RECI COMM LIKE – COMPLY	1x	4:[4]

[1] The number before the colon refers to the book, the number(s) [between brackets] refer to the scenario number in that book.
Book 1 = *The Art of Deception: Controlling the Human Element of Security* (Mitnick & Simon, 2002);
Book 2 = *Ghost in the Wires: My Adventures as the World's Most Wanted Hacker* (Mitnick et al., 2011);
Book 3 = *Hacking the Human: Social Engineering Techniques and Security Countermeasures* (Mann, 2008);
Book 4 = *Social Engineering: The Art of Human Hacking* (Hadnagy & Wilson, 2010);
AUTH: In more than half of the attack steps, Authority was used as single step;
AUTH RECI: All 27 double principle scenarios contained Authority;
AUTH RECI LIKE: Liking is used in 91% of the triple principle scenarios;
RECI: Commitment, Conformity and Reciprocity are only executed once as single principle;
OTHER: 17 occurrences being non persuasion principles;

2.3.4 Q2: "Which persuasion principles are used in the context of social engineering attacks?"

Fisher's Exact test showed that there was a significant difference in the use of principles during social engineering attacks, [$p = .000$]. The occurrence of the 6 persuasion principles is *i*) 63% for Authority, *ii*) between 10 and 13% for Liking, Reciprocity and Commitment and *iii*) under 2% for Scarcity and Conformity (refer to Figure 2.4).

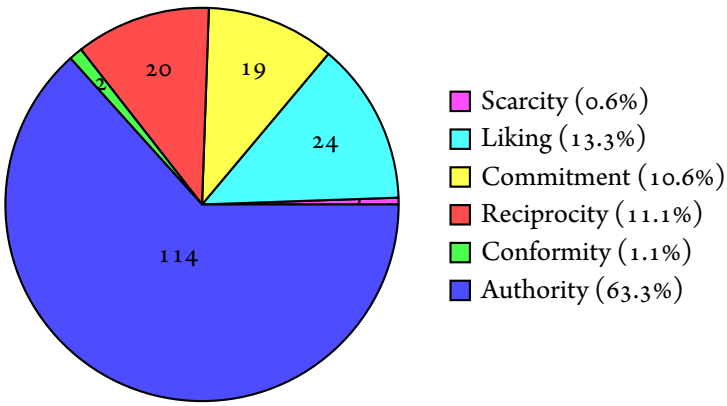


Figure 2.4: Prevalence of persuasion principles used in the four books that were analysed.

2.3.5 Q3: "How are the persuasion principles combined in social engineering attack steps?"

Q3.1: How many persuasion principles are used in an attack step?

In total 142 attack steps contain social influences, of which 125 include persuasion principles and the remaining 17 contain some other social influence. Single attack steps contain up to four different persuasion principles. The average number of persuasion principles used per attack step is $M = 1.44$ (SD = .723). There was a statistically significant difference in the number of occurrences of persuasion principles used in a single social engineering attack step, [$p = .000$].

Regarding simultaneously used principles, single principles are used considerably more often whilst quadruple principles considerably less often, refer to Figure 2.5.

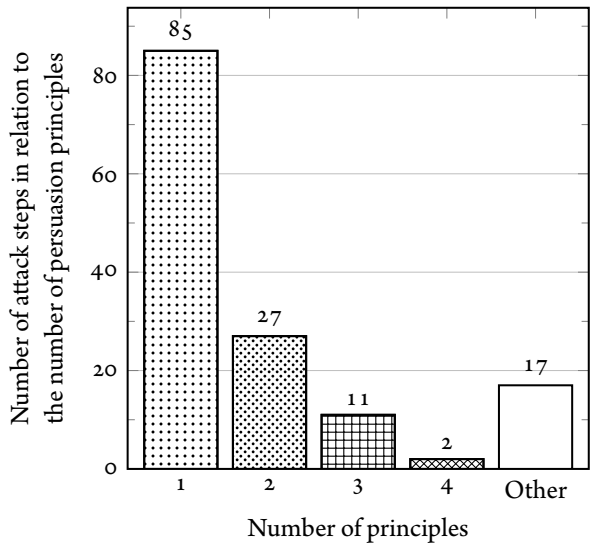


Figure 2.5: Number of principles used per interaction.

Q3.2: How many attack steps are used in an crime script?

In total 74 scenarios contain 142 attack steps. The shortest attack path contained a single step, whereas the longest attack path consisted of six attack steps. On average the attack path has $M = 1.92$ (SD = 1.311) steps. There was a statistically significant difference in the number of attack steps used in crime scripts, [$p = .000$]. Regarding combining attack steps, single step attacks are used considerably more often whilst attacks containing six steps considerably less often, refer to Figure 2.6.

Q3.3: How are persuasion principles used in the course of a crime script?

The combined principles were decomposed to obtain the occurrence of each principle (for absolute and relative occurrences refer to Figure 2.7 and 2.8 respectively). In the first step before compliance, there were 101 occurrences of persuasion principles, in the second step this decreased to 38, whilst in the sixth step before compliance there were only 2 occurrences. There was a statistical significant difference in the use of Authority [$p = .000$], Liking [$p = .005$], Commitment [$p = .003$] and Reciprocity [$p = .002$] over time. Authority, Liking, Commitment and Reciprocity are used considerably more in the 1^{st} step before compliance whilst these are used considerably less in the 6^{th}, 5^{th} and 4^{th} step before compliance. Due to the limited occurrence of the Scarcity and Conformity principles, it was not possible to perform a statistical test.

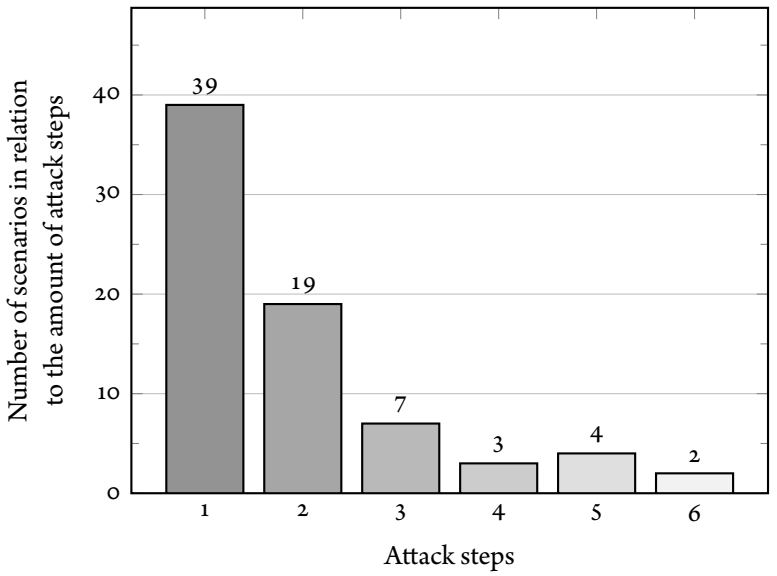


Figure 2.6: Number steps in an attack.

2.3.6 *Q4*: "What is the similarity between the different attack steps in the crime script?"

Q4.1: Do principles differ over the course of an attack?

In a multiple step attack, the average number of steps is $M = 1.92$ (SD = 1.311), therefore the last three attack steps before compliance comprise 88% of all attack steps. In the remaining 17 steps there are eight single principle steps, which all contain Authority; there are four double principle steps, which all combine Authority with some other principle (i.e. Commitment and Reciprocity). Furthermore, there is one triple principle step consisting of Authority, Commitment and Liking, and one quadruple one containing Authority, Conformity, Commitment and Liking. Finally, there are three steps containing another form of social influence than the six persuasion principles.

Single principle attack steps are more popular than combined principle attacks (refer to Figure 2.5). Authority as a single step is used in 85.7% of the steps immediately before compliance; 90.9% of the second steps before compliance; 91.7% of the third steps before compliance. The use of Authority (i.e. in isolation) is relatively stable over time. The absolute use of principles over time is shown in Figure 2.9, whilst the relative use in Figure 2.10.

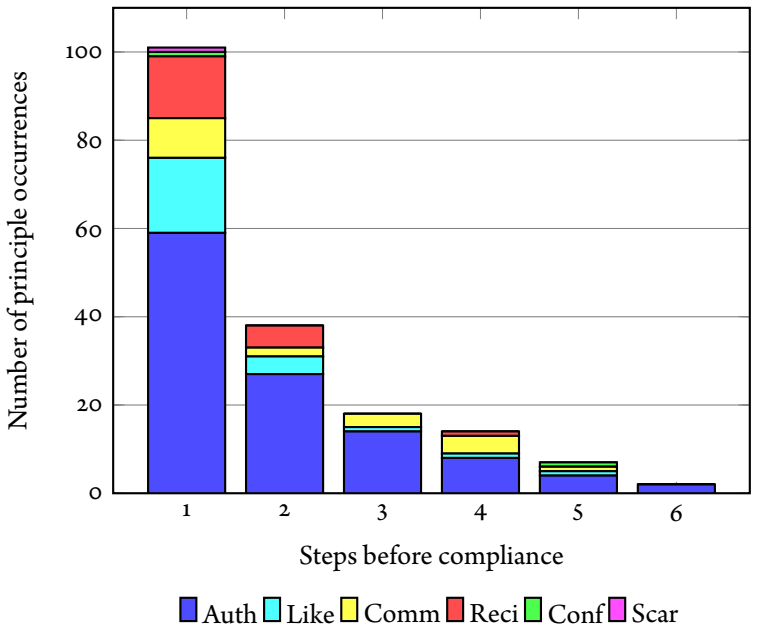


Figure 2.7: Absolute principle use over time.

Despite the small numbers, in double principle attack steps the combination Authority + Commitment (3 times), Liking (4 times) or Reciprocity (7 times) are recurring combinations (refer to Table 2.4). In the final step before compliance, 67 attack steps used persuasion principles versus 7 that used some other social influence. The number of scenarios containing persuasion principles in the last step before compliance is different from zero [$N = 74$, $\chi^2 = 29.108$, $df = 1$, $p = .001$].

Table 2.5 shows an increase of attacks using multiple principles towards the end of the attack. Towards the moment of compliance, the relative use of single persuasion principles in an attack step drops from 81.3% to 56.8%, while the use of multiple persuasion principles in attack steps increased from 12.6% to 33.9%. This tendency is summarized in Figure 2.11.

Q4.2: Comparison between two consecutive steps.

In total there are 68 consecutive attack steps identified in the 74 scenarios. The number of principles in the former attack step does not differ from that of its latter step, [$p = .768$].

There are 45 attack steps that contain persuasion principles; 39 (57.4%) attack steps began with Authority, 26 (38.2%) steps succeeded with a single Authority attack step whilst 10 (14.7%) of them consist of a combination of Authority + Commitment,

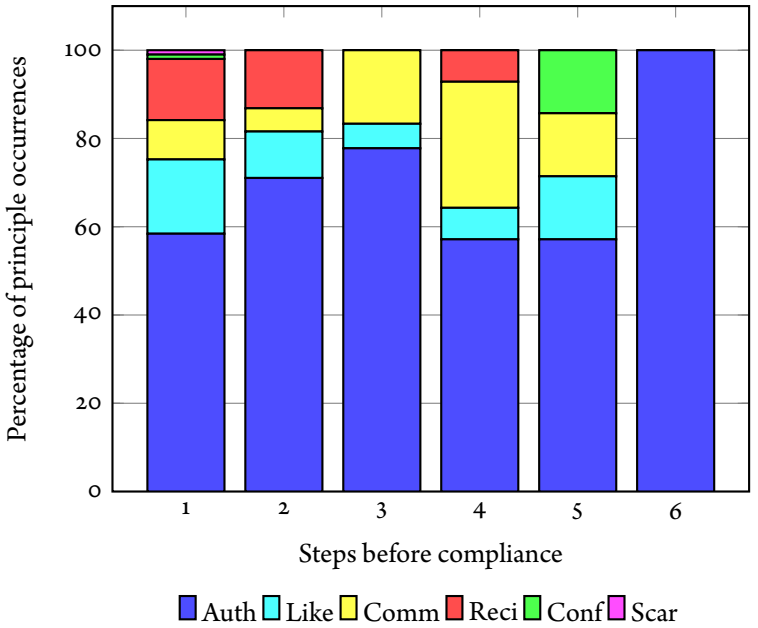


Figure 2.8: Relative principle use over time.

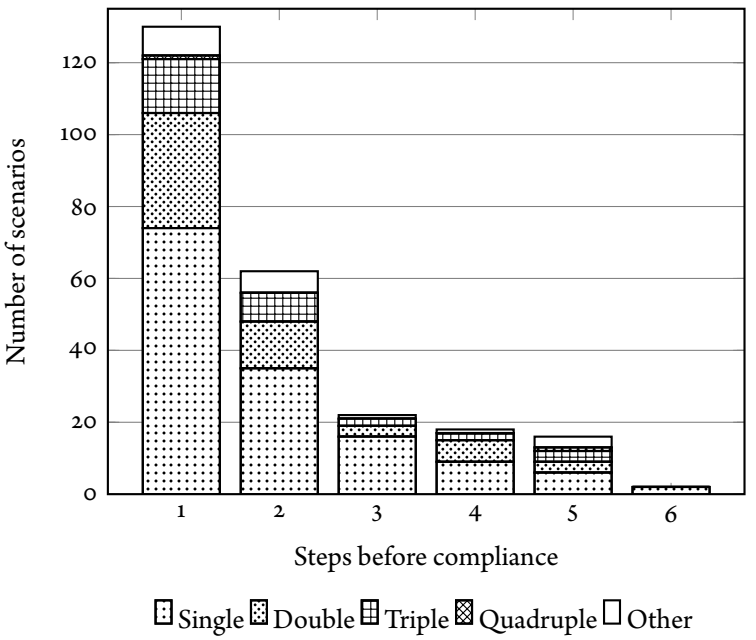


Figure 2.9: Absolute principle combination for each step before compliance.

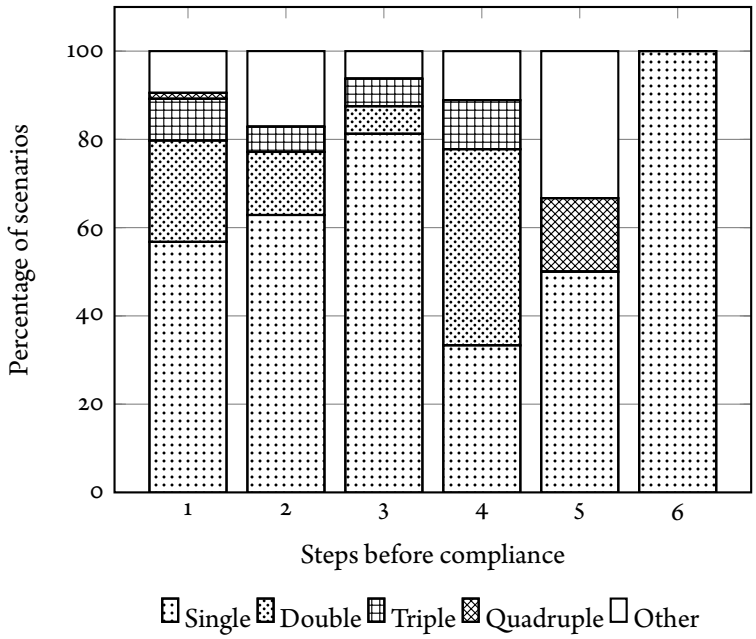


Figure 2.10: Relative principle combination for each step before compliance.

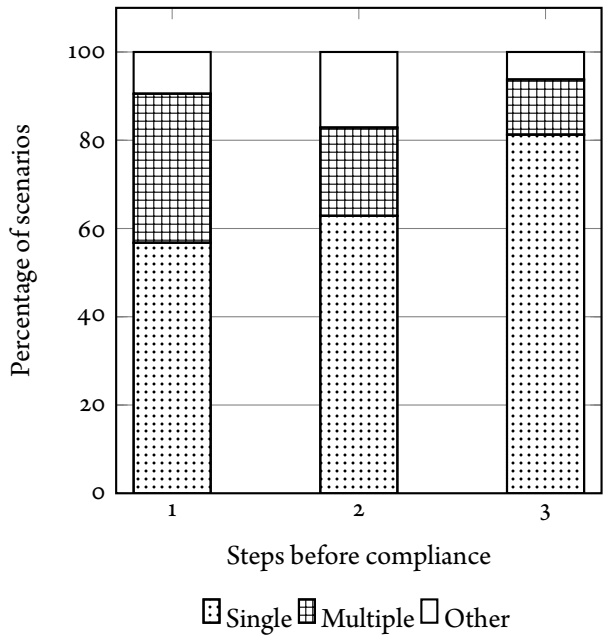


Figure 2.11: Use of principle combinations last 3 steps before compliance.

Table 2.5: The number of combined principles in an attack step over time before the target complies with the request (N=142).

		STEPS BEFORE COMPLIANCE						
		1	2	3	4	5	6	TOTAL
PRINCIPLES	SINGLE	42 (56.8%)	22 (62.9%)	13 (81.3%)	3 (33.3%)	3 (60%)	2 (100%)	85
	DOUBLE	17 (23.0%)	5 (14.3%)	1 (6.3%)	4 (44.4%)	-	-	27
	TRIPLE	7 (9.5%)	2 (5.7%)	1 (6.3%)	1 (11.1%)	-	-	11
	QUADRUPLE	1 (1.4%)	-	-	-	1 (20%)	-	2
	OTHER	7 (9.5%)	6 (17.1%)	1 (6.3%)	1 (11.1%)	2 (40%)	-	17
	TOTAL	74 (100%)	35 (100%)	16 (100%)	9 (100%)	6 (100%)	2 (100%)	**142**

Reciprocity or Liking. Only 3 consecutive steps end with something other than Authority. Finally, there are 13 steps that either start or end with a social influence other than a persuasion principle.

The combined principles in an attack step were further decomposed. A total of 115 consecutive individual principles were found in the 74 scenarios. There are 17 steps which either begin or end (i.e. succeed) with a social influence other than the six persuasion principles. The principles used in the former step do not differ with respect to those in the latter step [$p = .630$]. Out of the remaining 99 consecutive persuasion principles, 74 (74.7 %) consecutive principles begin with Authority, whilst 49 (49.5 %) of the consecutive principles also end with this principle. Furthermore, there are only 7 (7.07 %) consecutive persuasion principles which don't involve Authority.

Q4.3: Relation first and last step before compliance.

There are 26 scenarios with more than one attack step. The number of persuasion principles in the first step does not differ from those in the final step, [$p = .515$].

The combined principles in an attack step were further decomposed. In total there are 55 scenarios with individual persuasion principles. The number of persuasion principles used in the first attack step does not differ from those in the final step, [$p = .797$]. There are 37 (66%) scenarios starting with Authority, whilst it is used 24 times (42.9%) as the last step in the scenario. Only 6 (10.7%) scenarios do not involve this principle as either the first or last principle.

2.4 Conclusion

In social engineering attacks, offenders use persuasion principles to change the odds of their target complying with their request. This study investigated how persuasion principles were used in successful social engineering attacks based on the accounts of social engineers.

The dissection into crime scripts shows that the anatomy of social engineering attacks consists of: *i*) Persuasion principles (refer to Q2), *ii*) Other social influences (refer to Q2), *iii*) Deception, *vi*) Real time communication and *v*) Telephone operation (refer to Q1). The heart of the social engineering attacks (based on numerical frequency) is shown in orange in Figure 2.3. This visualization contains the key elements of a social engineering attack since: *i*) Approximately 80% of the crime scripts consist of 1 or 2 attack steps (refer to Figure 2.6); *ii*) Approximately 80% of the attack

steps consist of 1 or 2 persuasion principles (refer to Figure 2.5 and Table 2.5); *iii*) The most frequently used persuasion principles are Authority and Liking (refer to Figure 2.4); *vi*) Besides persuasion principles, other social influences are used (refer to Q2); *v*) Combined persuasion principles in attack steps always contain Authority.

This study gives an unique insight into how offenders use social engineering successfully to perform their criminal act. In 88% of the attack steps persuasion principles were used by the offender. Hence, in 12% of the attack steps another social influence was used. We can therefore conclude that (based on their accounts) social engineers make frequent use of persuasion principles as social influences to make their targets comply with their request. Some principles are used more frequently than others. Given the similarity between the ranking of principles based on the success-rates in the meta-analyses (refer to Table 2.1) and the ranking of principles in this study (refer to Figure 2.4), we believe that the occurrence of principles reflects their effectiveness in social engineering.

In order to draw conclusions about effectiveness, an experiment would need to be performed to verify the present conjectures. The experimental conditions can be controlled in an experiment (i.e. the use of the persuasion principles). Furthermore, the effectiveness can be determined since the number of subjects who comply and don't comply is known. When the experimental design and context are kept constant, the effectiveness of the persuasion principles can be calculated. Such an experiment could, for example, involve a so-called technical support scam. This involves a telephone fraud scam where the offender impersonates technical support service personnel (Arthur, 2010). The *Modus Operandi* (MO) often includes the offender informing the target that there is a problem with their PC. To resolve the problem the caller recommends buying a small software tool to prevent further damage. The use of individual or multiple persuasion principles can be included in the telephone script used by the offender. One instance of this experimental design is demonstrated in Chapter 3.

It was found that there are significantly more social engineering attacks executed via the telephone, compared to other modalities. We believe that this is because of the lower effort and risk that such an attack entails compared to one that involves being physically present. Physically going to meet the target implies additional risks: *i*) the risk of being caught at the scene, *ii*) body language could hamper the plan and, *iii*) the attacker can only attack once as his face might have been recognized.

The outcome of an attack can be influenced if offenders are able to apply persuasion principles. The literature suggests that all principles can be effective. However, the success of principles depends on the context, operationalization and final goal. Mil-

gram showed that there was a significantly lower effect of Authority when switching to the telephone modality (Milgram, 1965). However, a 'nurse experiment' showed a high rate of compliance when Authority was applied over the telephone (Hofling, Brotzman, Dalrymple, Graves, & Pierce, 1966). The present study shows that Authority was the most frequently successful used principle. The reason it is most commonly executed over the telephone probably relates to its relative low effort. Regarding Authority, there are several factors that explain its effectiveness.

One of the pre-conditions for Authority is the institutional framework of modern society. Since childhood people are taught how to operate within an institutional system. This framework is initially in the form of regulated parental-adult Authority, later through the Authority of teachers in school and finally through a boss in a company or commander in the army (Milgram, 1963). Another pre-condition for Authority is rewards for compliance to Authority, while failure to comply results in a punishment. Authority (from a psychological point of view) is when a person is perceived to be in the position of social control for a particular situation (Milgram, 1963). The person claiming to be the Authority will succeed if: *i*) someone expects an Authority, *ii*) appropriate dress or equipment is used (e.g. lab-coat, tag, uniform), *iii*) there is absence of competing Authority and *iv*) there is absence of conspicuous factors (e.g. child of 5 years old claiming to be a pilot). Unless contradictions in the information or anomalies appear, the Authority will likely suffice. People respond to the appearance of Authority, rather than the actual Authority (Milgram, 1963). Another finding of this study is that only very few social engineering attacks begin or end with the Scarcity principle; we assume that this reflects low success rates. Scarcity seems easy to operationalize, since this is a frequently used technique in sales and television advertisements (e.g. 'only 25 products left' or 'order **today** and get a 50% discount on the second item'). Furthermore, it seems that single step attacks only containing Commitment, Scarcity or Conformity are rare. The data shows that instead, Commitment and Reciprocity are used in combination with Authority. This could indicate that this combination of principles strengthens each other.

Knowledge about the principles, principle combinations and the time line in social engineering attacks is useful for designing countermeasures. The topic of countermeasures is discussed later in this section.

The use of a single principle occurs in almost 60% of the attack steps. The use of combined persuasion principles is used in less than 30% of the attack steps. This suggests that it is easier to operationalize an attack step consisting of a single principle compared to multiple principles. Although it is likely that by combining multiple principles in one step the effectiveness of that step increases, the operational complexity

could outweigh its benefits. To the best of our knowledge, the issue of principle combination had not been discussed in the literature until now.

The results of this study show that the average number of attack steps (interactions) is two. This means that the crime script is short and that in order to make a target comply with the offender, only information or support from one other employee is needed. In the final step before compliance, the number of combined principles increases. This could indicate that to make the target comply, a boost is needed in the final attack step and that this is being achieved by combining principles in one attack step.

The results of this study show no difference between the number of principles used in the first attack step compared to those in the last attack step. Furthermore, there is no statistical difference in the number of principles used comparing the former and latter of two successive attack steps within a crime script. The results indicate that a single principle attack step is more likely to occur after a single principle attack step. Similarly, it is more likely that the use of Authority is followed by Authority. From this we can conclude that offenders, like all other humans, might be creatures of habit in the sense that they stick to the method they initially chose. We assume offenders choose their method based on successful past attempts.

Moreover, regarding countermeasures to defeat social engineers, results showed that successful social engineering attacks most often use Authority. However, since our society is built on the authority paradigm, it would be extremely difficult to counteract Authority by itself. We believe that it is a better approach to use situational countermeasures. These could involve 4 mechanisms; *i*) Procedural, *ii*) Environmental, *iii*) Technical and *iv*) Behavioural.

i) Procedural countermeasures could, for example, involve the use a classification system for all organizational data, including employee and PC names, schedules and software versions. Data above a certain classification threshold should not be allowed to leave the organisation. Furthermore, in the case of someone receiving a request, it should be determined if the request is legitimate (Mitnick & Simon, 2002). This can be done by verifying the identity of the requester (e.g. call-back policy or shared secret). An employee who receives a request for information should verify the identity of the requester and initiate the communication via a channel that is verified by the organisation. If someone claims to be a colleague, the employment status should be confirmed (e.g. lookup in employee directory or contact their manager for verification). After verifying the employment status, check the knowledge needs of that person (e.g. check level of classification employee or contact his manager). It is important that the verification does not become a time-consuming process because em-

ployees might then ignore this.

ii) Environmental countermeasures adjust the environment to encourage a desired behaviour. Research results show that social engineers are likely to operate via the telephone. One environmental countermeasure could involve placing stickers on telephones; to remind the user each time they use the telephone.

iii) A technical solution could include using white and black lists. Telephone calls from numbers on the white list are connected whilst, those on the black list get terminated.

iv) Behavioural countermeasures relate to adjusting the human; *a*) by making employees aware that there are social engineers that use social influences to make people comply with their requests and they should realize that they are vulnerable (Muscanell et al., 2014), *b*) by training people to spot a social engineering attack, *c*) by making employees aware of why this is dangerous and what the implications are for the individual and the organisation and *d*) by distributing guidelines about what to do or not to if they are under a social engineering attack.

Muscanell et al. (2014) describe the best practices to resist social influences (i.e. persuasion principles). Best practices result in six questions to counteract the individual persuasion principles, those are described in Chapter 1.1.2.

The short-term effects of informing employees has been demonstrated by a group of students using social engineering to obtain office keys from university personnel. Those who received an information campaign complied significantly less frequently than those who were not informed, refer to Chapter 3. A related relevant issue would be to assess how learning impacts compliance over time. We expect that 'quick' interventions which are repeated on a regular basis are effective. Such an effect is already shown in the context of Cardiopulmonary Resuscitation (CPR) skill retention. In this study, the subjects were given a 4 minute CPR training every 1, 2 and 3 months after the start of the study. The final result (after 6 months) was an increase from $\pm$ 20% to $\pm$ 70% of the subjects performing a perfect CPR (Sutton et al., 2011).

Finally, both the implementation and the effectiveness of the procedure should be tested. This can be achieved by trying to social engineer one's own employees in a controlled environment. If this is done regularly, the employees will most likely remain responsive and alert.

When conducting experiments on humans (e.g. employees) some ethical concerns must be taken into account. The Belmont Report (1979) defines 3 ethical principles for the protection of humans during testing: *i*) respect for persons, *ii*) beneficence and *iii*) justice.

"*Respect for persons incorporates at least two ethical convictions: first, that individuals*

should be treated as autonomous agents, and second, that persons with diminished autonomy are entitled to protection" (Belmont Report, 1979). Respect for persons means that people are free to participate or to decline participation in research. Furthermore, people who are not capable of making competent decisions by themselves should be guided by a capable guardian. Beneficence is defined as: "*Persons are treated in an ethical manner not only by respecting their decisions and protecting them from harm, but also by making efforts to secure their well-being*" (Belmont Report, 1979). This means that one should not harm the participants. Moreover, the possible benefits of participating should be maximized, and the potential harm should be minimized. "*Who ought to receive the benefits of research and bear its burdens?*" (Belmont Report, 1979) relates to the justice principle. Both the risks and benefits of the study should be equally distributed within the subjects.

One ethical challenge is the use of deception since it conflicts with the 'respect for persons' principle. The use of deception might be acceptable if: *i*) the experiment does not involve more than minimum risk (i.e. harm or discomfort should not be greater than those experienced in daily life) (Code of Federal Regulations, 2005), *ii*) the study could not be performed without deception (subjects in laboratory studies may behave differently than they normally would or their behaviour may be altered because of the experimental setting), *iii*) the knowledge obtained from the study has important value and *iv*) when appropriate, the subjects are provided with relevant information about the assessment afterwards (i.e. debriefing) (Finn & Jakobsson, 2007).

Dimkov, Pieters, and Hartel (2009) described the 5 R^* requirements in penetration testing research (which often uses social engineering): *i*) Realistic (the test should resemble a real life scenario), *ii*) Respectful (the test should be done ethically) *iii*) Reliable (the test should not cause productivity loss of employees) *iv*) Repeatable (repeating the test should result in similar results) *v*) Reportable (all actions should be logged).

Dimkov et al. identified conflicting requirements and noted that designing a penetration test involves finding a balance in the requirements. For a discussion of the three major ethics standpoints (i.e. *i*) Virtue Ethics, *ii*) Utilitarianism and *iii*) Deontology) refer to (Mouton, Malan, Kimppa, & Venter, 2015). Finally, it should be noted that in the current study the authors did not have any control regarding the ethical concerns in the scenarios as they are literature-based.

2.4.1 Future work

Finally we summarize four recommendations for future research. The details of all recommendations are already presented in the discussion section, therefore only a brief summary is presented in this section, for the purpose of referring the reader to additional works. First, the analysis of the four books shows that social engineering works. However, all scenarios in the books involved success stories, therefore it is unclear what the success rate of a social engineering attack is. Experiments should therefore be used to find the success rates. Experiments to find such success rates are discussed in Chapters 3, 4 and 5. Second, a useful follow up study could involve investigating if these social engineering attacks can be blocked or their effects reduced. The effectiveness of countermeasures will be discussed in Chapters 3 and 5. Furthermore, one should investigate the effect of persuasion principles when used in combination in an attack step to identify which ones have the likelihood of succeeding. Authority, one of the persuasion principles is investigated in more detail in Chapter 3. Finally, it is possible that compliance depends on cultural aspects. The deployment of social engineering experiments in different countries could allow to identify cross country differences. The effect of culture on the outcome of a social engineering attack is discussed in Chapter 4.

It ain't what you don't know that gets you into trouble.
It's what you know for sure that just ain't so.

Mark Twain

3

Face-to-face social engineering

This chapter is based on (Bullée, Montoya, Pieters, Junger, & Hartel, 2015a) and (Bullée, Montoya, Junger, & Hartel, 2017a).

3.1 Introduction

Chapter 2 focused on the social influence techniques successfully used by offenders in social engineering attacks. A total of 74 success stories from the social engineering literature were analysed, showing how offenders exploit the human element in security. This chapter describes a Face-to-Face (F2F) social engineering experiment with an identical *Modus Operandi* (MO) as in the previous chapter. In particular, the effect of an intervention and the usage of authority will be tested.

Many attempts at unauthorized access involve exploitation of human weaknesses. Whereas this so-called social engineering had previously been associated with the social sciences, its terminology has caught on among computer and information security professionals (Anderson, 2008). Social engineering is a non-technical type of attack based on human interaction and often involves tricking others into breaking security policies. Examples of social engineering include persuading targets to run malware-infected email attachments as well as phishers trying to convince people to disclose sensitive information. In the field of email social engineering (i.e. phishing), emails often lead to fake websites that resemble legitimate websites, and lure people into disclosing sensitive information (Kumaraguru, Sheng, Acquisti, Cranor, & Hong, 2010). Security experts propose that as our culture becomes more dependent on information technology, social engineering will become the greatest threat to any security system (Rouse, 2006).

These examples show that social engineers know that humans are a weak link in cyber-security, and therefore try to trick people into violating security policies. The actions of social engineers are designed to appear harmless and to look legitimate (The Federal Bureau of Investigation, 2013). For some time, social engineering has been part of many organisational security testing programs. Such programs typically involve an outside party conducting *i*) technical tests such as vulnerability assessments, penetration tests, audits, and *ii*) user-oriented tests such as social engineering and phishing tests. These actions can be simple ones such as calls using invented scenarios (i.e. pretexting, tailgating) or complex such as sophisticated make-up and appropriate acting. Many organisations choose to conduct both on-site and remote social engineering testing to elevate the staff's ability to recognise social engineering attacks and to find out how they respond (Cross, 2011). However, little scientific literature has shown how persuasion techniques influence the success of a social engineering attack or the effect of countermeasures. This chapter therefore explores to what extent *i*) persuasion techniques influence the outcome and *ii*) an intervention reduces the effects of social engineering.

3.1.1 Persuasion

Humans are susceptible to persuasion by nature and in some circumstances resisting is almost impossible. Once a person is a target, the offender can use persuasion techniques to change the odds in his/her favour. Six principles of persuasion can be used to increase the offender's probability of success: Reciprocity, Conformity, Liking, Scarcity, Commitment and Authority (Cialdini, 2009). For details regarding the persuasion principles, refer to Chapter 1.1.2.

In particular, the *Authority* principle describes people's tendency to obey the request of authoritative figures. If people are unable to make a well-informed decision, the responsibility to do so is transferred to the group or person they believe is in charge. Crisis and stress activate the behavioural trait of responsibility transition. The Authority principle can be operationalized in multiple ways; the three most common are: *i*) prestigious titles such as Professor, Doctor or Lawyer give strength to an argument of authority; *ii*) stylish and expensive outfits carry an aura of status and position, as do trappings such as jewellery and cars; *iii*) the physical appearance of a person is an indicator of authority (Doob & Gross, 1968). There are differences in perception of authority towards someone wearing a police uniform, a casual outfit, a business suit or work clothing (Bickman, 1974; Lefkowitz, Blake, & Mouton, 1955).

The most famous study illustrating authority is the classical experiment of Stanley Milgram, which tested obedience to authority. Details on the effectiveness of authority in this experiment are discussed in Chapter 2.1.1.

The operationalization of authority using clothing was demonstrated by means of an experiment (Bickman, 1974; Lefkowitz et al., 1955). A stranger requested a small donation for the payment of a parking meter. When a police officer asked for a contribution, 92% were willing to contribute, compared to 42% in the case of a civilian (Bickman, 1974). A difference also was found between the number of people willing to follow someone who was crossing illegally at an intersection, depending on whether the person was casually or formally dressed. If the pedestrian was wearing a well-tailored suit, 3.5 times more people followed and crossed illegally than when the pedestrian wore casual clothing (Lefkowitz et al., 1955). The authority principle, operationalized by formal clothing will be used as experimental condition in this Chapter.

3.1.2 Behavioural changes to countermeasure persuasion

As stated before, offenders are aware of the susceptibility of humans to persuasion. Interventions are a common way to promote behavioural change and make poten-

tial targets more resistant against psychological manipulation. A wide range of approaches, strategies and theories can be used. Many of them are well studied and are the result of extensive research. Well known strategies and theories on attitude change include: *i*) shift of focus from the Protection Motivation Theory (Rogers, 1975), *ii*) the effect of social comparison shown by Festinger (1957), *iii*) correction of misconception from the Theory of Planned Behaviour (Ajzen, 1988, 1991), *iv*) model learning by observing role models from the Social Learning Theory (Bandura, 1986), *v*) persuasive communication from the Elaboration Likelihood Model (ELM) (Petty & Cacioppo, 1986). The ELM of Persuasion describes how information is processed and can be tailored to the receivers. According to the ELM, people process information via either the peripheral or the central route. The peripheral route of information processing is used when there is minimal attention to the message and can involve superficial cues, such as the attractiveness of the message presented. One may like the sound of a person's voice, or that person might have gone to the same university as one did. The central route, on the other hand, involves persuasion on the basis of the message content, such as voting for the political party with the best arguments (Petty & Cacioppo, 1981). Consistent with the ELM theory, attitudes obtained via the central route last longer, are less vulnerable to contra-argumentation and are better predictors of human behaviour. Furthermore, the effect of persuasive communication increases if the message is relevant to the audience and if surprises and repetitions are used (Petty & Cacioppo, 1984, 1986).

Persuasive communication by means of leaflets has been thoroughly studied and proved to be an effective mechanism for administering an intervention. Studies about successful leaflets have focused on increasing the knowledge of the general public (Humphris, Duncalf, Holt, & Field, 1999; Stubbings et al., 2000), as well as more specific groups such as patients (Barlow, 1998; Hawkey & Hawkey, 1989), parents (Ghaderi, Adl, & Ranjbar, 2013) and customers (Shim et al., 2011). Other studies found that leaflets influenced behaviour (Ershoff, Mullen, & Quinn, 1989; Hart et al., 1997) and reduced anxiety about an illness (Humphris, Ireland, & Field, 2001; Robb, Miles, Campbell, Evans, & Wardle, 2006). However, conflicting results were reported by Carré et al. (2008). They argued that the readability of their leaflet was probably not optimal for the audience, although the majority of the subjects found the leaflet fairly clear and interesting (Carré et al., 2008).

A meta-analysis on smoking cessation compared the distribution of leaflets against a control condition (i.e. no leaflet). In total 12 trials were conducted and a significant effect was found on smoking cessation ($N = 14\,787$, $p = .008$) (Lancaster & Stead, 2005). Similarly, experts argue that both training and education can help protect

users against phishing (Hight, 2005; Kumaraguru et al., 2010). Intervention materials currently available can be effective as long as the user actually reads the material (Kumaraguru et al., 2010). In the field of advertising, a distinction can be made between comparative and non comparative advertisements; the latter are peripherally processed (Lien, 2001). A meta-analysis on advertisements showed that comparative ads led to an increase in brand attitude ($k = 42$ studies, effect size $d = .23$), purchase intention ($k = 47$, $d = .20$) and purchase behaviour ($k = 6$, $d = .46$) (Grewal & Kavanoor, 1997). Information leaflets may offer an important contribution to raising long-term knowledge and awareness, but to be effective, these must be well presented and understandable by the target population (Krawczyk et al., 2012; Petty & Cacioppo, 1986).

Subtle reminders have shown to be a useful mechanism for remembering the contents of a leaflet within the context of promoting desired behaviour (Gisquet-Verrier & Riccio, 2012; Glanz, Rimer, & National Cancer Institute, 1997). Reminders (cues leading to action) have proved to be necessary in the promotion of healthy behaviour (Rosenstock, 1974). Reminders can either be internal (e.g. perception of bodily states) or external (e.g. medical advice, postcard reminder, television advertisement, warning label). Reminders are mentioned in the theory of Situational Crime Prevention as elements that Remove Excuses (Cornish & Clarke, 2003). Although the literature on reminder cues is limited, research suggests that these are effective (Flight, Wilson, & McGillivray, 2012).

Humour has been found to have a positive effect on the recall of information compared to neutral cues (Carlson, 2011; Gulas & Weinberger, 2006, p. 75). Schmidt (1994) describes three mechanisms that explain the effect of humour on information recollection. The first considers the effect of humour on physiological arousal. Humour is associated with an increase in heart and respiratory rates (McGhee (1983), (as cited in Schmidt, (1994)). Increased arousal rates during the presentation of humorous material are found to lead to long-term memory retention, compared to neutral arousal rates (Craik & Blankstein, 1975). Second, it is argued that humour increases attention towards a subject. Paying more attention to something therefore results in better recollection (Schmidt, 1994). A third mechanism is the effect of repetition caused by humorous material. Recollection is better when material is presented on more than one occasion (Schmidt, 1994).

3.1.3 Experimental context

Literature shows that social engineering attacks work (Mann, 2008), that they are effective (Schellevis, 2011) and that targets are unaware of being victimized (The Fed-

eral Bureau of Investigation, 2013; Hadnagy & Wilson, 2010). However, there is limited literature on success rates for such attacks. The books of Kevin Mitnick (2002; 2011) give an insight into how successful social engineering attacks are executed. Although this is interesting anecdotal information, the prevalence and success rate of social engineering attacks are unknown. Penetration testing reports occasionally surface, but these represent the proverbial needle in a haystack and are uncontrolled. To the best of our knowledge, there are no other studies that combine persuasion principles with an intervention.

3.1.4 Collecting data

Surveys are a convenient method to collect data. However, they can give biased outcomes, due to the tendency of people to provide socially desirable responses. There is typically an over-reporting of 'good behaviour' or an under-reporting of 'bad behaviour' (Crowne & Marlowe, 1960). In political science, this phenomenon is referred to as the Bradley effect, Wilder effect or Dinkins effect and relates to the discrepancy between voter opinion polls and the outcome of elections (Payne, 2010). Social sciences also acknowledged the discrepancy between verbal compliance and behavioural compliance. Among others, this was investigated in the context of fear of crime. The researchers first asked how people would react if, unexpectedly, a stranger ringed their door. In a later stage, the researchers would go to their homes, ring their door and see if their reaction matched the verbal compliance. The researchers found a statistically significant difference between the verbal and behavioural compliance (p = .001) (Van Dijk & Nijenhuis, 1979). Although there is a difference between verbal and behavioural compliance, both a survey and an experiment were conducted to measure the success and differences of social engineering attacks. Only the success rate was measured, prevalence was not taken into account.

The success of a social engineering attack might be substantially influenced by the context. Context can influence and can be influenced by both the target and the offender to obtain a more desirable outcome. The offender may use the knowledge of the six principles of persuasion to achieve a higher probability of compliance. On the other hand, the target can be informed about social engineering via an intervention aiming at rejecting the offender's requests. In the current experiment, the context influenced by the offender is authority, whilst the context influenced by the defender is the intervention. The social engineering experiment was administered to university personnel with the goal of making them surrender their office key. The advantage of a key over a password is its appearance, since a key is physical and a password rep-

resents knowledge. It is therefore possible to surrender a key, but a password has to be shared. One cannot return a password without still knowing it, this additionally introduces the burden of changing one's password afterwards.

3.1.5 Research question

The objective of this research was to answer the following question: "*To what extent are people susceptible to a face-to-face scam*?" Five hypotheses were formulated:
H1) Previous research showed that there is a discrepancy between verbal and behavioural compliance. Therefore we hypothesise that there will be a difference between the self estimated and experimental observed compliance.
H2) Previous research on the effect of informing people showed an increase of their knowledge and a change in behaviour. We therefore hypothesise that in the intervention group, fewer people comply with the offender's request than in the control group.
H3) Previous research on the effect of authority showed an increase in compliance towards an authoritative figure compared to non-authoritative figures. We therefore hypothesise that in the experimental group where the researcher exercises authority (i.e. by wearing formal clothing), more people comply with the offender's request than in the control group.
H4) An intervention reduces the effect of authority in relation to the compliance with a request and therefore the third hypothesis is that authority has more effect if no intervention is exercised.
H5) Since there is (to the best of our knowledge) no literature available on the ideal placement of smart key lock activators in the context of a penetration test. We therefore hypothesise that the amount of physical effort influences the willingness of an employee to hand over his office key to a stranger.

3.2 Method

The sample consists of 162 subjects (who participated in a field study) and 49 subjects (who filled in a survey) of both sexes who work in two buildings on the University of Twente (UT) campus. Only people who were present in their office and whose office door had a specific type of lock (i.e. smart lock) were approached. The smart locks are electronic and manufactured by WinkHaus. A physical plastic token (resembling a key) with a chip inside is used to unlock the door (refer to Figure 3.1). The chip contains data that specifies which locks are allowed to be opened. A 128-bit challenge-

response between the key (chip) and the lock is used to validate the identity of the key. When the key is authorized, the lock pulls a small pin and allows one to open the door (Schneier, 2005). The validity of the credentials on the key lasts until midnight and thus the key has to be reactivated on a daily basis. The activation facilities are located next to the entrance to the buildings so that staff can activate the key upon arrival. Staff who forgets to activate must return to the entrance.

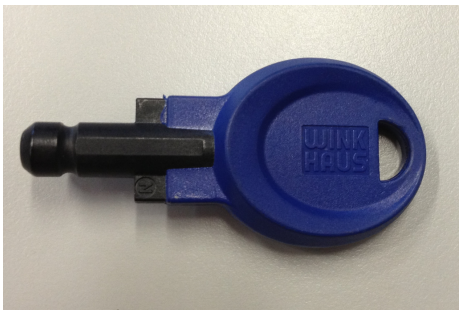


Figure 3.1: The smart office key that is in use in the buildings where the experiment was performed.

3.2.1 Subjects

Professors, secretaries and laboratory staff were excluded from the experiment in order to minimize disruption of the main university activities. The sample only consists of one third of all possible targets; the sex distribution is comparable to the overall sex distribution of the buildings, whilst the experimental sample is slightly younger (34 vs 40 years). Although there are some small differences with respect to a number of characteristics, the sample deviates little from the general population of these buildings. Those who filled out the survey also had a comparable sex distribution (χ^2 = .031, df = 1, p = .861), although they were younger; 30.37 years (SD = 6.23).

3.2.2 Researchers

The researchers (i.e. the "attackers") consisted of 35 Bachelor/Master students (12 female and 23 male). The average age of the researchers was 21.43 years (SD = 1.38). The researcher who administered the surveys was a 19-year-old Dutch male bachelor student. There was no restriction in terms of approaching subjects of the same sex.

3.2.3 Field study procedure

The Institutional Review Board (IRB) of the University approved the study before collecting data. One week before the researchers approached the subjects, half of the subjects were exposed to an intervention. Departments were randomly selected and all their staff was exposed to the intervention, in an effort to avoid disturbing the normal activities of the department.

No targets from the control group reported having been aware that an intervention had taken place, whereas those in the intervention group reported to have received the intervention. This was ascertained during the debriefing when subjects were directly asked. The target group consisted of university scientific staff and it was assumed that all of them were capable of processing the arguments presented in the intervention.

The intervention consisted of: *i*) a leaflet explaining what is social engineering, why is it dangerous, how to detect it and what to do if confronted with it (refer to Figure 3.9), *ii*) a black key chain with the university logo on one side and the text "Don't give me to a stranger" on the other side (refer to Figure 3.2a and 3.2b respectively) and *iii*) a poster containing a humorous quote and an explicit remark against password, key and PIN sharing (refer to Figure 3.10). The leaflets were designed so that they could be processed via the central route, which was achieved by the absence of peripheral cues in it (Petty & Cacioppo, 1986). The leaflet represents the information media, the key chain the subtle reminder of the intervention and the humorous poster the cue that helps to remember the leaflet better. Departmental secretaries were responsible for distributing the material, they were unaware that this was part of an experiment. They were only instructed to distribute the material within their research group.

The leaflet and the poster were distributed by email, while the key chain was distributed in person. It is unknown whether any of the subjects printed the posters and displayed them in their office, however there were no reports of the intervention material being displayed in public areas (e.g. hallways, coffee corners or lunch rooms). All subjects were individually approached by a researcher between 10 a.m. and 6 p.m. on a 'normal' Wednesday during term time. In order to avoid suspicion, researchers (i.e. attackers) never made consecutive visits to members of the same department. After each visit, they therefore had to come back to the base of operations (i.e. the first author's office) to obtain the name and location of the next target which was randomly selected from a list of all possible targets.

The researchers were randomly assigned to either the authority or the control condition, that is wearing formal or casual clothing respectively. Subjects were randomly assigned to one of 4 conditions: they were either exposed to *i*) both authority and

(a) Front View (b) Back View

Figure 3.2: The key chain that was used in the intervention.

intervention *ii*) authority but not intervention *iii*) intervention but not authority, or *iv*) they were exposed to neither authority nor intervention.

Each researcher approached the subject in their office using the script described in Scenario 3.1.

Scenario 3.1. Obtain office key

> Hi, I am `[Name]` and I work for Facility Management. I have a question regarding the door locks. We received several complaints about the door lock and the keys. Has unlocking the door ever been problematic for you? We have contacted the manufacturer about the malfunctioning and they had received other similar complaints. In order to solve the problem, the manufacturer sent us a measuring device to test the keys that are in use. I have to admit that I don't exactly know what the box measures, but the data collected is necessary for the manufacturer to analyse the situation and hopefully find a solution to the problem. Can I have your key for measurement?
> *After measuring the key*: I have to inform you that after reading your key, the key has been reset and needs reactivation downstairs. It is no problem for me to reactivate the key for you.
> **Request**: Is it OK with you if I do the reactivation of your key downstairs?

Each target was subjected to the same request. After the researcher obtained the key and walked away, he/she came back to return the key and orally debriefed the subject with regards to social engineering. During the debriefing session, the subject was asked some demographic information, their opinion of physical and digital security at the university, length of employment, their route to the activation point and, for those who had not complied with the offender, to explain why they had not handed the key over.

3.2.4 Survey procedure

All subjects were individually approached by a researcher between 10 a.m. and 6 p.m. on a 'normal' Thursday during term time. The researcher approached the subject and asked if they wanted to participate in a 10 minute survey regarding social engineering. Upon agreement, the subject received a study information sheet explaining the nature of the study and an informed consent form to sign agreement of participation. The survey is based on Scenario 3.1 and asked the subject their behavioural intentions how they would react when in that particular situation. Furthermore, the subject was asked some demographic information and length of employment.

3.2.5 Variables

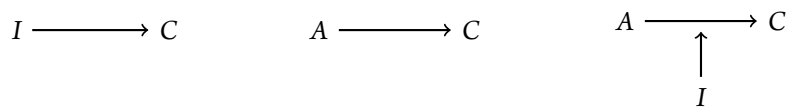
The variables used in the analysis were: compliance, intervention, authority, age, Years of Service (YoS), distance and floors. The dependent variable *compliance* measured whether the subject complied with the request of the offender to hand over the key. The dichotomous variable was dummy coded as 0 = did not comply, 1 = did comply. For those who filled in the survey the variable compliance measured the verbal compliance with the request (0 = would not comply, 1 = would comply). The independent variable *intervention* measured whether the subject was exposed to the intervention (0 = not exposed to the intervention, 1 = exposed to the intervention). The independent variable *authority* measured whether the offender (i.e. research assistant) wore casual clothing or wore formal clothing. Casual clothing was operationalized by wearing jeans and a t-shirt and formal clothing by a buttoned collar shirt and trousers (coded as 0 = informally dressed, 1 = formally dressed). The independent continuous variable *age* measured the age in whole years at the moment of the attack. The independent continuous variable *YoS* measured the seniority of an employee, operationalized as years of service (5 = 5 years of service). The independent continuous variable *distance* measured the distance the subject had to travel from the office to the activation point (1 = 1 meters of distance). The independent continuous variable *floors* measured the number of floors the subject had to travel from the office to the floor where the key activator is (1 = 1-floor difference). All variables regarding the field study were stored in the `f2f-se` dataset, whereas the variables regarding the survey (i.e. compliance, age and years of service) were stored in the survey dataset, refer to Table 1.2.

3.2.6 Analysis

The first three hypotheses were tested using Cross Tabulations and a Chi Square test. For the fourth and fifth hypothesis, logistic regression was used. The following three assumptions must be met for logistic regression analysis: *i*) sufficient sample size, *ii*) uncorrelated independent variables and *iii*) outlier-free dataset (Pallant, 2010). The Variance Inflation Factor (VIF) is 1.08 for both the authority and intervention variables, which is below the cut-off value of 10, indicating that there is no evidence of multicollinearity (Pallant, 2010). Only dichotomous variables were used, meaning that all variables are either 0 or 1, thus there are no outliers. To find out if intervention affects the relation between authority and compliance with the offender, a moderation analysis was carried out.

Compliance was predicted on the basis of *i*) model 1: (I) for intervention, *ii*) model 2: (A) for authority and *iii*) model 3: (AxI) for authority, intervention and the interaction between the two. The third model is needed to test whether the simultaneous influence of two variables on a third is non-additive. The relevance is that if authority and intervention interact, the relationship between each of the interacting variables and compliance depends on the value of the other interacting variable.

Finally, the three models (refer to Figure 3.3a, 3.3b and 3.3c) were compared to each other on the basis of the variance and likelihood ratio.



(a) Model 1. (b) Model 2. (c) Model 3.

Figure 3.3: Overview of the three models.

3.2.7 Buildings

Building 1 in this analysis is a modern building and has a circular lay out. This building has 5 floor levels, refer to Figure 3.4 for a blue print of the third floor of the building. The main entrance is near activator C, the side entrance is near activator D and there is an entrance from a passage from an other building near activator A.

Building 2 is a building in the traditional sense, there is 1 straight corridor with offices on both sides. This building has 5 floor levels, refer to Figure 3.5 for a blue print of the third floor of the building. There are entrances near both activators.

3.3 Results

A total of 162 subjects were approached. No 'building occupied by target' effect on compliance (with the offender) ($N = 162$, $df = 1$, $\chi^2 = .728$, $p = .394$), 'offender's sex' effect on compliance ($N = 162$, $df = 1$, $\chi^2 = .621$, $p = .431$), 'target's sex' effect on compliance ($N = 161$, $df = 1$, $\chi^2 = 2.143$, $p = .143$), 'age' effect on compliance ($N = 155$, $OR = .989$, $p = .434$) or 'years of service' effect on compliance ($N = 109$, $OR = .990$, $p = .596$) were found and these are therefore not further mentioned.

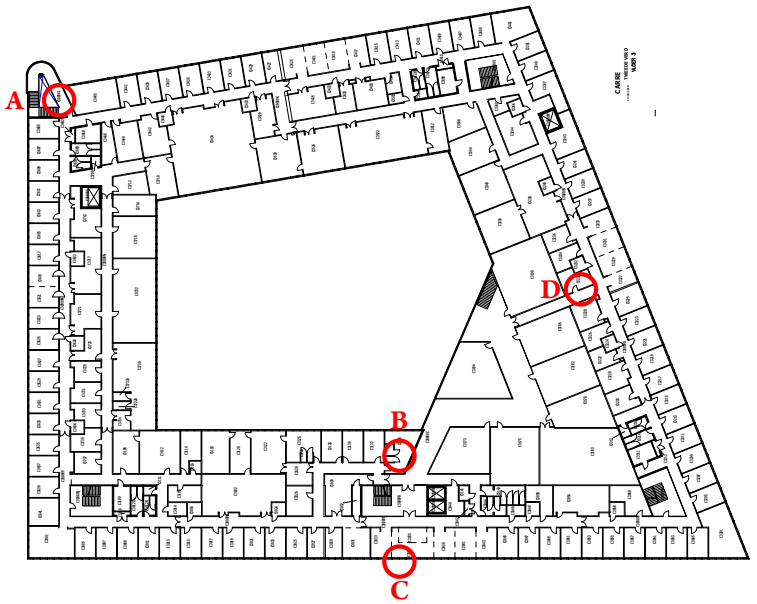


Figure 3.4: Layout building 1, the red circles indicate the location of the digital key activators. Activator A is on floor 2; B is on floor 3; C is on floor 1; D is on floor 1.

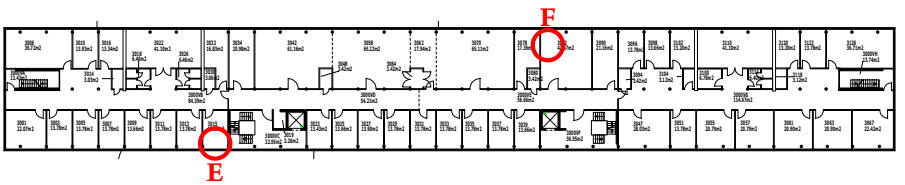


Figure 3.5: Layout building 2, the red circles indicate the location of both digital key activators on floor 1.

3.3.1 *H1*: "The compliance in the survey is unequal to the compliance in the experiment."

Those who were not exposed to an intervention and participated in the field study, 58.62% complied with the request of the offender compared to 3.1% of those who filled in the survey ($\chi^2 = 30.139$, $df = 1$, $p = .000$). Hypothesis *H1* is therefore accepted. Refer to Table 3.1 for descriptive statistics.

Table 3.1: Number of observations and percentages for each type of test.

		TEST		
		SURVEY	FIELD	TOTAL
COMPLIED	NO	30 (96.77%)	48 (41.38%)	78 (53.06%)
	YES	1 (3.23%)	68 (58.62%)	64 (46.94%)
TOTAL		31 (100%)	116 (100%)	147 (100%)

Group survey = field ($\chi^2 = 30.139$, $df = 1$, $p = .000$);

3.3.2 *H2*: "The subjects in the intervention and control groups comply unequally."

Of the targets who were not exposed to the intervention (control group), 58.62% agreed to the request of the offender compared to 36.96% of those in the intervention group ($\chi^2 = 6.199$, $df = 1$, $p = .013$). Hypothesis *H2* is therefore accepted. Those in the control group have 2.42 times higher odds of compliance (i.e. handing over their keys) than those who were exposed to the intervention ($OR = 2.42$, $CI\ [1.20, 4.88]$). Refer to Table 3.2 and Figure 3.6 for descriptive statistics.

Table 3.2: Number of observations and percentages per intervention condition.

		INTERVENTION		
		NO	YES	TOTAL
COMPLIED	NO	48 (41.38%)	29 (63.04%)	77 (47.53%)
	YES	68 (58.62%)	17 (36.96%)	85 (52.47%)
TOTAL		116 (100%)	46 (100%)	162 (100%)

Group control = intervention ($\chi^2 = 6.199$, $df = 1$, $p = .013$);

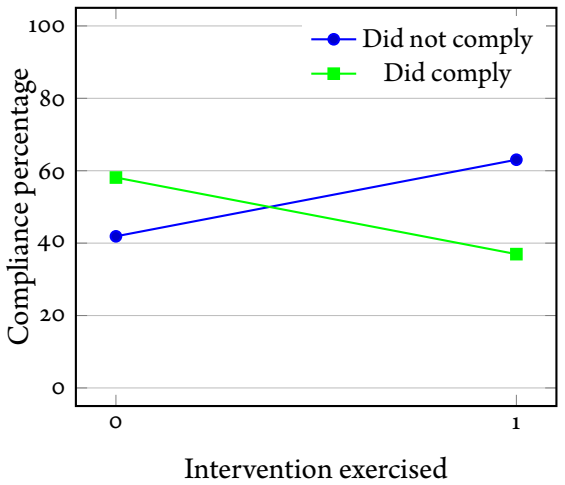


Figure 3.6: Predicted change of compliance due to the intervention, based on the regression coefficients.

3.3.3 H_3: "The subjects in the authority and control groups comply unequally."

In total 53% of the targets in control group (those approached by casually dressed offenders) complied compared to 51.61% of those in the authority group ($\chi^2 = .030$, $df = 1$, $p = .864$). H_3 is rejected in favour of the alternative hypothesis H_{3a}: "Authority and Control comply equally". Those in the control group have 1.06 times higher odds of compliance (i.e. handing over their keys) than those that were exposed to authority ($OR = 1.06$, $CI\,[0.56, 1.99]$). Refer to Table 3.3 and Figure 3.7 for descriptive statistics.

Table 3.3: Number of observations and percentages per authority condition.

		AUTHORITY		
		NO	YES	TOTAL
COMPLIED	NO	47 (47%)	30 (48.39%)	77 (47.53%)
	YES	53 (53%)	32 (51.61%)	85 (52.47%)
TOTAL		100 (100%)	62 (100%)	162 (100%)

Group control = intervention ($\chi^2 = .030$, $df = 1$, $p = .864$);

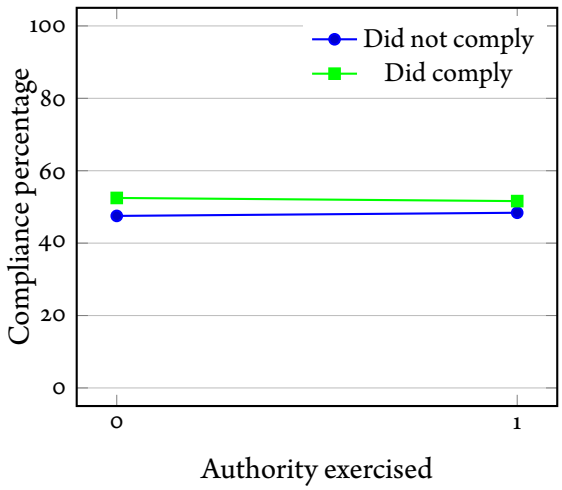


Figure 3.7: Predicted change of compliance with the offender due to authority, based on the regression coefficients.

3.3.4 *H4*: "An intervention reduces the effect of authority in relation to the compliance with a request."

It was tested whether the relation between authority and compliance was affected by the intervention. In the model with the interaction term (refer to Model 3: AxI in Table 3.4 and Figure 3.8), this variable is marginally significant ($p = .052$). The hypothesis *H4* is rejected in favour of the alternative hypothesis *H4a*, thus there is only a tendency for Intervention to moderate (i.e. affect) the relation between Authority and Compliance. Refer to Table 3.4 and Figure 3.8 for multiple regression results.

The comparison of the three models (refer to Table 3.4), shows a significant difference between model 1 & 2 and between 2 & 3. The pseudo R^2 shows zero percentage of explained variance for model 2 (authority), while in model 1 (intervention) this number increased.

3.3.5 H5: "The amount of physical effort influences the willingness of an employee to hand over his office key to a stranger."

Of the 116 targets, 68 (58.62%) handed over their key. There was no significant difference in the distance from the office to the activator for those who did not surrender their key ($M = 27.2$, SD = 21.2) and those who did surrender their keys ($M = 26.2$, SD = 23.3); $\chi^2 = .447$, $df = 1$, $p = .504$. For an overview of distances per building per floor, refer to Table 3.5. The distance was tested for each building individually; there

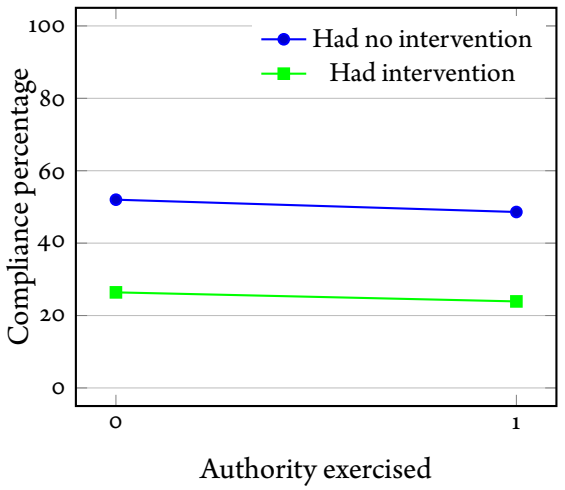


Figure 3.8: The effect of intervention on authority, based on the regression coefficients.

was no distance effect found for either building (Building 1: $\chi^2 = .222$, $df = 1$, $p = .637$, Building 2: $\chi^2 = 2.086$, $df = 1$, $p = .149$).

Combining distance, number of floors and building had no effect on compliance, refer to Table 3.6. Hypothesis $H5$ is therefore rejected in favour of hypothesis $H5a$: "The amount of physical effort does not influence the willingness to hand over your office key to a stranger".

3.4 Conclusion

This study investigated whether an intervention in the form of an awareness campaign affects the relation between the authority principle of persuasion and compliance with the request to hand over the office key to a stranger impersonating a facility management staff member.

An intervention composed of *i*) informing people about the risks of social engineering attacks *ii*) distributing a small key chain and *iii*) a humorous poster, proved to have a large impact on the likelihood of handing over office keys to strangers.

There was a difference between the rate of verbal compliance with the offender's request (3.2%) and the behavioural compliance (58.62%). This difference can be explained by the optimism bias; People believe that positive events are more likely to occur to them than to other people (Weinstein, 1980). The observed discrepancy between verbal and behavioural compliance should be taken into account when devel-

Table 3.4: Model comparison of the three models. The columns depict for each variable: the odds ratio (*OR*), its lower and upper 95% confidence intervals [in brackets] and its significance level.

	MODEL 1: (I)	MODEL 2: (A)	MODEL 3: (A*I)
INTERVENTION	0.41 [0.21, 0.84]*		0.44 [0.21, 0.90]*
AUTHORITY		0.95 [0.50, 1.78]	1.10 [0.56, 2.16]
AUTH*INT			0.23 [0.05, 1.01]a
CONSTANT	1.42 [0.98, 2.05]	1.13 [0.76, 1.67]	1.41 [0.92, 2.16]

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$; $^{a} = .052$
auth*int = interaction of authority and intervention;
Model 1 ($\chi^2 = 6.24$, $p = .013$), $N = 162$, Pseudo $R^2 = .028$;
Model 2 ($\chi^2 = 0.03$, $p = .864$), $N = 162$, Pseudo $R^2 = .000$;
Model 3 ($\chi^2 = 10.19$, $p = .017$), $N = 162$, Pseudo $R^2 = .045$;
Model 1=2 ($p = .045$); Model 1=3 ($p = .139$); Model 2=3 ($p = .006$);

Table 3.5: Overview of distances to the activator per building per number of floors for those who did and did not comply. The columns depict for each floor: the average distance, its standard deviation (in brackets) and its sample size.

		BUILDING 1		BUILDING 2	
		DIDN'T COMPLY	DID COMPLY	DIDN'T COMPLY	DID COMPLY
FLOORS	0	31.89 (19.34) 7	47.28 (31.20) 5	-	-
	1	39.12 (12.49) 5	28.40 (14.75) 12	21.70 (8.20) 3	18.60 (11.39) 5
	2	35.40 (40.57) 4	39.60 (25.54) 3	23.08 (13.07) 9	12.66 (6.26) 12
	3	37.20 (0) 1	40.80 (0) 1	20.67 (3.20) 6	11.37 (8.47) 6
	4	-	-	8.41 (5.28) 7	15.09 (8.36) 15
TOTAL		35.15 (22.31) 17	35.09 (21.17) 21	18.23 (10.56) 25	14.19 (8.17) 38

oping and distributing countermeasures. Since people don't see social engineering as an urgent problem that victimizes them, why would the accept any countermeasure? A likely thought of an employee could be: "I am less likely to be targeted than my colleague, and if I was targeted, I am better in resisting than my colleague. Therefore this countermeasure does not apply to me."

In total, 36.96% of staff exposed to the intervention versus 58.62% in the control group complied with the request to hand over their keys. Not exposing staff to the intervention means that the odds of them handing over their keys are 2.42 times higher. The priming effect is in line with the results from studies on Internet related crimes (Ferguson, 2005; Hight, 2005; Kumaraguru et al., 2010) and those in the medical sciences (Lancaster & Stead, 2005). The effect of leaflets has proved to be effective in the fields of promotion of healthy behaviour, advertisement and phishing (Lancaster

Table 3.6: Comparison of the four models. The columns depict for each variable: the odds ratio (OR), its lower and upper 95% confidence intervals [in brackets] and its significance level.

	MODEL 1: (D)	MODEL 2: (F)	MODEL 3: (DF)	MODEL 4: (DFB)
DISTANCE	1.00 [0.98, 1.02]		0.99 [0.97, 1.02]	0.99 [0.96, 1.02]
FLOORS		1.06 [0.80, 1.41]	1.04 [0.74, 1.44]	1.06 [0.71, 1.59]
BUILDING				0.88 [0.27, 2.83]
CONSTANT	1.43 [0.80, 2.58]	1.09 [0.54, 2.20]	1.55 [0.50, 4.86]	1.87 [0.25, 14.0]

* $p < .05$; ** $p < .01$; *** $p < .001$;
D = distance; F = floors; B = building;
Model 1 ($\chi^2 = .02$, $p = .896$), $N = 110$, pseudo $R^2 = .000$;
Model 2 ($\chi^2 = .19$, $p = .665$), $N = 112$, pseudo $R^2 = .001$;
Model 3 ($\chi^2 = .77$, $p = .681$), $N = 102$, pseudo $R^2 = .006$;
Model 4 ($\chi^2 = .82$, $p = .846$), $N = 102$, pseudo $R^2 = .006$;

& Stead, 2005; Grewal & Kavanoor, 1997; Kumaraguru et al., 2010). Existing literature shows that the processing of information in leaflets tends to cause behavioural change (e.g. healthier behaviour, increased purchase actions or a lower response rate to phishing emails). Our research confirms that informing people also leads to a change in behaviour. Knowledge about social engineering attacks helps to increase the safety of the organisation and decreases the susceptibility to victimisation.

Allowing offenders to make changes to the context by applying principles of persuasion should increase their probabilities of success. However, in our study the expected effect of authority, operationalized via formal clothing was not validated. Facility personnel at this particular university are formally dressed (since they wear black trousers and a blue shirt). It was therefore expected that informally dressed attackers would be viewed as having less power.

A possible explanation could be that countries differ in the perception of authority. To the best of our knowledge, the related research has all been carried out in The United States (Milgram, 1963; Bickman, 1974; Lefkowitz et al., 1955; Cialdini, 2009). It is possible that Dutch people are less sensitive towards authority than Americans. However, there is no evidence of such variation, with respect to sensitivity to authority. Research on cultural differences shows that both countries score equally (Hofstede, Hofstede, & Minkov, 2010).

Hofstede identified six cultural dimensions and made a comparison across 60 countries. The six dimensions are: *i*) Indulgence versus Restraint, *ii*) Power Distance (PDI), *iii*) Individualism versus Collectivism, *iv*) Uncertainty Avoidance, *v*) Long-term versus short-term orientation and *vi*) Masculinity versus Femininity (Hofstede et al., 2010). The cultural dimension 'Power Distance' was thought to be one that could

explain the perception of hierarchical structures and, accordingly the difference between the former and present findings. Three aspects indicate obedience to authoritative figures as in (Bickman, 1974; Lefkowitz et al., 1955; Milgram, 1963): *i*) "Hierarchy in organisations reflects the existential inequality between higher-ups and lower-downs", *ii*) "Subordinates expect to be told what to do" and *iii*) "Parents teach children obedience" (Hofstede et al., 2010). Citizens from The United States and The Netherlands score similarly on Power Distance (Hofstede et al., 2010) and consequently. However, in the present study not all staff was Dutch. Future research should involve controlling for the country of origin of subjects.

Perhaps the experimental setting of the present study (i.e. a university) explains the lack of effect of authority. In addition, this university has a hierarchical structure that is somewhat 'flat'; the distance between the 'ranks' is relatively small. The contact between staff and their supervisors is on a first name basis and it is common practice to walk into someone's office without an appointment. In contrast, studies from The United States found in the literature took place in real life situations on the street (Bickman, 1974; Lefkowitz et al., 1955) or a publicly available laboratory (Milgram, 1963).

A second alternative explanation might relate to the experimental design. Subjects who volunteer to participate in experiments might experience authority differently from a psychological point of view compared to those who are 'forced' to participate. Volunteer subjects might have developed a sense of commitment towards those running the experiment (e.g. researcher). According to Milgram (1974), commitment is the force that binds both the subject and authority to their role. The design of our experiment was different to the electro shock experiment, in which all subjects volunteered to participate (Milgram, 1963). The same reasoning applies to the illegal intersection crossing experiment (Lefkowitz et al., 1955), where the subjects volunteered to participate. In our experiment, on the other hand, subjects were randomly selected and 'authority' was applied directly upon entrance into the office. The subjects therefore had less psychological binding and commitment to the authority, and thus were less likely to comply with the request of the offender.

When authority and intervention were entered together in a model, authority still lacked predictive power. It is possible that this was caused by the age of the offender. The average age of the researchers was slightly above 20 years old, whilst studies from previous research indicated that the experimental authoritative figure was over 31 years old (Bickman, 1974; Lefkowitz et al., 1955; Milgram, 1963).

Given that 58.62% of staff handed over their office key, it seems that the currently used policies on this topic are in need of either clarification or proper dissemination.

The results of the experiment were discussed with security and facility management personnel, who were surprised about the compliance rates. Moreover, the assumption that there is no need to make a policy explicit, since it is common sense not to give away bicycle, home or car keys to strangers was, found to be wrong.

The likelihood of handing over the key for employees close to the activation point was similar to that of those who were further away. There was no difference between the two buildings. We therefore conclude that, in this context, there is no effect of effort on the compliance to a social engineering attack. An explanation for this could be that people do not consider the effort when deciding to surrender the key or not.

3.4.1 Future work

Finally, we present recommendations for future research. First, the intervention encompasses three components: *i*) a leaflet informing people about the risks of social engineering attacks; *ii*) a small key chain and *iii*) a humorous poster. Because each target in the intervention group was exposed to all three components, the measurement of the individual effect sizes was outside the scope of the current research and is hence considered as future research. The effect of each individual component (i.e. the key chain, the leaflet, the humorous poster) and their combination, should be tested in order to maximize the effectiveness of social engineering interventions. Second, a follow-up study in the form of a time-series analysis could be carried out to test whether the intervention effect is maintained over a period of time. If the effect does not hold, an evaluation of the time decay could provide practitioners in the field of security with relevant information for scheduling interventions. Furthermore, there is also a need to evaluate the effectiveness of reminders over time. Time as an function of compliance is tested in the study described in Chapter 5. Finally, it is possible that the likelihood of handing over the office key is related to the level of security inside the office. An extension of the study could involve identifying which of the staff have security features installed such as Kensington locks or locked cabinets.

Memo September 4, 2013

Do you share your PIN, key or password?

Your PIN, key & password, belong to you, are yours only and not meant to share. Sharing this is dangerous and therefore by many organisations forbidden, this also is forbidden by the University of Twente. Once this information are shared it is out of your control and could cause harmful situations with catastrophic consequences to both you and the organisation. Such as fraud or information theft as can be seen in the video[1].

You could also be victim of an information thief. Most dangerous part of these attacks is that you would not notice when you are under attack. Requests that look legitimately and could be explained rationally are hard to distinguish from unjustly requests. On the occasion that someone claims to be from the IT department and asks for you IP-address, this all sounds all justifiable. How to be sure that the person you are talking to is really who he claims to be? Challenge the requester with specific questions that validate the identity.

Information thieves more often exploit end users of a system to bypass security mechanisms. Technical security mechanisms grow stronger and become harder to abuse. Most employees are no expert in the area of computer systems, and this is known by thieves. This makes every employee interesting for attackers and this important in the security chain. Especially since we are helpful by nature.

Attackers use identities from someone else to trick targets into revealing sensitive information. Effective ways to do this is to pick someone who has authority, for example a manager, a lawyer or a doctor. Experiments show that a stranger was able to collect 60% of the username and password combinations within a medium sized company, just by asking employees for it.

Information thieves abuse employees to get access to computer networks. By impersonating people with authority, thieves are successful in obtaining user credentials and other useful information. This is all done, without the victim noticing it. If someone asks for information related to the computer system, validate their identity before you give any information. And never ever share your password, key or PIN with anyone. Help protecting yourself against information thieves with the special awareness key-chain. This key-chain is available for you at your secretary.

Do Not. . .	Do. . .
• give your credentials away	• challenge the requester
• say Yes to often	• be critical and suspicious
• share valuable information via a phone call	• ask for name + phone number and redial

[1] http://www.youtube.com/watch?v=IbXfCrCjHWA

UNIVERSITY OF TWENTE.

Figure 3.9: Memo used in the intervention.

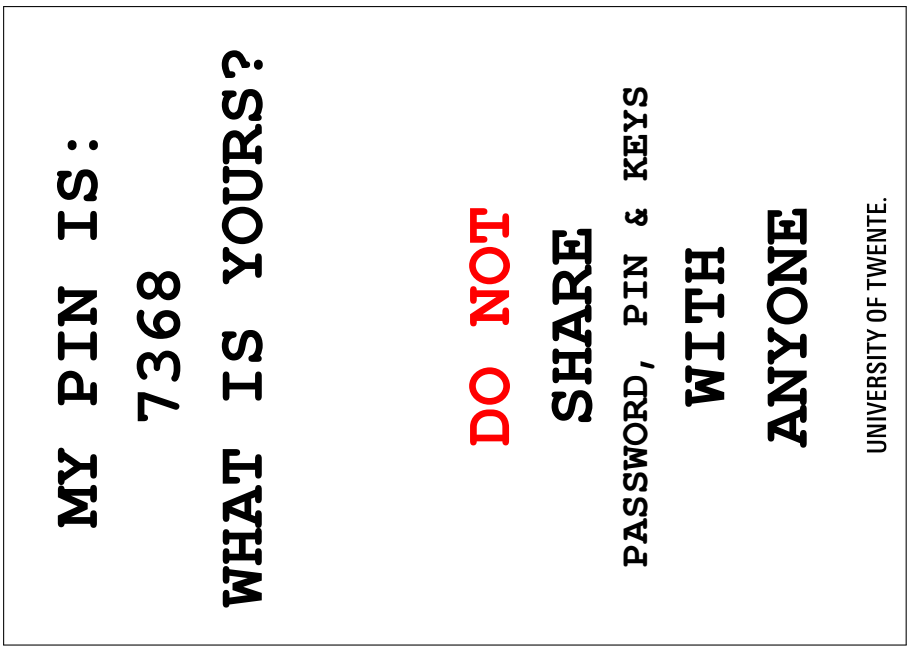


Figure 3.10: Poster used in the intervention.

Stay open minded. Things aren't always what they seem to be.

–Scottie Waves

4

Email social engineering

This chapter is based on (Bullée, Montoya, Junger, & Hartel, in press).

4.1 Introduction

In the previous chapter, a Face-to-Face (F2F) social engineering experiment was discussed. The result wast that the intervention has a significant effect on reducing compliance, but the same was not the case for authority and distance. This chapter describes an email social engineering experiment (i.e. phishing). Innovative in this chapter is the explanation of spear phishing using four target demographic variables i.e. gender, age, length of employment and cultural background.

Cyber security has for a long time primarily been treated as a technical problem (Rhee et al., 2009; Waldrop, 2016). However, cyber security incidents are often caused by human failure (Chan et al., 2005) rather than by technical failure (Schneier, 2000). Developing stronger digital security alone will not result in a viable long term solution against cyber security. Instead, the solution should involve solving human errors (Waldrop, 2016). Since people do not have the cognitive capacity to process all information, decision-making involves using rules of thumb (i.e. heuristics) (Cialdini, 2009). These mental shortcuts (resulting from experience and genetics) work well in most circumstances and increase the likelihood of solution on a task or problem. However, heuristics might not always lead to the correct solution and result in a cognitive bias instead (Gigerenzer, 1991; L. Harris, 2007; Lewis, 2012; Tversky & Kahneman, 1974). Offenders are well aware of the flaws in human logic and nudge the heuristics of their targets into systematic errors (i.e. cognitive biases) to make them comply (Bosworth et al., 2014; Dang, 2008; Kennedy, 2011; Luo et al., 2011; Twitchell, 2009). This kind of trickery is referred to as social engineering. This paper focusses on social engineering via email, also known as phishing. In particular, it aims to establish whether victimisation differs for general and spear phishing (i.e. targeted or personalised) emails in the context of an organisation. Secondly, it aims to establish whether sociodemographic characteristics of targets influence victimisation. The reason behind this is 3-fold: *i*) it finds out who is most vulnerable and could benefit most from training, *ii*) it reduces cost and time for those who give and receive training, *iii*) it prevents training fatigue and adverse training effects. It is argued that the possibility of adverse effects emphasises the need to study the effectiveness of interventions before their launch (Junger, Montoya, & Overink, 2017).

Some have argued, that most forms of cyber crime are not unique to the on-line world since they have long-established terrestrial counterparts (Grabosky, 2001; McCusker, 2006; Neve & van der Hulst, 2008) which pre-date the Internet but have found new forms of life on-line. For example, hacking activities could be seen as computer-aided versions of trespassing as the attacker is entering another person's

property without authorization. In addition, when a hacker purposely changes a website or destroys data, the action is comparable to vandalism. Similarly, phishing emails are comparable to the classical confidence tricks (e.g. scams or fraud) leading to theft (Montoya, Junger, & Hartel, 2013). The *Modus Operandi* (MO) include both building a trust relation with the target and then using psychological tricks (e.g. abuse the credulity of the target) to defraud it.

Data collection for both traditional and digital past crime experiences often involves finding case studies and filling in surveys, e.g. (J. Lee & Soberon-Ferrer, 1997; Titus, Heinzelmann, & Boyle, 1995). Although these are useful methods for data collection, the following three issues can be identified: *i*) the representativeness of the sample, *ii*) controlling for opportunity and *iii*) unawareness of victimization. For example, since there are case studies stating that older adults were victimised, it is therefore assumed that the elderly are more vulnerable. However, these individual case studies do not constitute a representative sample of victims (Ross, Grossmann, & Schryer, 2014). Carrying out a survey on people can be used to overcome this. However, simply asking people for their experience regarding fraud will bias the outcome. In many cases, there is no control for an opportunity (i.e. whether one receives an attempt) in the survey (Ross et al., 2014). Not being exposed to a fraudulent request will never make one become victimised. Only a few studies take this into account, e.g. Titus et al. (1995). Another drawback of surveying fraud is that some respondents were unaware of being defrauded, forgot the episode, misremembered or felt too ashamed to admit (Ross et al., 2014).

It is argued that it is best to do experiments and observe behaviour rather than to either ask subjects how they think they would behave in a given scenario or to recall a reaction (Petrova, Cialdini, & Sills, 2007). When experimenting in an organisational context, Pfeffer (1985) argues that inclusion of sociodemographic variables helps better understand the organisation. Employees are not a homogeneous group of entities and are hence diverse regarding e.g. age and Years of Service (YoS). The increase of women and different ethnic groups in the workforce has been further increased diversity (Shenhav & Haberfeld, 1992). It must be noted that conducting experiments regarding fraud and cyber crime requires careful planning and consideration. Since this typically involves conducting experiments on humans (e.g. employees), ethical considerations must be taken into account (Belmont Report, 1979). Particular challenging is the use of deception since it conflicts with ethical principles (Code of Federal Regulations, 2005). Furthermore, people who are aware of being in an experimental setting would be suspicious and hence biassed. It is unlikely that they would have similar levels of suspicion outside of the experiment (Anandpara, Dingman, Jakobsson,

Liu, & Roinestad, 2007; Furnell, 2007; Parsons, McCormac, Pattinson, Butavicius, & Jerram, 2013).

4.1.1 Phishing

"Phishing is a scalable act of deception whereby impersonation is used to obtain information from a target" (Lastdrager, 2014). The definition contains two parts: *i*) scalability relates to how easy it is for the offender to approach the targets it bulk. Therefore, all non-mass-media (i.e. F2F interaction or telephone calls) do not constitute phishing. *ii*) deception by impersonation to obtain information, is for example, claiming to be from someone's bank to obtain information. When the offender does not use impersonation, it is a non-fraudulent request for information, and therefore, it cannot be classified as phishing.

Spear Phishing

Spear phishing is a particular type of phishing, in which the target and context are investigated so that the email is tailored to receiver. If the process of personalization is scalable, spear phishing falls in the consensus definition of phishing (Lastdrager, 2014). The rationale behind putting additional effort in personalising the emails relates to the 'higher return on investment'. It is believed that spear phishing emails are successful because personalisation creates trust (Sparshott, 2014).

The Rational Choice Theory (RCT) (Cornish & Clarke, 1987) provides underpinning for our study. Actions in this theory are based on a conscious evaluation of the utility of acting in a certain way (i.e. cost vs benefit). In the field of crime and security, this translates into the weighing of the value to be gained by committing the act versus the negative consequences (Cornish & Clarke, 1987). J. H. Hong states that offenders are changing their focus from a wide to narrow range of targets (J. H. Hong, 2012). Offenders used to send out mass emails hoping to trick anyone; now they are more selective and use relevant context information in emails to trick specific targets (J. H. Hong, 2012). Based on this theory, we hypothesised that spear phishing is more attractive for two reasons: *i*) it involves targeting fewer people than in general phishing, the likelihood of a negative consequence is lower because the number of attempts the offender has to make is lower and *ii*) since the emails are personalised, the target is more likely to assume that the email is legitimate, translating into a higher compliance rate.

There is limited empirical work on the effectiveness of spear phishing. The Scopus query "`TITLE-ABS-KEY (spear phishing)`" returned 66 articles; the majority

of the studies discusses computer algorithms that aim to detect spear phishing emails, whereas there are 3 studies about sending (spear) phishing emails to human subjects. However, none of them investigated the success of spear phishing for different types of people. Nevertheless, studies that did investigate sociodemographic variables and 'general' phishing are discussed next.

In a field study, the effect of spear phishing was tested on 581 college students (age 18-24 years). Those in the control group ($N = 94$) received a phishing email from a fictitious person requesting to enter their login credentials on an untrusted website, whereas those in the experimental group ($N = 487$) received the same email supposedly from one of their friends. The subjects who received the email from a 'friend' entered their credentials 4.5 times more often than those who received the email from a stranger (16% vs 72%) (Jagatic et al., 2007).

In another field study, 158 employees distributed over five organisations in Sweden were approached under 2 conditions (Holm et al., 2014). All employees first received the general email (written in English) with the request to download software from an untrusted website. Later, all employees received another email (written in Swedish), using the name of the employee, the name of the organisation and the name of an executive to persuade them to download an add-on to the virus scanner. Those who received a spear phishing email were 5.3 times more likely to click the link in the email (27.2% vs 5.1%) and 2.8 times more likely to execute a binary (8.9% vs 3.2%) than those who received a general phishing email (Holm et al., 2014).

The following four sections discuss the influence of gender, age, years of service and culture on the compliance to phishing.

Age

A survey of a national representative sample including 957 adults in The United States (US) found a negative significant effect of age on victimisation, meaning that older persons are less likely to be victims of fraud (J. Lee & Soberon-Ferrer, 1997). It was not mentioned whether the survey asked if the subject had been exposed to attempted fraud, hence there was no control for opportunity. A national telephone survey, representing a relative probability sample of 1246 respondents also found a negative age victimisation relation (Titus et al., 1995). Titus et al. argued that older people (i.e. 65 and above) are more experienced regarding fraud and therefore less vulnerable (Titus et al., 1995). In a 400 respondent telephone survey, comparable to that of Titus et al., no age effect was found (Van Wyk & Benson, 1997). The latter two surveys did control for opportunity.

For phishing emails Sheng et al. (2010) collected data regarding phishing experience and victimisation in an online survey among 1001 respondents (containing a mix of US and non-US citizens, students and non-students) using Amazon.com's Mechanical Turk. Of their subjects, 52% indicated that they would click on links in the phishing emails. Furthermore, they found a negative relation between age and falling for phishing. People in the age group 18-25 were more likely to fall for phishing than people in other age groups (Sheng et al., 2010). The authors speculate that this effect is due to this age group having: *i*) a lower level of education, *ii*) fewer years on the Internet, *iii*) less exposure to training materials and *iv*) less of an aversion to risks. Research results for traditional and digital crime show either a negative or no effect of age. There is therefore not enough evidence to suggest that age should be discarded as a predictor of phishing. For a summary of the studies, refer to Table 4.1.

Gender

In traditional fraud there is generally no gender effect (J. Lee & Soberon-Ferrer, 1997; Titus et al., 1995; Van Wyk & Benson, 1997). However, a gender effect was found for specific types of fraud. Females were more frequently victims of lottery fraud and males of investment fraud (Deevy, Lucich, & Beals, 2012).

The 1001 respondents in the online survey conducted by Sheng et al. (2010) found that females fell more for phishing emails than males. The subjects in their sample had an average age of 30 and 48.25% ($N = 483$) were males (Sheng et al., 2010). After a training session females and males performed equally (Sheng et al., 2010).

In two studies, university students received a phishing email in their mailbox (Jagatic et al., 2007; Wright et al., 2014). In the control condition, 16% and 2.4% of the participants complied with the phishing email respectively. Furthermore, the outcome was that females were more likely to respond than males. Finally, there was one study in which the subjects received a variety of both legitimate and phishing emails (K. Hong et al., 2013). This latter design allows testing whether the participants in a phishing study become paranoid and classify all emails as phishing. The finding of K. Hong et al. (2013), based on seven phishing and seven non-phishing emails, was that 92.5% of their participants classified at least 1 phishing email as legitimate. The performance of females was worse than that of males at identifying phishing emails (K. Hong et al., 2013). These results suggest that females are overall more vulnerable than males to phishing emails (K. Hong et al., 2013; Jagatic et al., 2007; Sheng et al., 2010; Wright et al., 2014). For a summary of the studies, refer to Table 4.1.

Years of Service

In organisational research, the variable tenure or seniority is often included in the analysis and commonly operationalized as Years of Service (YoS) at an employer. Years of service correlates with job satisfaction in e.g. academic personnel (Oshagbemi, 2000) hospital employees (Mobley, Horner, & Hollingsworth, 1978) and insurance company clerks (Waters, Roach, & Waters, 1976). An explanation is that employees who are less satisfied will resign while those who are more satisfied will remain in a job. Years of service is also positively correlated with occupational commitments in e.g. nurses (Jafari Kelarijani, Heidarian, Jamshidi, & Khorshidi, 2014) and hotel employees (Sarker, Crossman, & Chinmeteepituck, 2003). Furthermore, there is also a relation with age, since the YoS cannot exceed the age of a person minus the years of education.

The relation between YoS and victimisation by phishing emails was investigated in (to the best of our knowledge) only one field study (Kearney & Kruger, 2014). The 213 employees involved were persuaded to validate their password on an untrusted website, 63.7% of the recipients responded to the email. A negative YoS age relation was found, meaning that employees who were hired more recently were more often victimised compared to those who were hired less recently (Kearney & Kruger, 2014). Despite the limited empirical evidence, this relation could be important in organisational research on phishing.

Cultural dimensions

Hofstede et al. (2010) identified six cultural dimensions and made a comparison across 60 countries. The six dimensions are: *i*) Power Distance (PDI) , *ii*) individualism versus collectivism (IDV), *iii*) uncertainty avoidance (UAI), *iv*) masculinity versus femininity (MAS), *v*) long-term versus short-term orientation (LTO), and *vi*) indulgence versus restraint (IND) (Hofstede et al., 2010). The cultural dimension Power Distance is one that could explain the different success rates of phishing emails between citizens of countries. Power Distance is defined as "the extent to which the less powerful members of organisations and institutions accept and expect that power is distributed unequally" (Hofstede et al., 2010). A higher PDI score suggests that there is a strongly enforced hierarchy. A lower score suggests that people question authority and strive to decentralise and distribute power more equally. Two aspects indicate obedience to the use of authority in phishing emails: *i*) "Hierarchy in organisations reflects the existential inequality between higher-ups and lower-downs" and *ii*) "Subordinates expect to be told what to do" (Hofstede et al., 2010).

In the context of information security behaviour, limited research has been conducted on cultural influences. The majority of studies have been conducted in Western countries, occasionally in Asia, whereas the rest of the world has been overlooked. Cross-cultural research is important since culture is likely to have a direct influence (Crossler et al., 2013).

In a survey, 50 US and 61 Indian participants were asked about phishing. Almost everyone in the sample had experienced a phishing attempt. In total 14% of the US and 31% of the Indian participants reported being victimised (Tembe et al., 2013). The results suggest that Indians are more susceptible to phishing emails. India has a higher PDI (i.e. 77) compared to the US (i.e. 40); this could explain the difference between the two countries. It must also be noted that the subjects from India were significantly younger than those in from the US sample, therefore an age effect is not excluded either (Tembe et al., 2013).

4.1.2 Research Question

The contribution of our work is two-fold. First, we provide an experimental design for measuring the effectiveness of two types of phishing emails, which provides real-life empirical data (as oppose to laboratory data or measuring intention). Second, our study gives insight into how the sociodemographic characteristics of victims predict compliance.

This research aims to answer the question: *"To what extent are people susceptible to phishing emails?"* Five hypotheses were formulated:

H1) Previous research showed that combining the name of the recipient, the name of the organisation, the name of a company executive and a translation to the native language was successful. However, no individual effect was investigated. We therefore hypothesise that the opening sentence of a phishing email influences its success.

H2) Previous research showed that females were more vulnerable to phishing emails than males. We therefore hypothesise that the success of a phishing email is influenced by the gender of the recipient and that females are more vulnerable.

H3) Previous research regarding phishing emails was inconclusive regarding the effect of age on compliance. We therefore hypothesise that the success of a phishing email is influenced by the age of the recipient and that older people are more vulnerable.

H4) Since there is limited research regarding the length of employment in the organisation in relation to phishing which involves the organisation, this variable is included in the analysis. We therefore hypothesise that the success of a phishing email is influ-

Table 4.1: Summary table of presented literature on phishing studies and sociodemographic predictors.

					MOST AT RISK						
N	TYPE	POPULATION[1]	OPPORTUNITY	DATA COLLECTION[2]	TYPE[3]	GENDER	AGE	YoS[4]	CULTURE	SUCCESS[5]	REF
957	Classical	Mix	Uncontrolled	Telephone Survey	-	No effect	Younger	-	-		(J. Lee & Soberon-Ferrer, 1997)
1246	Classical	Mix	Controlled	Telephone Survey	-	No effect	Younger	-	-		(Titus et al., 1995)
400	Classical	Mix	Controlled	Telephone Survey	-	No effect	No effect	-	-		(Van Wyk & Benson, 1997)
158	Digital	Staff	Controlled	Single exp. email	Spear	-	-	-	-	5.1	(Holm et al., 2014)
581	Digital	Students	Controlled	Single exp. email	Spear	Female	-	-	-	16	(Jagatic et al., 2007)
1001	Digital	Mix	Uncontrolled	Role-play task	-	Female	Younger	-	-	52	(Sheng et al., 2010)
2624	Digital	Student	Controlled	Single exp. email	-	Female	-	-	-	2.4	(Wright et al., 2014)
53	Digital	Students	Controlled	Multi exp. emails	-	Female	-	-	-	92.5[6]	(K. Hong et al., 2013)
111	Digital	Mix	Controlled	Single exp. email	-	-	Younger	-	High PDI	23.4	(Tembe et al., 2013)
490	Digital	Staff	Controlled	Single exp. email	-	-	-	Less	-	63.7	(Kearney & Kruger, 2014)

[1] Mix refers to a mix of students, and non students;
[2] 'single email' = the subject received 1 phishing email; 'multi emails' = the subject received multiple (non)-phishing emails;
[3] Type of phishing email; general phishing or spear phishing;
[4] YoS refers to Years of Service for the organisation;
[5] Success shows the success of the phishing email in the control group;
[6] Classified at least 1 phishing email as legitimate;

enced by the years of service of the recipient.

H5) Since there is limited research regarding the cross-cultural influences in security research, this variable is included in the analysis. We therefore hypothesise that the success of a phishing email is influenced by the cultural background of the recipient.

4.2 Methods

The sample consisted of 593 subjects of both genders who worked in The Netherlands. All employees from one faculty were approached.

4.2.1 Subject selection

The pool of subjects consists of Professors (Full, Associate and Assistant), Postdoctoral researchers, PhD-candidates and support personnel. The sample consisted of 24.5% females and 75.5% males. The average age of the employees is 39.46 (SD = 12.20) years, ranging between 22 and 76, whereas females were younger (38.08 vs 39.92 year). Regarding YoS, the average is 5.72 (SD = 7.82) years, ranging between 0 and 42. Females had slightly more years of service compared to males (6.23 vs 5.80). Two third of the employees were Dutch ($N = 380$), whereas 196 (34.03%) originated from elsewhere. In this latter group, 77 employees were from Europe, 90 from Asia, 8 from Africa, 2 from North America, 16 from South America and 1 from Oceania. More details can be found in Table 4.2.

Table 4.2: Origin of subjects (380 Dutch subjects, PDI = 38, are excluded from this overview).

continent	countries	*N*	pdi range	avg[1]
Africa	5	8	[49, 80]	*67.88*
Asia	12	90	[54, 104]	*71.57*
Europe	12	77	[35, 93]	*48.23*
North America	2	5	[39, 40]	*39.75*
Oceania	1	1	[36]	*36.00*
South America	7	16	[63, 85]	*71.88*
Total	39	196	[35, 104]	*61.44*

[1] = Weighed average PDI;

4.2.2 Procedure

Before data collection, our research was approved by the Institutional Review Board (IRB) of the University. All employees were approached by email on a Monday evening in a regular term week. The subjects had until Thursday evening to respond to the email; the data collection stopped thereafter.

Two types of email were randomly assigned to the subjects. Those in the control group received an email with the opening 'Dear employee', whereas those in the experimental group received an email with the opening 'Dear `[name]`'. Each employee received the following email:

Scenario 4.1. Password Synchronisation

> Dear employee,
>
> Due to recent changes to the UT computer system, some complications emerged between our database servers. This system, which contains your username and password, is not correctly synchronised.
>
> Your data were not compromised, and the problems are already fixed. To avoid complications in the near future a complete synchronisation between the servers is scheduled. If your account is not correctly synchronised, you cannot login anymore.
>
> The password for your IT-account needs to be resynchronized before 29-10-2015. This can only be done via login.utwente.nl. Click on Sync password, and your password will be synchronised automatically. Cannot find Sync password? This means your password has already been synchronised. If you do not resynchronize your password with this link, you will no longer be able to use the central IT facilities.
>
> The IT-account is used for computer log on, email, WiFi, VPN connection and also logging on to various UT web applications.
>
> Synchronise your password within a period of 3 days.
>
> To synchronise your password, click here.
>
> Kind regards,
>
> Jort Welp
> Security Manager IT-Helpdesk.

Each employee was subjected to the same request. Following the data collection, employees received a debriefing statement explaining that the phishing email was part

of an awareness training.

Legitimate vs illegitimate email

The differences between the legitimate and the experimental email and website are discussed. A legitimate email from the ICTS department has the following characteristics:
i) Since there are 34% foreign employees in the organisation, important communication from the organisation is sent in two languages (i.e. Dutch and English).
ii) The sender of the email is always either the ICTS or Facility Management department. In case the ICTS department is the sender, the logos of both the university and the department are used.
iii) The signatory of the email is never a person, but instead the contact details of the ICTS department are provided, even if it was sent by the Facility Management staff.

The illegitimate email had three characteristics: *i*) Both the URLs in the email redirected to `http://login.utvvente.nl` rather than to `http://login.utwente.nl` (note the difference between in `utVVente` vs `utWente`). The website was hosted by a third party which did not relate to the organisation. For a photo of the website used in the experiment and its legitimate equivalent used in the organisation, refer to Figure 4.1 and 4.2 respectively.
ii) The signatory was a fictitious person, hence not a university employee.
iii) IT-Helpdesk was mentioned as the organisational unit in charge of IT rather than ICTS.

Figure 4.1: The fake website that was used in the experiment.

4.2.3 Variables

The variables used in the analysis were: compliance, spear, gender, age, Years of Service (YoS), nationality and Power Distance. The dependent variable *compliance* measured whether the subject complied with providing the requested Personally Identifi-

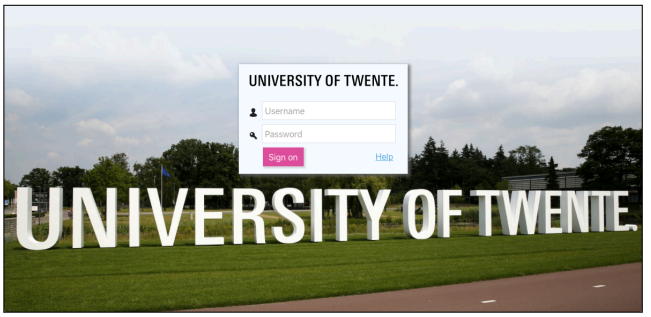


Figure 4.2: The legitimate website of the organisation.

able Information (PII) in a web form. The dichotomous variable was dummy coded as 0 = did not comply, 1 = did comply. The independent dichotomous variable *spear* measured the opening the phishing email used (0 = general, 1 = personalised). The independent dichotomous variable *gender* measured whether the subject was a female or a male and was dummy coded (0 = female, 1 = male). The independent continuous variable *age* measured the age of the subject (25 = 25 years old). The independent continuous variable *YoS* measured the seniority of an employee, operationalized as years of service (5 = 5 years of service). The independent categorical variable *nationality* measured the subject's country of origin. The independent continuous variable *PDI* measured the extent to which a society accepts that power is unequally distributed (low score indicates the tendency to distribute power equally, whereas high scores indicate a confirmation of the hierarchy) (Hofstede et al., 2010). The *PDI* score is based on the variable nationality, for each nationality the *PDI* values were retrieved from Hofstede et al. (2010), the scores range between 11 and 104 (38 = The Netherlands). The cultural scales are validated and correlate with the dimensions from the World Values Survey (P. B. Smith & Schwartz, 1997; Hofstede, 2001). Finally, it was tested whether the different type of phishing emails were randomly distributed among the four independent variables (i.e. gender, age, YoS and PDI), no statistical significant differences were found. All variables were stored in the `mail-se` dataset, refer to Table 1.2.

4.2.4 Analysis

The first hypothesis was tested using Cross Tabulation and Chi Square. The remaining hypotheses were tested using Logistic Regression. The following two data assumptions must be met for Chi Square analysis: *i*) independence and *ii*) minimum frequency of 5 observations per cell in the cross-tabulation (Field et al., 2012). Independence relates to putting a single observation in only one cell. In case one assumption

is not met, the Fisher's Exact test should be used instead. Both assumption were met since *i* there was categorical data used in the analysis and the number of observations exceeded the required minimum.

The following three assumptions must be met for logistic regression analysis: *i*) sufficient sample size, *ii*) no outliers and *iii*) no multicollinearity (Pallant, 2010). First, the dataset should contain at least 10 Events Per Variable (EPV), which is considered as a minimum required for running a logistic regression (Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996). The dataset contained 593 observations and was therefore considered sufficient. Second, for dichotomous variables, one value was placed in exactly one category. Third, the Variance Inflation Factor (VIF) (Variance Inflation Factor) is 1.25 which is below the cut-off value of 10, indicating that there was no evidence of multicollinearity (Pallant, 2010). For an overview of the VIF statistics refer to Table 4.3.

Table 4.3: Summary of VIF statistics.

VARIABLE	VIF	TOLERANCE	R^2
spear	1.00	0.997	0.003
gender	1.02	0.984	0.016
age	1.58	0.632	0.368
YoS	1.52	0.657	0.344
PDI	1.10	0.906	0.094
MEAN	1.25		

One characteristic of age is that it is often nonlinearly correlated with other variables. In this study a u-shaped relation was found with compliance, suggesting a nonlinear relation (refer to Figure 4.3). To overcome this, a straight line needed to be transformed into a curved line (i.e. adding a quadratic coefficient). Carrying out a nonlinear regression is as simple as transforming the independent variable and adding it (i.e. age^2) to the equation (Miles & Shevlin, 2001, p. 138). Note that *age* and age^2 are not functionally independent, but linearly independent (Greene, 2011). Addition of the squared term means that the two *age* coefficients cannot be interpreted separately (European Social Survey Education Netu, 2013). Furthermore, we tested whether the simultaneous influence of two variables on a third was non-additive. The relevance is that if two variables interact, the relationship between each of the interacting variables and compliance with the offender depends on the value of the other interacting variable. The YoS of an employee is restricted by their age, hence their correlation $[r = 0.587, p = .000]$, therefore their interaction was tested. Other interactions

involve the type of email and sociodemographic variables, where different groups of employees have different responses to the different types of email.

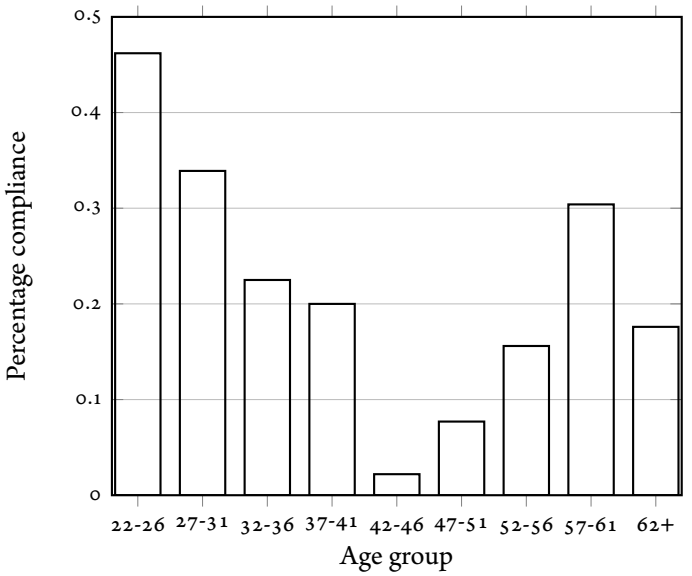


Figure 4.3: Percentage compliance per age group.

4.3 Results

4.3.1 H1) "The opening sentence of a phishing email influences its success."

Compliance of those who received a general phishing email was 19.3%, compared to 28.9% for those who received a spear phishing email ($\chi^2 = 7.368$, $df = 1$, $p = .007$). Hypothesis 1 is therefore accepted. Those in the spear phishing group had 1.693 times higher odds of compliance (i.e. providing their PII) than those exposed to a general phishing email ($OR = 1.693$, $CI\,[1.16, 2.48]$). Refer to Table 4.4 for descriptive statistics. After controlling for sociodemographic variables, those who received a spear phishing email still had higher odds of compliance ($OR = 2.418$, $CI\,[1.22, 4.79]$), refer to Table 4.5.

Compliance rates in previous research for general phishing emails ranged between 2.4% (Wright et al., 2014) and 92.5% (K. Hong et al., 2013), with a weighted average of 21.4%, refer to Table 4.1. In this study the compliance rate for general phishing was 19.3% and compared to previous studies, was considered as average.

Table 4.4: Number of observations and percentages per phishing condition.

		SPEAR PHISHING		
		NO	YES	TOTAL
COMPLIED	NO	238 (80.7%)	212 (71.1%)	450 (75.9%)
	YES	57 (19.3%)	86 (28.9%)	143 (24.1%)
TOTAL		295 (100%)	298 (100%)	593 (100%)

Group general = spear ($\chi^2 = 7.368$, $df = 1$, $p = .007$);

Since those who received a spear phishing email have higher odds of compliance, we tested whether the odds of compliance differed among people with different characteristics. In particular we tested whether the relation between sociodemographic variables and compliance was influenced by the type of phishing email (i.e. *spear*). The relevance is that if the type of email and the sociodemographic variables interact, the relationship between each of the interacting variables and compliance with the offender depends on the value of the other interacting variable. Therefore, for each hypothesis we also tested whether *spear* moderates that variable.

4.3.2 H2) "The success of a phishing email is influenced by the gender of the recipient."

No main effect of gender on compliance was found while controlling for sociodemographic variables ($OR = .825$, $CI\,[0.48, 1.43]$, $p = .492$). Refer to Table 4.5 for the full regression model. Hypothesis 2 is therefore rejected in favour of the alternative hypothesis H2a: "The success of a phishing email is not influenced by the gender of the recipient." The interaction effect between *spear* and *gender* was not statistically significant (*spearXgender*: $OR = 0.806$, $CI = [0.27, 2.42]$, $p = .700$).

4.3.3 H3) "The success of a phishing email is influenced by the age of the recipient."

No main effect of age on compliance was found while controlling for sociodemographic variables (*age*: $OR = 0.865$, $CI = [0.72, 1.04]$, $p = .126$ and age^2: $OR = 1.002$, $CI = [1.00, 1.00]$, $p = .117$). Hypothesis 3 is therefore rejected in favour of the alternative hypothesis H3a: "The success of a phishing email is not influenced by the age of the recipient."

A significant interaction effect was found between *spear* and *age* (interaction term *spearXage*: $OR = 0.925$, $CI = [0.88, 0.97]$, $p = .003$). These results suggest that younger

and older employees react differently to the two type of emails, whereas the variable *spearXage* indicates how different this is, refer to Figure 4.4. Compliance to spear phishing emails is higher compared to general phishing emails for younger emlpoyees. This difference decreases with age, approaches zero, and the age group 61+ employees is the most vulnerable to general phishing emails. The Odds Ratio (OR) tells us that as age increases by 1 year, in combination with a general email becoming a spear phishing email, the change in odds of providing PII is .925. In particular, *spearXage* is the difference in OR corresponding the change in type of phishing email (from general to spear) in two age homogeneous groups which differ by 1 year.

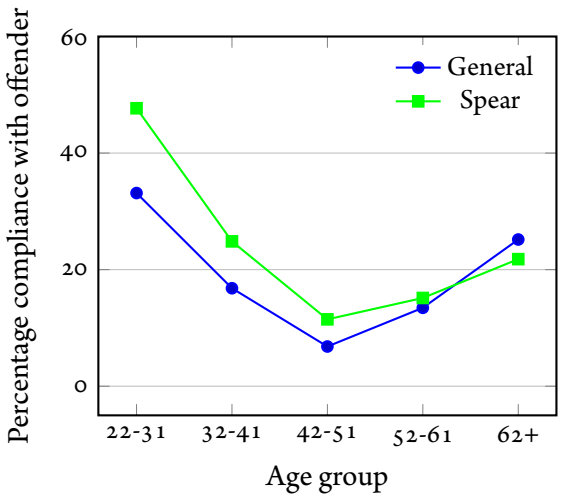


Figure 4.4: Compliance percentage for 5 age groups, for two types of phishing emails. The sample size for the age groups that received a general phishing email are 89, 52, 47, 37 and 11. The sample size for the age groups that received a spear phishing email are 97, 59, 51, 25 and 16.

4.3.4 H4) "The success of a phishing email is influenced by the years of service of the recipient."

The employees who worked longer for the organization were less vulnerable to phishing emails ($OR = 0.855$, $CI = [0.77, 0.94]$, $p = .002$). Hypothesis 4 is therefore accepted.

A significant interaction effect was found between *age* and YoS (*ageXyos*: $OR = 1.005$, $CI = [1.00, 1.01]$, $p = .041$). These results suggest that new employees and those with more YoS react differently to phishing emails for different ages. The OR tells us that as the age of an employee increases by 1 and the years of service increases,

the change in OR for complying compared to not complying is 1.005.

Furthermore, a significant interaction effect was found between the variables *spear* and YoS (*spearXyos*: $OR = 1.259$, $CI = [1.09, 1.46]$, $p = .002$). These results suggest that new employees and those with more YoS respond differently to the two types of phishing emails, refer to Figure 4.5. The OR tells us that as the type of email changes from a general to a spear email, in combination with an increase in YoS, the change in OR of compliance with the offender compared to non compliance is 1.259.

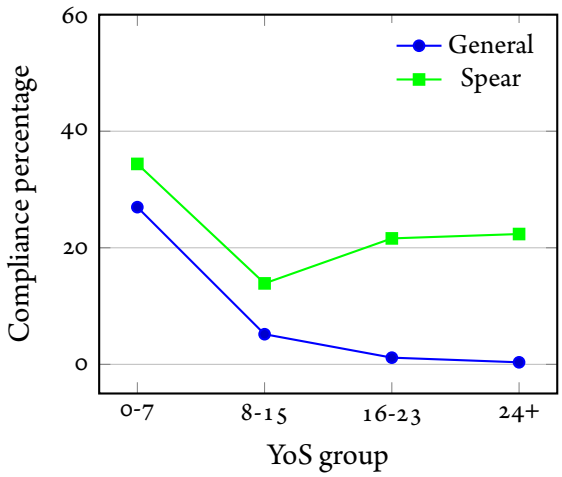


Figure 4.5: Compliance percentage for 4 YoS groups, by two types of phishing emails. The sample size for the age groups that received a general phishing email are 171, 29, 20 and 11. The sample size for the age groups that received a spear phishing email are 178, 42, 14 and 12.

The effect of having more years of service is different depending on the age and whether the employee received a general or a spear phishing email. For instance, for the youngest age group (age 22-31), in both the general and spear condition, compliance decreases when YoS increases. For those who received a general phishing email, compliance was 38.5% for those with less that 1 YoS compared to 1.18% for those with 8 YoS, whereas for those who received a spear phishing email compliance was 52.07% for those with less that 1 YoS and 23.19% for those with 8 years of service. This trend is comparable for the age group 32-41. For employees in the age group 42-51 the who received a general phishing email, compliance decreases when the YoS increases, whilst the compliance does not change for those who received a spear phishing email. For those in the age group 52-61 and 61+ who received a general phishing email, compliance decreases when the YoS increases, whereas the compliance increases for those who received a spear phishing email. For an overview of the predicted compliance

rates for *age* and *YoS* per type of phishing email, refer to Figure 4.6a for general phishing emails and 4.6b for spear phishing emails.

4.3.5 H5) "The success of a phishing email is influenced by the cultural background of the recipient."

Those with a higher PDI have a higher probability of filling out PII in a phishing email (OR = 1.025, CI = [1.01, 1.04], p = .000). Hypothesis 5 is therefore accepted. Furthermore, no interaction effect between *spear* and *PDI* was found (*spearXpdi*: OR = 1.014, CI = [0.99, 1.04], p = .326).

Table 4.5: Model comparison. The columns depict for each variable: the odds ratio (OR), its standard error (between parentheses), its lower and upper 95% confidence intervals [in brackets] and its significance level.

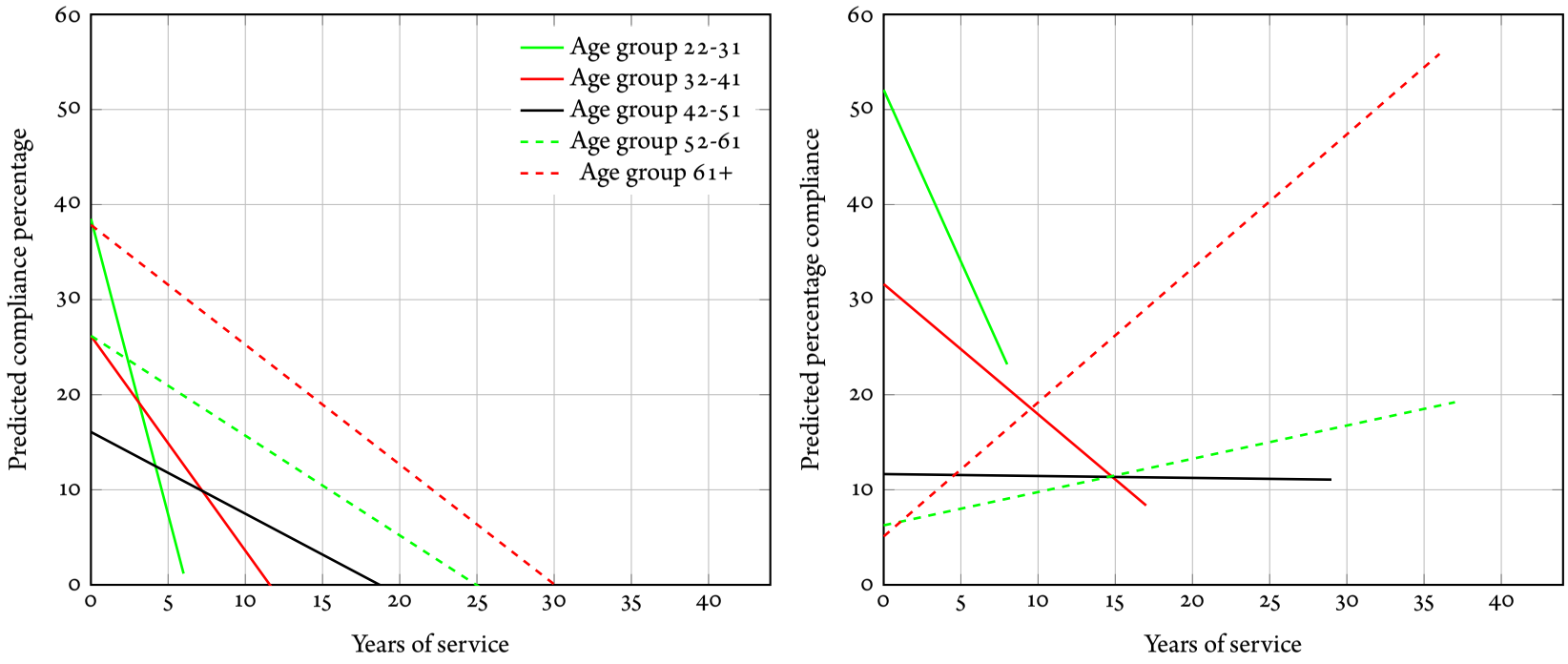
VARIABLE	MODEL 1: SPEAR	MODEL 2: FULL
Spear	1.693 (*.331*) [1.16, 2.48]**	2.418 (*0.843*) [1.22, 4.79]*
Gender		0.825 (*0.230*) [0.48, 1.43]
Age		0.865 (*0.082*) [0.72, 1.04]
Age²		1.002 (*0.001*) [1.00, 1.00]
YoS		0.855 (*0.043*) [0.77, 0.94]**
ageXyos		1.005 (*0.002*) [1.00, 1.01]*
PDI		1.025 (*0.007*) [1.01, 1.04]***
spearXgender		0.806 (*0.452*) [0.27, 2.42]
spearXage		0.925 (*0.024*) [0.88, 0.97]**
spearXyos		1.259 (*0.093*) [1.09, 1.46]**
spearXpdi		1.014 (*0.014*) [0.99, 1.04]
Consstant	0.239 (*.035*) [0.18, 0.32]***	1.879 (*3.775*) [0.04, 96.4]

* $p < .05$; ** $p < .01$; *** $p < .001$;
ageXyos = interaction of Age and Years of Service (YoS);
Model 1($\chi^2(1) = 7.41$, $p = .007$), N = 593, pseudo R^2 = .011;
Model 2($\chi^2(11) = 87.16$, $p = .000$), N = 462, pseudo R^2 = .167;
Model 1=2 ($p = .000$);

4.4 Discussion

This study investigated the susceptibility to two type of phishing emails and the personal characteristics that influence the probability of compliance.

Personalised phishing emails proved to be more successful than general phishing emails. Those who receive a spear phishing email have 1.7 times higher odds of compliance with the offender than those who receive a general phishing email. These results are in line with Sparshott (2014).



(a) Received a **general** email

(b) Received a **spear** email

Figure 4.6: Predicted compliance based on the full model (i.e. spear, gender age, age^2, YoS, ageXyos, pdi, spearXgender, spearXage, spearXyos and spearXpdi) for 5 age groups by type of email (i.e. spear).

Previous research operationalized spear phishing by supposedly using a friend of the receiver as the sender (Jagatic et al., 2007) or a combination of writing in the native language, using the name of the receiver, using the name of the organisation and the name the organisation's executive (Flores, Holm, Nohlberg, & Ekstedt, 2015). These have proven to be successful. Our results show that by only using the name of the recipient, a strong spear effect can be established.

Since spear phishing emails are more successful, it is suggested that the offender can generate similar benefits by sending fewer emails. Note that there is more effort required in constructing spear phishing emails. Furthermore, there is a potential risk in obtaining the personalized info. However, there is less exposure since the lower number of emails. In sum, there is a nett gain in effort-to-reward ratio for the offender in spear phishing. From a rational choice perspective, spear phishing is, therefore, a bigger threat than general phishing.

No main or interaction effect of gender was found when controlling for sociodemographic variables i.e. females and males are equally vulnerable. These results suggest that females and males are equally capable of identifying (spear) phishing emails. We therefore conclude that incorporating a gender-based approach into awareness campaigns or training is unnecessary. One explanation for this finding could be that males are considered to be more tech-savvy, whereas females are considered more risk averse. It was found in a quantitative study of 1 058 subjects that females had more negative attitudes and more anxiety towards ICT use compared to males (Broos, 2005). Furthermore, of 120 final year polytechnic students, it was observed that males had acquired more ICT skills compared to females (Murgor, 2013). Regarding females being more risk averse, experiments have shown that females make more risk averse choices than males, whereas this effect declines with age (Riley & Chow, 1992; Hersch, 1996; Barsky, Juster, Kimball, & Shapiro, 1997; Halek & Eisenhauer, 2001). Furthermore, people also tend to perceive and predict that females as more risk averse than males (Ball, Eckel, & Heracleous, 2010). Although it was not tested, it could be that these two effects cancel each out.

No main effect of age was found when controlling for sociodemographic variables. This result contradicts previous experimental findings (Sheng et al., 2010; Tembe et al., 2013) which found younger people to be more vulnerable. Although Figure 4.3 suggests there is an effect, this effect disappears once the quadratic term was included in the model. One explanation could be that the other studies did not check or mention whether the relation between age and victimisation was u-shaped and assumed a linear rather than a quadratic model. When we tested the relation between age and

compliance in a linear model, we also have found a strong age effect with younger people being more vulnerable. Moreover, although previous research found that younger people are more susceptible than older people, none of them included YoS as a variable in their analysis (J. Lee & Soberon-Ferrer, 1997; Titus et al., 1995; Sheng et al., 2010; Tembe et al., 2013). Perhaps YoS is a better predictor for victimisation than age. Another explanation could be that older people have more experience regarding fraud (Titus et al., 1995) and phishing, whereas younger people are born with this technology ubiquitously. Both age groups have an advantage which possibly complements one another.

The relation between age and compliance with the offender is moderated by the type of email that was received. A 'general' explanation for the moderation effect of email type could be that general phishing and spear phishing are two different types of crime and should be seen as such. This is comparable to the absence of a gender effect in fraud victimisation (Deevy et al., 2012), whereas zooming in on the specific types of fraud did find gender effects.

It was found that those who worked longer for an organisation were less likely to be victimised by a phishing email. The phishing email used in this study was related to the organisation. Our explanation is that those who worked longer for the organisation are more aware and familiar with the rules and procedures and therefore less likely to fall for an organisation-related phishing email. Furthermore, it is important to include the variable YoS in future research involving organisational penetration testing.

The relation between YoS and compliance was moderated by age. Since YoS and age are correlated, these two have a relation. In the context of an organisation, it makes sense that employees with more YoS are more familiar with the organisational procedures and customs, whereas those with less YoS do not have this experience. The results suggest that young employees with few YoS are the group that is most vulnerable to phishing emails. A (Human Resource Management (HRM) strategy that favours focus on short temporary employment contracts, therefore implies a security challenge! It is our view that this cost reducing approach not only increases management but also security prevention costs. This trend of focusing on temporary employment contracts also surfaces in the physical security branch of e.g. a national airport. In a report of the Dutch Federation of Trade Unions, 176 airport service agents were interviewed regarding their workload and safety culture of the airport (FNV, 2016). It was almost unanimously (99%) stated that it is important for safety and security to have experienced personnel, whereas 73% stated that uncertainty regarding employment negatively affects airport security. The conclusion was that the workload and the security risks are partially caused by high staff turnover and job uncertainty in

combination with inexperienced personnel (i.e. interns and temporary staff).

Regarding PDI, the results showed that those with high power distance cultural backgrounds are more vulnerable to phishing emails than those with a low score background. The rationale for this finding is that those with high PDI scores are more inclined to follow those higher in the hierarchy. The signature of the phishing email was from the Security Manager, someone clearly high in the hierarchy. Furthermore, the PDI scale describes "Subordinates expect to be told what to do." as a characteristic for those with a high score (Hofstede et al., 2010). The phishing email had a clear instruction on what to do; "To synchronise your password, click here".

4.4.1 Implication for practice

This research has the following implications for the practitioners: *i*) Focus awareness: recently hired personnel and employees with cultural backgrounds that have a high PDI are more vulnerable. Special attention should be paid to this group. *ii*) Content of an awareness campaign: spear phishing emails were more successful than general phishing emails. This knowledge should be included in the awareness training. *iii*) It is advisable to give employees an intervention at the start of their employment.

4.4.2 Limitations

This study has several limitations: *i*) The cultural dimension Power Distance was based on nationality. The variable reflects the country-based average rather than the individual's score. Future research could involve the inclusion of the subject's cultural perception (Hofstede, 1980; C. Lee, Pillutla, & Law, 2000; Yoo, Donthu, & Lenartowicz, 2011). *ii*) The dataset that was used in the analysis had missing values for some variables. It should be noted that the outcomes could be different if there were no missing values.

4.4.3 Future work

Finally, we present five recommendations for future research: *i*) Expansion of the experimental design to include an awareness campaign that counters the effect of phishing emails. An example of such a design is discussed in (Wright et al., 2014).*ii*) The usage of a heterogeneous sample for more generalizable results over a homogeneous academic sample. *iii*) The usage of persuasion principles (refer to (Cialdini, 2009) for varying experimental conditions. *iv*) Time decay effects of an intervention, as described in (Sutton et al., 2011). *v*) The level of seniority was included as years of service. An additional variable to include could be the level of education.

R2D2, you know better than to trust a strange computer.

C3PO – Star Wars

5

Telephone social engineering

This chapter is based on (Bullée, Montoya, Junger, & Hartel, 2016).

5.1 Introduction

Chapter 4 explained victimisation of email social engineering (i.e. phishing) by type of phishing email and 4 target demographic variables. Factors that influenced the success rate of a phishing email (either positive or negative) were type of phishing email, years of service and cultural background. This chapter describes a telephone social engineering experiment. In particular, the effect of an intervention over time will be investigated. This experiment was repeated 3 times with different convenience samples.

Since 2008 a scam has been carried out by employees claiming to belong to Microsoft's technical department (Arthur, 2010). A phone call is received unexpectedly at home; the caller introduces himself and proceeds to inform the home owner that there is a virus on the PC or that the PC is distributing spam emails. To verify the claim from the caller, the victim is persuaded to open a remote desktop session to review some warnings in the system log files. In order to resolve the problem, the caller advises the victim to buy a small software tool to prevent loosing valuable data. The solution can be bought on their website and payment is possible via either credit card or PayPal. When the victim next checks the bank account, he/she discovers that the savings have disappeared. The victim thinks that he/she is securing the system whilst in fact the opposite has taken place.

Since 2011, the Dutch Fraud Help Desk (an organisation that collects fraud data) has received 6 000 reports of this Microsoft technical support scam in The Netherlands. In 2014, there were 856 people who filed a complaint, and 88 (10.28%) of them admitted to have paid the scammers. In 2015 there were 1 629 complaints filed, of which 218 (13.4%) involved payment. The total reported damage in 2015 was €113 270. In 2016 there were 1 536 incident reports, of which 129 (12.6%) involved payment. The total reported damage in 2016 was €220 398. This constitutes an increase in the the reported damage and damage per person. These numbers indicate that people are vulnerable to social engineering via the telephone.

Information security has been treated as a technical problem for many years, resulting mainly in technical solutions (Rouse, 2006) and overlooking the human aspect (Rhee et al., 2009). Since information technology becomes more integrated into our daily activities, security experts propose that social engineering will be the greatest threat to any security system (Rouse, 2006). One example of social engineering is the technical support scam (Harley, Grooten, Burn, & Johnston, 2012).

This chapter therefore explores: *i*) the extent to which people are susceptible to telephone-based social engineering attacks when they are persuaded to go to a website

and download a software and, *ii*) the effectiveness of an intervention to reduce the effects of social engineering over time.

5.1.1 Informing people to change behaviour

The Elaboration Likelihood Model (ELM) of Persuasion describes how information is processed and can be tailored to the receivers (Petty & Cacioppo, 1986). For a discussion on how the ELM can be used in information leaflets and gadgets to change behaviour, refer to Chapter 3.1.2.

5.1.2 Retention

In 1885 Herman Ebbinghaus demonstrated the existence of memory decay over time (Ebbinghaus, 1913). In addition, he described the 'forgetting curves', which refer to the amount of new information one is able to retain with each repetition of the same content. The amount of new information learned after each repetition decreases less steeply, meaning that more and more information stays in one's memory (Ebbinghaus, 1913). Retention can relate to knowledge but it can also relate to skill and is not limited to a single context as shown in the examples below.

Knowledge

Ebbinghaus showed, using a list of 3 letter nonsense syllables (starting and ending with a consonant and a resonant in the middle), that over time, more syllables are forgotten (Ebbinghaus, 1913). Since the publication of this work, other research areas have looked into this topic as well, such as in the fields of aeronautics and medicine.

The aeronautical knowledge of 60 pilots was tested with a multiple-choice test, with items randomly selected from the Federal Aviation Administration (FAA) private pilot item bank of questions. A negative correlation was found between test scores and number of months since each pilot's last flight review (r = -.44, df = 18, t = 1.96, $p < .05$). This means that pilots who recently completed their flight reviews were associated with higher scores, compared to those with a longer time since the review (Casner, Heraldez, & Jones, 2006).

In the field of medicine it was shown that knowledge regarding Cardiopulmonary Resuscitation (CPR) dropped significantly over a period of 10 weeks. In this test 19 health care professionals (i.e. nurses) were tested via a 26 open item CPR theoretical questionnaire (Broomfield, 1996).

Skill

The flight manoeuvres of a group of 192 pilots were tested in a full-motion flight simulator 6 and 12 months after training. During the test the pilots had to perform 12 emergency manoeuvres and 25 normal manoeuvres. The results showed there was a significant decay in performance between the 6 and 12 month group (Hendrickson, Goldsmith, & Johnson, 2006).

18 medical students were assessed on their CPR performance at three points in time: a) a pre-test to establish the baseline score, b) a post-test after a training and c) a re-test after 10 weeks. The test scores between the post- and re-test differed significantly, confirming skill decay (Madden, 2006).

In the context of a phishing test, the skill retention of adult subjects was measured 7 days after an information campaign (Kumaraguru et al., 2010). During the test, the subjects had to classify 6 emails (as phishing emails or not) at 3 points in time (i.e. pre-test, post-test and re-test). This study did not find a significant decay of performance in classifying phishing emails.

This suggests that skill after training remains stable on the short term, but not on the long term. A further question to be answered is: how do the effects of training decay between 1 and 10 weeks?

5.1.3 Research question

The objective of this research was to find out: "*How susceptible are people to a technical support scam?*" Two hypotheses were formulated:
H1) Previous research (refer to Chapter 3.1.4) showed that there is a discrepancy between verbal and behavioural compliance. Therefore we hypothesise that there will be a difference between the self estimated and experimental observed compliance rate.
H2) Previous research showed that knowledge decays over time. Therefore we hypothesise that time influences the relation between compliance and intervention.

5.2 Method

The sample consisted of 92 subjects (who participated in a field study) and 31 subjects (who filled in a survey) of both sexes who worked in one particular building on the campus of the University of Twente. All subjects were scientific personnel. Only approached were those who *i*) had an office work space and *ii*) were present to answer the telephone.

5.2.1 Subject selection

Professors, secretaries, support and laboratory staff were excluded from the experiment in order to minimize disruption of primary activities. The pool of subjects therefore consisted of PhD-candidates, Post-Doc researchers as well as Assistant and Associate professors. The sample consisted of 34.29% of all possible targets; the nationality distribution was comparable to the overall nationality distribution of the faculty, while those in the experimental sample were slightly younger (39 vs 42 years).

5.2.2 Researchers

The researchers (i.e. the "offenders") consisted of 5 bachelor students (3 female and 2 male). The age of the researchers ranged between 21 and 24, the average age was *M* 22.4 years (SD = 1.14). The researchers were 80% (i.e. 4 out of 5) Dutch nationals and 20% (i.e. 1 out of 5) were German nationals, all subjects were profound in speaking the Dutch language. The researcher who administered the surveys was a 19-year-old Dutch male bachelor student. There was no restriction with regards to approaching subjects of the same sex.

5.2.3 Field study Procedure

The Institutional Review Board (IRB) of the University approved the study before collecting data. One-third of the potential subjects was exposed to an information campaign two weeks before the experiment whilst another one-third was exposed one week before. The research departments were randomly selected and all their staff was exposed to the intervention.

The intervention consisted of two parts: *i*) a leaflet informing staff about what constitutes a scam and describing how scammers operate, how to detect them and what to do (refer to Figure 5.1) and *ii*) a reminder in the form of a semi-transparent card holder with the university logo on one side and the text "Beware of scams. Verify all requests. Report all incidents." on the other side.

The leaflets were designed using story telling to ensure that non experts could understand it as well (Rader, Wash, & Brooks, 2012). The leaflet represents the information medium and the card holder represents a cue to remember the message. Departmental secretaries were responsible for distributing the materials and they were unaware that this was part of an experiment. The leaflet was distributed via email, whereas the card holder was distributed in person. All subjects were approached via telephone between 9.30 p.m. and 5 a.m., on a 'normal' Monday during term time.

Figure 5.1: Information leaflet.

The researchers were randomly assigned to a target, however if the researcher recognised a target, this target person was randomly assigned to another researcher. Each researcher approached the subject using the script described in Scenario 5.1.

Scenario 5.1. Download and install software

> Hi this is [name]. We discovered that the PC you are using is distributing spam emails. This is caused by a malicious program that is running in the background. Did you notice that your PC was a bit slower lately? There is nothing to be ashamed of, there are other people who have the same problem. I already helped 3 people to fix this earlier this morning. Luckily this is easy and quick to solve. Do you have 2 or 3 minutes time, so that we can remove it together right away? Please click the link that appears in the chat window. URL: `http://removespam.utwente.info`. To proceed to the download, please enter the validation code; this is your complete employee number. The complete number can be found on the back of your employee card. Please save the file to your Desktop and execute it. After the program is finished, could you read out the completion code?

All targets were subjected to the same script and request. After the target indicated that the downloaded file was installed, the debriefing procedure was started. During

the debriefing procedure the target was told that this was an experiment, and asked some demographic information, employment length, some computer characteristics and their reasons for or against downloading and installing the software. Finally, the importance of not sharing any information about the experiment with colleagues was explained; all subjects acknowledged this and agreed not to disclose any information. This was checked during the debriefing and none of the subjects stated having had prior knowledge of the experiment.

5.2.4 Survey procedure

All subjects were individually approached by a researcher between 10 a.m. and 6 p.m. on a 'normal' Thursday during term time. The researcher approached the subject and asked if they wanted to participate in a 10 minute survey regarding social engineering. Upon agreement, the subject received a study information sheet explaining the nature of the study and an informed consent form to sign agreement of participation. The survey is based on Scenario 5.1 and asked the subject their behavioural intentions how they would react when in that particular situation. Furthermore, the subject was asked some demographic information and length of employment.

5.2.5 Variables

The variables used in the analysis were: compliance, intervention, age, offender, sex and Years of Service (YoS). The dependent variable *compliance* measured whether the subject complied with the request of the offender to download and install the software package. The dichotomous variable was dummy coded as 0 = did not comply, 1 = did comply. For those who filled in the survey the variable compliance measured the verbal compliance with the request (0 = would not comply, 1 = would comply). The independent variable *intervention* measured whether the subject was exposed to the intervention (0 = not exposed to the intervention, 1 = exposed to the intervention 1 week before, 2 = exposed to the intervention 2 weeks before). The independent continuous variable *age* measured the age of the subject (25 = 25 years old). The independent dichotomous variable *sex* was measured for both the subjects and the researchers and was dummy coded (0 = female, 1 = male). The independent continuous variable *YoS* measured the seniority of an employee, operationalized as years of service (5 = 5 years of service). All variables regarding the field study were stored in the `phone-se` dataset, whereas the variables regarding the survey (i.e. compliance, age and YoS) were stored in the SURVEY dataset, refer to Table 1.2.

5.2.6 Analysis

Both hypotheses were tested using Cross Tabulations and a Chi Square test.

5.3 Results

A total of 92 subjects were contacted. No 'building' effect on compliance (with the offender) ($\chi^2 = 1.557$, $df = 1$, $p = .212$), 'offender sex' effect on compliance ($\chi^2 = .133$, $df = 1$, $p = .715$), 'age' effect on compliance ($OR = .98$, $p = .275$) and 'YoS' effect on compliance ($OR = .97$, $p = .222$) was found and therefore these are not further mentioned. A marginal effect of 'target sex' on compliance ($\chi^2 = 3.381$, $df = 1$, $p = .066$) was found, meaning that females have the tendency to comply more often compared to males.

5.3.1 Manipulation check

The debriefing procedure was used to verify that the subjects were coded correctly (i.e. had received an intervention or had not received an intervention). The subjects who recalled receiving the intervention material were coded as intervention group whilst those who could not recall having received any intervention material were coded as control group.

5.3.2 *H1*: "The compliance in the survey is unequal to the compliance in the experiment."

Those who were not exposed to an intervention and participated in the field study, 40.0% complied with the request of the offender compared to 0.0% of those who filled in the survey ($\chi^2 = 17.911$, $df = 1$, $p = .000$). Hypothesis *H1* is therefore accepted. Refer to Table 5.1 for descriptive statistics.

Table 5.1: Number of observations and percentages for each type of test.

		test		
		survey	field	total
complied	no	49 (100%)	21 (60.0%)	70 (83.3%)
	yes	0 (0.0%)	14 (40.0%)	14 (16.7%)
total		49 (100%)	35 (100%)	84 (100%)

Group control = intervention ($\chi^2 = 23.52$, $df = 1$, $p = .000$);

5.3.3 *H2:* "Time influences the relation between compliance and intervention."

The compliance for the control group was 40% compared to 17.2% when measured 1 week after exposure to the campaign. The compliance of those exposed to the campaign 1 week prior was 17.2% compared to 42.9% when measured 2 weeks after the exposure to the campaign. The compliance for the control group was 40% compared to 42.9% when measured 2 weeks after the exposure to the campaign. A difference was found between the control and the 1 week group, furthermore a difference was found between the 1 week and 2 week group. However no difference was found between the control and the 2 week group. Hypothesis *H2* is therefore accepted. Refer to Table 5.2 and Figure 5.2 for descriptive statistics.

Table 5.2: Number of observations and percentages per intervention condition over time.

		INTERVENTION			
		NO	1 WEEK	2 WEEKS	TOTAL
COMPLIED	NO	21 (60.0%)	24 (82.8%)	16 (57.1%)	61 (60.4%)
	YES	14 (40.0%)	5 (17.2%)	12 (42.9%)	31 (33.7%)
TOTAL		35 (100%)	29 (100%)	28 (100%)	92 (100%)

Group control = 1 week ($\chi^2 = 3.935$, $df = 1$, $p = .047$);
Group control = 2 week ($\chi^2 = 0.052$, $df = 1$, $p = .819$);
Group 1 week = 2 week ($\chi^2 = 4.466$, $df = 1$, $p = .035$);

5.4 Conclusion

This study investigated whether an information campaign influences the compliance with a telephone request to download and install software available on the internet, over time.

An information campaign consisting of *i*) informing employees about the dangers of telephone scams and *ii*) distributing a card holder with a reminder text was effective in the short term to reduce the vulnerabilty of employees to follow a stranger's request to perform actions on their PC.

There was a difference between the rate of verbal compliance with the offender's request (0.0%) and the behavioural compliance rate (40.0%). The results are similar to those in Chapter 3, those who filled in the survey underestimated the social engi-

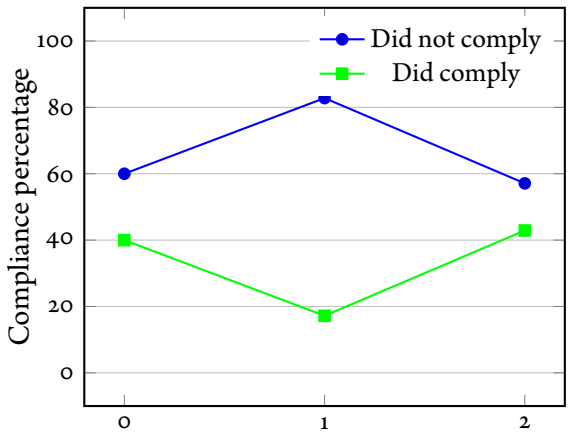


Figure 5.2: Change in compliance with the offender over time.

neering threat too. Therefore a similar reasoning applies regarding the distribution of countermeasures. People believe that positive events are more likely to occur to them than to other people (Weinstein, 1980). Therefore, they will underestimate the importance of the countermeasure.

In total 17.2% of the employees exposed to an information campaign (1 week prior to the attack) compared to 40% in the control group complied with the request to download and execute software available on the internet. Those not exposed to the campaign (1 week prior) have 3.2 higher odds of complying with the offenders request. These findings are in line with the results in Chapter 3 in which university personnel was approached by strangers and asked to hand over their keys.

The employees exposed 2 weeks prior showed no difference to the control group. The CPR studies of Madden (2006) and Broomfield (1996) both showed a significant decay since the training, however they measured their decay over a period of 10 weeks time.

One explanation for the difference could be the modality of information that is transferred, (i.e. visual leaflet). The effect of different modalities on memory has been shown in an experiment where the subjects had to remember and recall a list of words or auditory representations. The results showed that the auditory representations had both a significant better *i*) recall and *ii*) recall order of the presented stimuli (Drewnowski & Murdock, 1980). Glenberg showed that this auditory modality effect is also present in long term memory, this means that auditory stimuli were better remembered in the long term compared to their visual counterparts (Glenberg, 1984).

An alternative explanation could be in the process of creating a memory. There are 3 processes that constitute memory processing: *i*) Encoding, *ii*) Consolidation and *iii*) Retrieval. Encoding is the process of forming mental representations of the materials one wants to put in memory. The process of encoding is enhanced by elaboration (interpretation of the materials and connecting them to other materials) and trying to repeat it to oneself (Craik & Lockhart, 1972). Attention is important as well; when attention is divided as encoding will be weaker and later attempts to remember are likely to fail (E. Smith & Kosslyn, 2008, p. 202). The consolidation process modifies the mental representation in such a way that it becomes stable in memory. Consolidation of the declarative memory (explicit memory containing experiences and information) can be affected by sleep (Prehn-Kristensen et al., 2011; Ashworth, Hill, Karmiloff-Smith, & Dimitriou, 2014), caffeine intake (Favila & Kuhl, 2014) and age (Cahill, Prins, Weber, & McGaugh, 1994). Finally, retrieval is important to access the knowledge in the brain via cues. The context (e.g. physical surrounding) in which a memory is created is important for retrieval. In the context of the information campaign, a possibility is that the subjects could not identify themselves with the material because they thought that it was not applicable to them, since there are others who were more vulnerable (this line of thought can be explained by the Optimism Bias, refer to Chapter 1.1.3 for more details). In both cases the subjects will not have the appropriate attention to properly encode the material to make it last in the memory.

A third explanation could be that the information campaign is too abstract and therefore an ambiguous memory cue is created for the material. Once the subject is 'attacked' there is no proper recollection of the cue to memory and he/she fails to recollect the materials in the information campaign. This could explain the difference between the 1 and 2 week groups as the cues are simply forgotten over time. It would be too simple to say that the subjects forgot about the information campaign since there were questions to control for this, since the subjects stated they recalled having received and read the materials.

5.4.1 Limitations

This study has 2 limitations: *i*) The current study has a limited number of observations, therefore only a limited number of variables could be tested. *ii*) All subjects were from the same organisation, replication in other organisations would be needed.

5.4.2 Other work

The experiment described in this chapter was repeated 4 times in different organisations and departments. In these four samples only the control group was included. For an overview of the subjects, their origin, the sample size, compliance rate refer to Table 5.3.

Table 5.3: Overview of alternative populations that were tested using Scenario 5.1.

COUNTRY	ORG.[1]	SUBJECT TYPE	*N*	DID COMPLY	MODALITY[4]
NL	UT	Support - IT	8	1 (12.5%)	Mobile phone
NL	UT	Support - Finance	12	4 (33.3%)	Mobile phone
NL	TUD	Scientific - EWI[2]	15	5 (33.3%)	Mobile phone
SGP	SUTD	Scientific - Mix[3]	15	5 (33.3%)	Desk phone

[1] Organisation: SUTD = Singapore University of Technology and Design; TUD = Delft University of Technology, UT = University of Twente;
[2] EWI = Electrical Engineering, Mathematics and Computer Science;
[3] Mix = a mixture of technical and non technical backgrounds;
[4] MODALITY = the type of phone used by the offender;

5.4.3 Future work

In the current research the subjects in the experimental groups received an information campaign and a gadget. Although the gadget worked well as an intervention, it represents a physical item while the attack happens in the cyber sphere. A suggestion for future research would be to devise a "cyber intervention". A key benefit of this would be that distribution would be easier compared to that of physical items. The subjects in the current study all originate form the same organisation, faculty and building. However, the outcomes could differ among organisations, faculties and buildings. The suggestion for future research would be to approach subjects from different organisations, faculties and buildings. Finally, the results showed that there is an decay of knowledge over time. A useful follow up study could involve investigating if this decay effect can be countered. We expect that 'quick' booster interventions which are repeated on a regular basis would be effective, as in the CPR study (Sutton et al., 2011). The results showed that people are vulnerable to social engineering attacks and that an information campaign is effective in the short term. Organizations need to acknowledge the problem, put the issue on the agenda, conduct or commission research into it and find ways to reduce the threat. In this case, for example, the organization could issue a call-back policy for unknown incoming phone calls.

May the odds be ever in your favor!

Suzanne Collins – The Hunger Games

6

Social engineering in attack trees

This chapter is based on (Bullée, Montoya, Pieters, Junger, & Hartel, 2015b).

6.1 Introduction

Social engineering experiments using three modalities (i.e Face-to-Face (F2F), email and phone) were described in Chapters 3, 4 and 5 respectively. Factors that influence the success rate of the attack were described, refer to Table 1.1. This chapter describes a methodology that allows to include the results from social engineering experiments as input in a quantitative risk assessment using attack trees. In information security, attack trees are used to assess security. To do a quantitative analysis, values need to be assigned to the leaf nodes of the tree. Typically, these values are based on estimations using a 3-point scale (e.g. low-medium-high) (Bagnato et al., 2012). Although this is an elegant method, it is prone to cognitive biases, e.g. discussed in Chapter 1.1.3. The methodology discussed in this chapter allows to include data based on experiments to overcome the limitations.

The complexity of attacks on critical systems increases with the complexity of the systems themselves (Kordy, Mauw, Radomirović, & Schweitzer, 2011). To evaluate the safety of systems, fault trees were first developed in the 1960s (Vesely, Goldberg, Roberts, & Haasl, 1981). Attack trees were popularized by Bruce Schneier (Schneier, 1999) and constitute similar tree structures which have been used since the 1990s to assess security. The root node of an attack tree depicts the goal of the attacker (e.g. Obtain Exam). The children of a node in the tree are refinements of the node's goal into sub-goals. The leaves of the tree represent the basic actions to be executed by the attacker. Relations between siblings can either be: *i*) AND-relations for which all sub-goals have to be executed to satisfy the parent node; or *ii*) OR-relations for which any of the sub-goals has to be executed to satisfy the parent node (refer to Figure 6.1 for an example of an attack tree).

The quantification of attacks is a key component in security risk evaluation and mitigation. The leaf nodes of attack trees can be annotated with quantitative information. Values such as probability of success, costs or frequency of occurrence can be estimated for each leaf node and propagated up to the root node. However, the mathematical operations corresponding to the AND and OR relations differ. In the case of an AND gate, the probability of success of an attack is represented by the product of the associated leaf nodes whilst for OR gates the MAX-function (i.e. the largest of the given numbers, e.g. MAX(2, 3) gives 3) is applied to the leaf nodes. In fact, other propagation rules exist as well. For the purpose of this chapter, there is no strict need to use a particular set of propagation rules.

The annotation of leaf nodes is usually done using expert knowledge. Typical an-

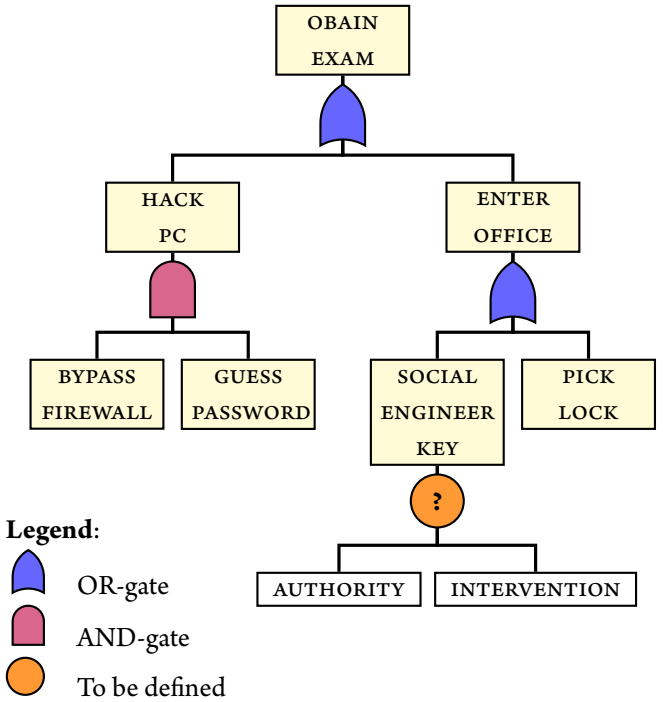


Figure 6.1: An attack tree for the node 'OBTAIN EXAM'. This goal can be achieved either by hacking into the PC *or* to physically enter the office. To hack into the PC, one must bypass the firewall *and* guess the password. To enter the office, either social engineer the key *or* pick the lock.

notations used are dichotomies (e.g. yes-no) or ordered categories such as 3 or 5-point, or Likert scales (e.g. low-medium-high). Although such methods suit experts because of their elegance, they have some drawbacks. In the case of ordered categories, reliability can be hampered due to biases such as the optimism bias (Weinstein, 1980), anchoring bias (Tversky & Kahneman, 1974) or the overconfidence effect (Pallier et al., 2002). Another factor to be considered is that the parent-and-child relation being modelled is context-dependent. This situation is typical of data from the social sciences involving human behaviour, such as that of social engineering experiments, in which the probability of success depends on the attacker's changes to the context by e.g. applying persuasion principles (Cialdini, 2009), which aim to maximize the probability of success.

To illustrate the modelling difficulties, a scenario consisting of an attacker who wants to steal an exam is used (refer to Figure 6.1). In the scenario we assume that the exam is in the office of the lecturer; physically in a printed form and digitally on the PC. The office is on the second floor of the building and can only be accessed by

the lecturer who has the key. Furthermore, the PC is connected to the internet and is protected by a password that is only known to the lecturer. Finally, when the lecturer is not in the office, the office is locked. The attacker can obtain the exam by either: *i*) hacking into the PC of the lecturer OR *ii*) obtaining a physical copy from the office of the lecturer.

To hack into the PC, the attacker needs: *i*) to bypass the firewall AND *ii*) then guess the password of the lecturer. The key, on the other hand, can be obtained by: *i*) manipulating the lecturer (social engineering) in order to obtain access to the office OR by *ii*) picking the lock of the lecturer's office. The success regarding the manipulation of the lecturer by means of social engineering can be influenced by: *i*) the attacker using authority; OR *ii*) the target having received a preventive intervention; OR *iii*) by using both authority and intervention; OR *iv*) with neither the use of authority nor intervention.

The example describes both properties in control of the attacker (e.g. Bypass Firewall, Authority and Pick Lock) and properties that are in control of the target (i.e. Intervention). Therefore, the example can be interpreted as an Attack Defence Tree (ADTree) (Kordy et al., 2011). ADTrees and traditional attack trees have a comparable structure. However, there is a difference in the children; independent of the AND and OR-gates in ADTrees. Actions in ADTrees can have refinements (children of the same type) and countermeasures (children of a different type). For each refinement, there can be one sibling of a different type that counteracts the parent. In the ADTree formalism, Authority would be modelled as an attack node, whereas Intervention as a defence node, refer to Figure 6.2.

The challenge is to model the node which contains human interaction (i.e. the 'SOCIAL ENGINEER KEY' node). To manipulate the context of the attack, the attacker may use his knowledge of one or several of Cialdini's persuasion principles: authority, commitment, liking, conformity, reciprocity and scarcity (for a detailed discussion, refer to Chapter 1.1.2). The 'authority' principle describes the likelihood of obeying requests from authoritative figures (e.g. requests made by a boss or person with a well-defined and known task). On the other hand, the potential targets can make changes to the context as well. They can be educated to protect themselves against social engineering attacks. The 'intervention', refers to an awareness campaign that helped preventing targets against social engineering attacks (e.g. describing how such an attack looks like and how to prevent becoming a victim). Such manipulations of context constitute attack steps themselves. However, this does not represent the simple type of relation which is modelled in traditional attack trees using AND or OR relations.

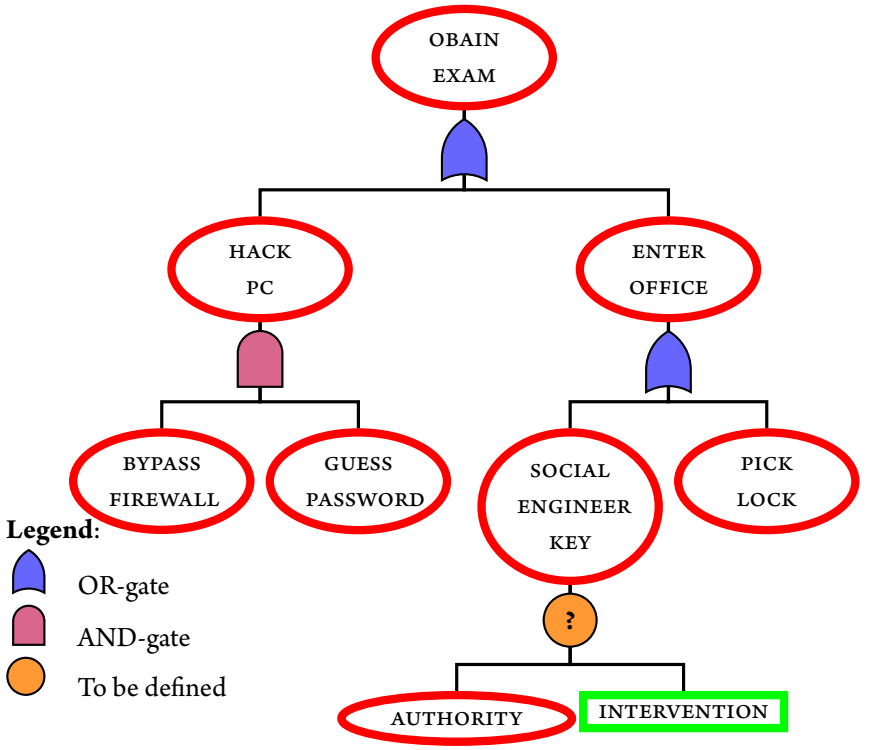


Figure 6.2: An ADTree for the node 'OBTAIN EXAM'. The ellipses with a red line resemble attack nodes, whereas the rectangle with a green line resembles a defence node.

In existing attack tree formalisms, a refinement of the node 'SOCIAL ENGINEER KEY' could have, for example, the children: 'deploy intervention', 'get aware of the risk', 'go to office', 'increase authority' and 'request key'. In the ADTree formalism, the children 'go to office', 'increase authority' and 'request key' would be modelled as a refinement, whereas the children 'get aware of the risk' and 'deploy intervention' would be modelled as counter measures. However, the node 'SOCIAL ENGINEER KEY' would have to be designed as either an AND or an OR node. In the former case, all 'children' would have to be executed in order to satisfy the goal of the 'parent' node (i.e. SOCIAL ENGINEER KEY), whilst in the latter case, the attacker would need to execute only one of the 'children'.

Such AND or OR form of reasoning does not apply to social engineering since it would be possible to succeed only by requesting the key (i.e. applying none of the 'children') but also by requesting the key and executing one or more of the 'children'. The particular nature of social engineering data (as opposed to technical data) implies that all 'children' correlate with the 'parent' to some extent; therefore, modifying any of the 'child' nodes affects the outcome (i.e. probability of success). Therefore, in an

attack tree's social engineering node, combinations of 'child' nodes should yield various probabilities of success. This also means that it is possible to have a probability of compliance with the request when none of the context variables are used. A different approach is therefore needed to incorporate social engineering human behaviour into attack trees. We therefore propose a 'regression node' to model the 'parent' on the basis of the 'children' parameters based on correlation coefficients as a method for incorporating social engineering human behaviour nodes in attack trees.

Benefits of this method are: *i*) the capability to incorporate context variables that influence probabilities, *ii*) data from experiments can be used to annotate the tree and *iii*) the context variables are incorporated in an compact way.

This chapter is structured as follows. First, a brief overview of the proposed attack trees extension(s) is given in Section 6.2. Section 6.3 describes the proposed regression node for attack trees. Furthermore, an example to illustrate the use of the proposed regression node is presented in Section 6.3.1. Finally, conclusions are drawn and suggestions for future research are made in Section 6.4.

6.2 Related work

Scholars already have made progress in the extension of attack trees. For example, multiple gates, relations and nodes have been proposed for various specific problems (Kordy, Piètre-Cambacédès, & Schweitzer, 2014). The extensions are separated in two groups: extensions that could model our scenario, and extensions that are interesting for future research, in particular defence nodes. For each extension method a short description is given.

6.2.1 Possible alternatives

Seven alternatives are discussed: *i*) Ordered AND-gate satisfies the parent node if all children are executed according to a given order; the children can be both leaf and non-leaf nodes. The Priority AND (Brooke & Paige, 2003; Khand, 2009) and the Sequence AND (Bistarelli, Fioravanti, & Peretti, 2006; Lv & Li, 2011) are similar types of gates and can be used in Intrusion Detection Systems (IDS) to detect attacks that are in the execution phase but that have not yet been completed (Camtepe & Yener, 2007).
ii) A Conditional Subordination (CSUB) gate is as extension of an AND-gate (Khand, 2009). It acts as an AND-gate, with an additional side input that is prioritized over the children. The parent node is satisfied if: *a*) all children are executed or *b*) an additional side leaf which has higher priority is executed.

iii) The Time Based Order Connector satisfies the parent node when the child nodes are executed within a predefined time frame (Wang, Whitley, Phan, & Parish, 2011). The number of child nodes executed is at least one and the order of execution is of no importance.
iv) The Inhibit gate is a special case of an AND-gate within the Fault Trees for Security formalism (Brooke & Paige, 2003). The parent's goal is satisfied if: *a*) all the child nodes are executed; and *b*) a predefined condition is met. The condition has to be of an environmental nature, such as temperature.
v) XOR gate, which originates in the Fault Trees for Security, indicates that exactly one of the children must be executed to satisfy the parent's goal (Brooke & Paige, 2003). This definition differs from some of the definitions of the 'standard' OR gate (i.e. at least one of the children must be executed). In the literature, these three definitions are used interchangeably. OR gates are defined as 'any child' by (Schneier, 1999; Bistarelli et al., 2006; Zonouz, Khurana, Sanders, & Yardley, 2009), 'only one child' by (Mauw & Oostdijk, 2006; Lv & Li, 2011; Yager, 2006; Roy, Kim, & Trivedi, 2010) and 'at least 1 child' by (Kordy et al., 2011; Brooke & Paige, 2003; Camtepe & Yener, 2007; Fovino, Masera, & Cian, 2009). It should be noted that the difference in definition affects the propagation rules.
vi) OWA operators are part of OWA trees (Yager, 2006). The aim of OWA is to handle fuzzy sets of executed child nodes in order to satisfy the parent's goal. This implies situations that lie between 'all children' and 'one child', such as 'most of the children' or 'at least half of the children' must be executed to satisfy the parent's goal. This means that OWA allows the modelling of situations with probabilistic uncertainty based on the number of children that must be executed to satisfy the parent's goal.
vii) In a *k*-out-of-*n* gate, the parent's goal is satisfied if a pre-defined subset (k) of all children (n) is executed, whereas the order is not important (Khand, 2009; Roy et al., 2010). A common way to present this is as a $\frac{k}{n}$-gate, where $k \geq 1$. In the case of $k = 1$, the function is the same as an OR-gate. A similar gate is the Threshold Based Connector, where every combination of exactly k children out of n is possible (Wang et al., 2011).

None of the extensions were able to model satisfying the parent node when none of the children are executed. Extensions *i*, *ii*, *iii* and *iv* assume that all children are executed in order to satisfy the parent, whereas extensions *v*, *vi* and *vii* assume a minimum of 1 child is executed in order to satisfy the parent node. It is therefore not possible to model all the 'possible' changes in context of the 'SOCIAL ENGINEER KEY' attack step. Our approach differs because attack trees get enriched by placing basic attack actions (leaf nodes) in a particular context. The context of the attack can change. The focus

is not on weighting actions *per se*, but rather on how attacks change by actions one is able to control.

6.2.2 Defence nodes

Next to possible alternatives for modelling social engineering in attack trees, defence nodes may be combined with social engineering aspects to model the effect of interventions such as awareness campaigns. ADTrees are already discussed in the Introduction. Related to this are Attack Responses, these are similar to the countermeasures in an ADTree, but they approach the problem from a different perspective (Zonouz et al., 2009). Attack trees are based on all possible attack scenarios that are able to satisfy the goal of the attacker. However, Attack Responses are based on the attack consequences, (e.g. a SQL crash). The goal of Attack Responses is to find only those attack consequences that lead to the violation of an asset's security properties therefore, knowing all possible attack scenarios is not necessary. The Countermeasure gate can not be used to model our scenario for the same reason that the Attack Responses can not be used.

The final possible alternative is the Bayesian Belief Network (BBN). This is an approach that uses a graphical representation of prior probability distributions, represented in a Directed Acyclic Graph (DAG) (Heckerman, 1995). Each directed edge represents a dependence relation between 2 variables, meaning that the variable (B) is stochastically dependent on variable (A), written as $P(B \mid A)$. Each node in the graph includes a table containing conditional probabilities quantifying the influence strength of the other variables (Heckerman, 1995). Since the BBN approach is 'further' away from the approach involving the design of a new kind of node or gate, we chose not to follow the BBN option.

6.3 Regression node

In traditional attack trees, all 'children' (AND-gate) or any 'child' (OR-gate) must be executed to satisfy the parent node. In this chapter, we propose a 'regression-node' to model the 'leaf node' of an attack tree on the basis of the 'contextual' parameters, based on correlation coefficients.

Regression analysis is a technique that predicts an outcome from a model, based on the relation among input variables (Field et al., 2012). In the 'social engineer key' node presented in Figure 6.1, this would translate into estimating how context variables that the attacker is able to exercise or that describe him/her affect the compliance with the request to hand over the office key.

Logistic regression is used to predict binary outcomes, whereas a continuous outcome is predicted by linear regression. This means that the outcome of logistic regression is limited to the range between 0 or 1, whereas the outcome of a linear regression is any number between $-\infty$ and $+\infty$. Since the outcome of social engineering is either complying or not complying with a request, there is a need to limit the predicted outcome to a value that is either 0 or 1, thus the need to use the natural logarithm of the odds of the predictor variable (Gelman & Hill, 2007, p. 79-80).

In order to run a logistic regression, the dataset must fulfil three assumptions: *i*) Sufficient sample size, *ii*) no multicollinearity and *iii*) no outliers. The dataset should at least contain 10 Events Per Variable (EPV), which is considered as a minimum required for running a logistic regression (Peduzzi et al., 1996). The Variance Inflation Factor (VIF) statistic below the cut-off value of 10 indicates absence of multicollinearity (Peduzzi et al., 1996). In the case of dichotomous variables this means that one value should be placed in exactly one category.

In the regression node, the regression equation will replace traditional AND and OR-gates. A single regression equation consisting of: the outcome variable (i.e. dependent variable) and predictor variables (i.e. independent variables) will be used to estimate the compliance probabilities. The outcome variable is considered the construct of measurement, in the case of social engineering this is Compliance (whether or not a target complies with the request of the offender). The predictor variables are the variables that influence the outcome variable, in the case of social engineering this could be offender using authority or the target having received an intervention.

The basic logistic regression equation is:

$$LN\left(\frac{p}{1-p}\right) = \beta_0 + [\beta_1 \cdot x] \tag{6.1}$$

Equation 6.1 can be also written as Equation 6.2. However, for readability purposes the format of Equation 6.1 is preferred.

$$P(y) = \frac{1}{1 + e^{-(\beta_0 + [\beta_1 \cdot x])}} \tag{6.2}$$

where:
β_0 is the intercept (i.e. constant) and
β_1 is the coefficient of the predictor variable x to the outcome y.

A general mapping from regression equation to attack tree is as follows: *i*) the outcome variable of the regression equation (y) corresponds to the annotation of the parent node (e.g. Social Engineer Key) and *ii*) all regression predictor variables (e.g. x) correspond to the child nodes, whereas a combination of predictor variables

corresponds to one specific child node, as shown in the Example (refer to Section 6.3). Only one combination of context variables applies to the attack situation; this value (probability of success) represents the final value of the regression node and is used in the propagation towards the root node. The regression node is designed as a leaf node that does not allow further refinements. Unlike in other leaf nodes, the children in the regression node do not constitute atomic actions, instead they resemble context and are used in the calculations to adapt the outcome to specific situations.

The use of the regression node will be illustrated by means of a data set from a social engineering experiment.

Social engineer office key

The variables compliance, authority and intervention from the `f2f-se` dataset (refer to Table 1.2) were used to populate the regression node. The outcome variable *compliance* measured whether the subject complied with the request of the attacker to hand over the office key, coded as 0 = did not comply and 1 = did comply. The predictor variable *authority* measured the level of formality of the attacker's clothing (e.g. jeans, t-shirt) or formal clothing (e.g. suit, tie). The variable was coded as 0 = informally dressed and 1 = formally dressed. The predictor variable *intervention* measured whether the potential targets have been exposed to a social engineering awareness campaign. The variable was coded as 0 = did not receive an intervention and 1 = received an intervention. Refer to Chapter 3 for details regarding the experiment.

The probabilities of compliance are modelled based on the equation:

$$LN\left(\frac{p}{1-p}\right) = \beta_0 + [\beta_1 \cdot x] + [\beta_2 \cdot z] \tag{6.3}$$

where:
β_0 is the intercept (i.e. constant),
β_1 is the coefficient of the predictor variable x (i.e. *authority*) to the outcome y (i.e. *compliance*), when z (i.e. *intervention*) = 0 and
β_2 is the coefficient of the predictor variable z (i.e. *intervention*) to the outcome y (i.e. *compliance*), when x (i.e. *authority*) = 0.

This equation can be easily extended to include extra predictor variables. The mapping from regression equation to attack tree is done in the following way: *i*) The outcome variable of the regression equation (y) corresponds to the parent node 'social engineer key' (refer to Figure 6.1); *ii*) The children of the node 'social engineer key' correspond to the predictor variables. Depending on the context that applies to the attack, the outcome of Equation 6.3 is the 'final' value of the regression node and

propagates upwards.

6.3.1 Example

Regression analyses aim at making a model based on a given dataset. To illustrate the procedure and interpret the results of the regression node the dataset from a real social engineering experiment in a university environment is used, refer to Chapter 3.

Social engineer office key

Chapter 3 explored the extent to which *i*) an intervention reduces the effects of social engineering (e.g. the obtaining of access via persuasion) in an office environment and *ii*) the effect of authority is of influence during such an attack. In total, the offices of $N = 162$ employees were visited by thirty-five different 'attackers' who asked each employee (on the basis of a script) to hand over their office key. Authority, one of the six principles of persuasion, was used by half of the attackers to persuade a target to comply with his/her request. The authority condition was operationalized by clothing: the attacker wore either casual clothing (i.e. jeans and a t-shirt) or wore formal clothing (i.e. buttoned collar shirt and trousers). This particular dress code was used to mimic facility management personnel. Prior to the visit, an intervention was randomly administered to half of the targets to increase their resilience against attempts by others to obtain their credentials. The intervention contained *i*) an informing leaflet about the risks of social engineering attacks, *ii*) a small key chain, and *iii*) a humorous poster.

Among the employees that received an intervention, 36.96% handed their keys while 58.62% of those who were not exposed to it handed their key over. The intervention significantly reduced the compliance but authority did not significantly increase, refer to Chapter 3.

The F2F-SE dataset (refer to Table 1.2) fulfilled all three assumptions needed to run a logistic regression: *i*) There are at least 17 events per variable which is more than the required minimum of 10, *ii*) the VIF statistic for both authority and intervention is 1.08 which is below the cut-off value of 10, indicating absence of multicollinearity and *iii*) since there are only dichotomous variables used, the dataset is free of outliers.

The outcome of the logistic regression analysis is shown in Table 6.1. Of interest are the β coefficients, which are input for the regression equation (refer to Equation 6.3). The outcome of the equation are the probabilities of success for the attack tree's action 'social engineer key', and range between 36% and 60% (refer to Table 6.2).

Table 6.1: Output regression analysis for the node 'SOCIAL ENGINEER KEY' with dichotomous predictor variables ($N = 162$).

	β	SE	p-VALUE	95% CI
AUTHORITY	.071	.334	.831	(-.585 – .727)
INTERVENTION	-.894	.364	.014	(-1.607 – -.182)
CONSTANT	.324	.219	.139	(-.105 – .754)

Although the aspects of cost/benefit are not incorporated in the regression node, one can assume that the attacker would make the rational choice to maximize the probability of success. On the basis of the Rational Choice Theory (RCT), which is used to understand human behaviour, it is assumed that: *i*) the offender is a rational actor, *ii*) the offender makes an end/means or cost/benefit calculation and *iii*) the offender chooses to perform the behaviour based on rational calculations (Cornish & Clarke, 2014; Gül, 2009). Even if the goal of the attacker seems irrational, the methods and choices to achieve it are rational from the point of view of the attacker (Cornish & Clarke, 2014; Winstok, 2013). Therefore if the attacker manipulates the context by choosing not to apply authority in the attack, the probability of succeeding is either 36% or 58%. Assuming that the attacker is lucky and that the target did not receive an intervention, the probability of Social Engineering the Key successfully the will be 58%.

Table 6.2: Probabilities of success for the node 'SOCIAL ENGINEER KEY' with dichotomous predictor variables.

AUTH	INTERV	REGRESSION EQUATION	SUCCESS
0	0	$LN(\frac{p}{1-p}) = .324 + [.071 \cdot 0] + [-.894 \cdot 0]$	58
0	1	$LN(\frac{p}{1-p}) = .324 + [.071 \cdot 0] + [-.894 \cdot 1]$	36
1	0	$LN(\frac{p}{1-p}) = .324 + [.071 \cdot 1] + [-.894 \cdot 0]$	60
1	1	$LN(\frac{p}{1-p}) = .324 + [.071 \cdot 1] + [-.894 \cdot 1]$	38

Using a logistic regression has additional benefits regarding the representation in the tree. Alternatively, this would be an OR-gate with the Cartesian product of the context variables, resulting in an explosion of child nodes, refer to Figure 6.3. In the hypothetical case where there are 3 categorical variables with 3 options each, this would result in $3 \times 3 \times 3 = 27$ child nodes.

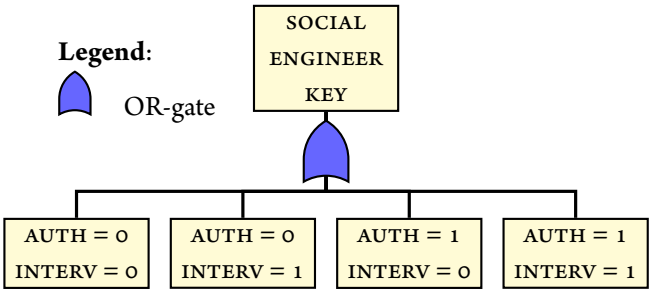


Figure 6.3: An (exploded) attack tree for the node 'SOCIAL ENGINEER KEY'. By putting each combination of refinements as an individual node, the number of nodes increases drastically.

6.3.2 Further Propagation

This section provides an example of how the Regression node result can be used to propagate to the root node. For the annotation of the other nodes in the scenario (refer to the attack tree in Figure 6.1) expert knowledge is used. Here we assume: *i*) 80% chance to bypass the firewall and *ii*) 60% chance to guess the password. Since 'HACK PC' is connected with an AND-gate to its 'children', the probability of success is calculated as the product: 80% × 60% = 48%. The probability succeeding to pick the lock is estimated at 50%.

By applying the OR-gate to 'ENTER OFFICE', the final result of this gate is 60%, obtained by the MAX-function of 60% (from the 'SOCIAL ENGINEER KEY') and 50% ('PICK LOCK'). Due to the higher probability of success, one would assume that the attacker would choose the option Social Engineer Key where authority is executed and (hopefully) the target got no intervention, rather than choosing to pick the lock. The leaf nodes are subsequently annotated with probabilities of success (refer to Figure 6.4).

This regression node enables: *i*) annotate leaf nodes in the tree with context variables, *ii*) incorporate data from a social science experiment in the tree, *iii*) limit the explosion of terms.

6.4 Conclusion

This chapter contributes to the field of socio-technical vulnerability assessment with an attack trees extension which allows the use of social science data such as that of social engineering experiments. This approach enables annotating the leaf nodes of an attack tree with success probabilities for context dependent properties which can

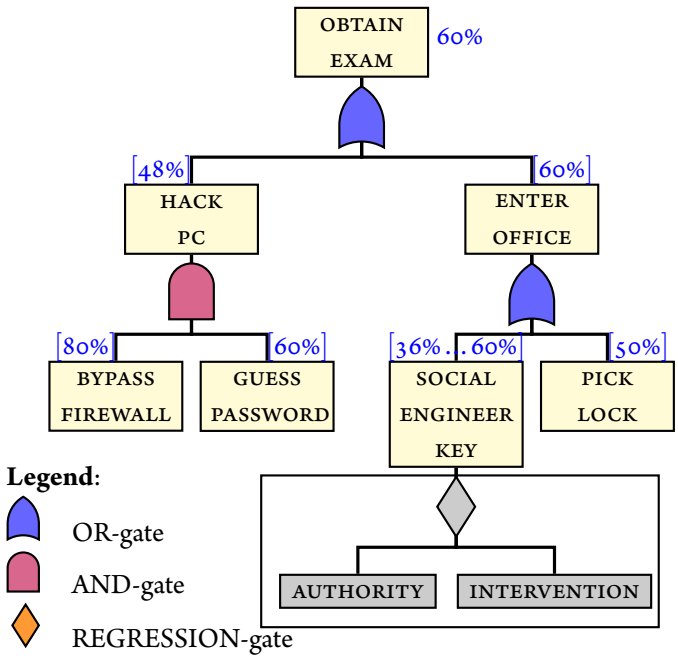


Figure 6.4: An attack tree for the node 'OBTAIN EXAM' annotated with probabilities of success. The box shows the inner working of the Social Engineering node. The percentage of the node 'SOCIAL ENGINEER KEY' depends on the activation of the refinements.

either involve none, one, multiple or all children. The process from experimental observations towards probability of success for the root of the attack tree was illustrated using a dataset obtained from a social engineering experiment. The proposed regression node is demonstrated to work in the context of a social engineering experiment.

One of the main advantages of this approach is that from a graphics point of view, it can handle the explosion of terms resulting from having to display all possible combinations of context variables. Moreover, using a conventional attack tree depiction, it would not be possible to deal with continuous predictor variables since there is an infinite number of possible combinations.

6.4.1 LIMITATIONS

The proposed extension has three limitations: *i*) In its current state, the regression node constitutes a leaf node. Its output is a probability of success that can be propagated upwards using propagation rules. *ii*) Each experiment applied to the leaf nodes should be independent of other experiments. Therefore, since in many cases experimental data is population specific, generalization across the attack tree is not possible.

This would rule out designing a single experiment to develop success probabilities for several nodes of the attack tree and. *iii*) It is possible that some social science experiments are deemed unsuitable, not from a research design but from a regression assumption viewpoint since not meeting data assumptions means that the results will be unreliable and hence useless.

6.4.2 Future work

Finally, we make recommendations for future research. So far, a regression node for dealing with social science experimental data has been proposed. However, a follow up study should assess whether the regression node, rather than being an extension of attack trees, could constitute a means to generalize attack trees. Gates could therefore represent functions of the input. In the already proposed regression node, it is already possible to represent AND-gates. When all variables are dichotomous, it is possible to make a logical AND relation between parent and child, where the parent node becomes one if–and only if–all child nodes have of value 1. In other words, this logical AND would represent an interaction coefficient. A first step towards generalizing trees to regressions is to make the regression node available as a normal node within the tree, instead of being a leaf node.

Furthermore, the integration of the costs for the attacker into the regression gate should be also explored. This is relevant since leaving the costs out of the node means that the attacker would select the highest probability of success. This is not ideal since using persuasion principles (e.g. buying a suit) have monetary and time implications. By including these variables in the regression, the analysis becomes a cost-benefit analysis. For example, the initial authority level of the attacker with respect to a particular person in the system could be 3, and that the additional attacker cost for increasing authority by 1 unit (i.e. to 4), would carry a cost of 100.

I know you've deceived me, now here's a surprise
I know that you have 'cause there's magic in my eyes
If you think that I don't know about the little tricks you've played
And never see you when deliberately you put things in my way

–"I Can See for Miles", The Who

7

Conclusion

This chapter is based on preceding.

This thesis investigated social engineering, in an organisational context, with the aim to find factors that explain or prevent the phenomena. Across different empirical studies, the targets were persuaded to comply with the requests of the offender. There were ten factors investigated, seven were controlled by (or a property of) the target, the other three belonged to the offender. Four factors had an influence on the outcome of a social engineering attack (i.e. culture, Years of Service (YoS), intervention and spear). The remaining six factors (i.e. target sex, offender sex, age, building, distance and authority) had no effect. For an overview of the factors, modalities and effect, refer to Table 1.1. Furthermore, data from the experiments was used as input for an extension of attack trees to do a quantitative risk assessment. For each chapter a brief summary is given:

Chapter 1 gave an introduction to social engineering and illustrated this with an example. Social engineering was explained by crime science and social psychology theories. Furthermore, the research question was defined, contributions were given and the outline of this thesis was discussed.

Chapter 2 contained a literature review, where 74 successful social engineering scenarios were dissected. In particular, the availability of persuasion principles within each scenario was investigated. The finding was that *i*) persuasion principles are often used in social engineering attacks, *ii*) Authority (one of the six persuasion principles) was used considerably more often than others and *iii*) single-principle attack steps occurred more often than multiple-principle ones. The scenario analysis illustrated how to exploit the human element in security. The findings supported the view that security mechanisms should include not only technical but also social countermeasures.

Chapter 3 discussed a Face-to-Face (F2F) social engineering experiment. The effects of authority, intervention and effort were investigated. The result was that a total of 36.96% of the employees who were exposed to the intervention complied with the offender while 58.62% of those who were not exposed did. The intervention has a significant effect on compliance, but the same was not the case for authority and effort. Awareness-raising about the dangers, characteristics, and countermeasures associated with social engineering proved to have a significant positive effect on neutralising the attacker. The likelihood of handing over the keys for employees close to a key activator was similar to those who were further away.

Chapter 4 described a social engineering via email (i.e. phishing) experiment. 593 employees received phishing emails requesting Personally Identifiable Information (PII). A spear phishing email opening was randomly used in half of the emails. 19.3% of the employees provided their PII in a general phishing email, compared to 28.9% in the spear phishing condition. Employees that have a high power distance cultural

background were more likely to provide PII compared to those with a low one. There was no effect of age on providing the requested PII when the recipient's YoS within the organisation is taken into account. This research shows that success is higher when the opening sentence of a phishing email is personalised. The resulting model allows practitioners to focus awareness campaigns to maximise their effect.

Chapter 5 described a study that aims to get insight into the effectiveness of an information campaign to counter a social engineering attack via the telephone. A total of 40% of the employees not exposed to the intervention followed the instructions of the offender. This was significantly different to those exposed to an intervention one week before the attack (17.2%); However, there was no effect for those exposed to an intervention two weeks before the attack (42.9%). This research suggests that scam awareness-raising campaigns reduce vulnerability only in the short term.

Chapter 6 provided a methodology to include the results of a social engineering experiment in risk analysis tools. In particular, an extension to attack trees, involving a regression-node was described. By allowing the annotation of leaf nodes with experimental data from social science, the regression-node enables the development of integrated socio-technical security models.

7.1 Implications for theory

This section discusses four implications for theory:

i) Previous research developed crime scripts for various offences (Tremblay et al., 2001; Rowe et al., 2012; Thompson & Chainey, 2011; Beauregard et al., 2007; Chiu et al., 2011), however there was no crime script for social engineering found. Chapter 2 describes crime scripts for social engineering based on 74 successful social engineering scenarios. These crime scripts allowed to identify the social influences that were used in successful social engineering attacks and contribute to a better understanding of the inner working of social engineering attacks.

ii) Previous field tests on social engineering via email (i.e. phishing) found that younger people were more likely to be victimised (Sheng et al., 2010; Tembe et al., 2013). The results in Chapter 4 suggests that there is no relation between age and victimisation. Careful investigation of the data showed that *a*) there is a quadratic relation between age and victimisation and *b*) including YoS in the model the effect of age disappears.

iii) Books by social engineers describe social engineering as single successful executed scenarios, e.g. (Mitnick & Simon, 2002) and (Mann, 2008). These accounts provide no details on the success rate of the attack and how vulnerable people are to

the threat. In Chapter 3 and 5 are two accounts of social engineering systematically tested. The results provide insight into the success of social engineering attacks.

iv) In Chapter 3 and 5 both a survey and a field test were used to assess the success of social engineering. The use of surveys is a convenient way to collect data, whereas conducting a field test requires considerably more effort. This effort relates to the ethical requirements when conducting experiments that involve human subjects and deception (Belmont Report, 1979). The result of the survey showed that only a few people would become a victim, whereas the field test showed that 40% and 58.62% (in Chapter 3 and 5 respectively) were victimised. When assessing organisational security, the most accurate view of the situation can be obtained by using field test.

7.2 Implications for the practice

The results of this research have the following implications for practitioners:

i) Awareness-raising about the dangers, characteristics, and countermeasures that are associated with social engineering reduces the probability of victimisation (refer to Chapter 3 and 5). Regarding the contents of an awareness raising training; The authority principle is most frequent used in successful social engineering attacks (refer to Chapter 2). How to recognise this and how to react should be part of training for example. Related to this is the aspect of challenging the identity of the party that initiated contact (refer to Chapter 2).

ii) Awareness raising training was only effective for a short period (refer to Chapter 3 and 5). Therefore, one should remember that a single round of awareness training is insufficient. High levels of repetition of the same message is not the solution either; this can produce adverse results (Stewart & Martin, 1994). The solution is likely to lie somewhere in the middle.

iii) Caution should be exercised with regard to the distribution of awareness raising campaigns (refer to Chapter 3 and 5). For e.g. F2F social engineering, 58.62% of the employees was victimised, while 3.2% thought that they would be victimised. Notice the discrepancy between the actual behaviour and intended behaviour (refer to Chapter 3). Therefore, if people do not see the urgency of the problem, why would they accept any countermeasure? One explanation for this is explained by the optimism bias. A likely thought of an employee could be: "I am less likely to be targeted than my colleague, and if I was targeted, I am better in resisting than my colleague. Therefore this countermeasure does not apply to me." Measures that reduce the optimism bias should therefore be an integral part of the awareness raising training.

iv) Vulnerable groups should get special attention (refer to Chapter 4). It was

found that recently hired personnel and employees with a high Power Distance (PDI) cultural backgrounds are more vulnerable. An easy way to reduce the vulnerability is to provide awareness training for those two groups, refer to Chapter 4. It was shown by Price Waterhouse Coopers that the organisations that provided security awareness programs for new employees had less financial loss caused by cyber security incidents compared to those who did not provide such training for their new employees, 162.000 and 683.000 United States Dollar (USD) respectively (PWC, 2014). There was no effect of sex or age found. Therefore specific training for males, females, younger and elder is not necessary (refer to Chapter 3, 4 and 5).

7.3 Evaluation of methods

The empirical part of the thesis (i.e. Chapters 3, 4 and 5) uses surveys and field studies to collect data. For each method, the benefits, limitations and experiences are briefly discussed.

Surveys are an easy and practical tool to collect data in a relatively cost effective way. With limited resources, respondents can be asked about their thoughts and opinions. However, it is also argued that surveys are prone to socially desirable responses (Crowne & Marlowe, 1960). The latter is shown in Section 3.3.1 and 5.3.2.

The surveys used in Chapters 3 and 5 were based on the scenarios that were used in the related chapters. Since the scenarios were already developed, transforming them to a questionnaire was of low effort. Each step in the scenario (relating to an action) was translated into an action. In case of the telephone survey, the first action involved picking up the phone. Given the context of the field study, the first question in the survey was: "You receive an anonymous phone call at your office phone. What would you do?" *a*) Answer the phone. *b*) Not answer the phone.

The Institutional Review Board (IRB) accepted the survey study without any problems. The experimenter who did the data collection received much positive feedback and the vast majority was willing to fill in the survey. In a short period a sufficient amount of data was collected.

Regarding the field studies, those took a considerable amount of time in the preparation phase. The biggest hurdle in my experience is the IRB approval. Careful planning and consideration are required, since social engineering, by definition involves deception (refer to Section 1.1.1). The most accurate results (regarding security testing) are obtained when the subject is tested in the natural environment and unaware of being tested. The combination of these two makes it difficult, refer to (Belmont Report, 1979; Code of Federal Regulations, 2005).

Second step is informing the parties that are involved. In the case of the F2F social engineering field study, refer to Chapter 3, this included campus security, facility management, building maintenance, the managing director of the faculty, the secretary of the faculty, the IT service desk and the IRB. One needs to carefully think how the subject could react and try to anticipate on that. It is important to stress that the existence of the field test should not be shared with the potential subjects.

The majority of the over 900 subjects who participated in the field studies provided positive feedback. There was only one complaint was received by the IRB with the request to be removed from the field study. The given reason was that the person felt uncomfortable participating in the study and preferred to be left alone.

There is one particular positive situation that I remember vividly and like to share. The day after the data collection of the telephone experiment, I did a F2F evaluation of the study with some of the subjects. When asking the subject about their experience in participating in security research, this was the response: "I absolutely liked being part of this study. I never thought I would become a victim of a telephone scam. I sometimes read these accounts in news papers or popular media, and have to laugh. How could someone be that stupid to fall for something an obviously fake? Now I am a victim myself. This experience made me realize that when not focussing on the situation, it is easy to become a victim." What I like about this situation is that it reflects what was found in the results: people think that they would not become a victim while where is a real chance that they would become a victim.

7.4 Practitioners feedback

The results from the experiments (refer to Chapter 3, 4 and 5) were discussed with 5 practitioners in the organisation. The pool of practitioners consisted of 2 Information Technology (IT)-department security managers, two Facility Management (FM)-department managers and the Campus Security manager. A 1-to-1 F2F interview was used to discuss the following topics: *i*) Their reaction towards the results, *ii*) What should be changed in the department or organisation, *iii*) How the results help in their job, *iv*) What result is most relevant, *v*) Results that are missing, *vi*) Where the results already used to make changes and *vii*) Their expectations regarding the results of a retest in 5 years. First F2F experiment will be discussed and second the email and telephone experiment.

Regarding F2F social engineering, the findings were discussed with the campus security manager and two members of facility management. The reaction from those of FM was that these results were 'non-surprising'. Two reasons were given: *i*) no

single system brings universal happiness, *ii*) there is a 'trust culture', people feel safe within the organisation. Regarding change, there were mixed reactions. At this moment there is no sufficient prevention, and it should be brought to people's attention more often. One suggestion was to make people more 'self-responsible'. A counter argument is that, due to these specific keys, the window of opportunity for offenders is limited. Once a key is reported as 'missing', the key can be disabled centrally. The key (if activated) can only be used until midnight. Therefore nothing should be changed since this risk is acceptable. The results are useful and contribute in their job, in the sense that now there are facts and these help in discussions and decision making. So far, based on the results of these studies, the results are shared within the periodic department meetings. No concrete actions were taken. If the experiment was repeated in five years time; it is expected that (given the same conditions) results will be similar. What was suggested for future work is to find out how this particular type of key compares to other type of keys (e.g. traditional or RFID cards). Another suggestion was to change the offender clothing and mimic a campus security officer. Finally, it was noted that no requests were received to install additional activation points for the keys (e.g. one activator on each floor).

The findings related to the email and telephone social engineering experiments were discussed with two security managers from the IT-department, since these experiments relate directly to their department. They were satisfied with the results, in the sense that it is a confirmation of what was expected, but could not be proven. Both managers agree that something should be changed and came with three suggestions: *i*) The awareness campaigns should continue; however this should happen in an intelligent way, we have to carefully think about this. *ii*) The organisation should update its procedures, to make people validate the identity of others. *iii*) The communication originating from the IT-department to the organisation should change. If the root of email phishing is removed (i.e. email), then they believe the problem will disappear. It is acknowledged that the experiments support them in their job. It helps to create awareness for the management team. "Now we can support our claims with numbers, this will make future discussions easier." Another important aspect, it allows for reflection on the internal procedures. One of them said: "We learn from this too. For example, it is great that employees show initiative, but one should still follow the procedure." The result that stood out was the phishing victimisation rate. The experimental data showed much more victimisation than that was shown in the logging. It must be noted that the logging shows the abuse of credentials on the corporate network, the number of employees that gave their credentials away could be much higher. It was also surprising to see that the telephone modality had such a high victimisation rate

since there were only four reported incidents in the past. There were two suggestions for future work regarding the email social engineering experiment: *i*) include the level of seniority (i.e. professor, secretary, PhD candidate) as an independent variable and *ii*) another type of phishing email, one that contains grammatical errors. The latter was suggested since one of them felt that the email was too 'clean'. Their expectation regarding a repetition of the experiment in five years time, it is expected that all victimisation rates are lower compared to the current rates. One said: "I hope it will reduce; otherwise, I can better start doing something else."

7.5 Suggestions for corporate policies

The results of this research lead to six suggestions for corporate policies. Each policy is briefly explained; what is the policy, why this is relevant, how it could be used and who are involved. For an overview of policies and responsibilities, refer to Table 7.1.

i) *Secure communication.* The organisation should not communicate using email, but via a Secure Messaging Portal (SMP) instead. Since both authorities and non authorities use the same communication medium offenders can impersonate authorities and manipulate employees. Communication through a SMP overcomes this problem by removing the open communication medium. The line of thought is that if the organisation does not communicate via email, it is pointless to launch an attack that impersonates communication via a medium that is no longer in use. Furthermore, the messages received via the SMP are for sure from the organisation and of high-trust, whereas the messages in the general email box should be perceived as low-trust. If there is a need to communicate from the organisation to the employee, this message is placed in the SMP of the employee. This can be an online environment where the user has its private space. It is key to use this system only for business communications that originate from a limited set of authorized senders. If spam or unwanted messages end up in here, the acceptance of such a system will be low and is likely to fail. Therefore guidelines and regulations are needed that describe for what purposes this system can and can not be used. A comparable system is already in use by banks in The Netherlands. Those who are involved can be split into two groups; those who do the implementation and those who use the system. The IT-department is involved since this is a technical solution that requires an implementation. Those who use the system are those who send out the communications such as the FM, Human Resource Management (HRM), IT, Finance department, management and board of directors.

ii) *Security testing among employees.* The organisation should perform penetration tests, that involves their employees, on a regular basis. Testing the security within the

organisation provides insights into its strengths and weaknesses. The results of such a test points out where investment in countermeasures could be beneficial. It is important to this internally, since using data from other organisations is only realistic up to some extent. Besides the difficulty of obtaining data from comparable organisations, using your own data gives the most accurate insight. A penetration test that involves human test subjects includes the IRB and consent forms. One of the hurdles to overcome is employees being aware of being 'tested'. In the case that an employee is aware of being tested will bias the results. One way to overcome this is to make a consent which is valid for a longer predefined period, e.g. one year. It should be noted that this can introduce a sampling bias. Involved are the IRB and the Faculty Dean. The role of the IRB is to oversee the ethical aspects that are involved, whereas the Dean as the Faculty manager has to approve and oversee this structure.

iii) *Responsibility for employees.* Employees should be aware of the open characteristic of the organisation and the corresponding responsibilities. Due to the open characteristic of organisation (i.e. a university), it is possible for an outsider to walk up to the office of an employee. The door of the employee his office is (during office hours) the only line of defence. In many cases, the door is not even locked, since the employee is inside. Sometimes the doors are unlocked and open without the employee being in his room. There are no other defences such as a front or department door with access control. This makes the university an easy target for offenders. Therefore, the open characteristic of the organisation comes with additional responsibilities for the employee. Although employees think that e.g. theft from their office will never happen, this happens more often than expected. According to the Campus Security incident reports, on average (over the past five years) theft was reported more than once per week. A first step to overcome this, is to make employees aware that they can contribute to the security of the organisation. A second step would be to put this in writing and make it available to employees. Finally, a periodical systematic test to evaluate the current status. Involved are the HRM-department since they are involved with the new employee. Furthermore, this also relates to building and office security. Therefore Campus Security and the FM-department are involved in policy making and testing.

iv) *Do risk assessments.* Make social, digital and physical risk assessments mandatory before purchase or installation. Attacks evolve in parallel with the technology. Offenders adapt their strategies to stay in the business. In case there is a need to migrate or install a new system (e.g. smart locks or a VOIP telephone system) do a risk assessment. Price should not be the only consideration; a bad bargain is dear at a farthing. There are various tools and procedures available that can help in performing

different kinds of risk assessments. The owner of the system should take care of the risk assessment. In the case of smart locks, this is FM, whereas the VOIP telecommunication system relates to the IT-department.

v) *Multi-department involvement.* Organisational departments should communicate and share their policies and strategies. Providing transparency regarding departmental policies can prevent unintended negative effect on other departments. The responsibility of e.g. the IT-department is to keep the corporate network secure. The outcome of Chapter 4 shows that younger employees who have up to 7 YoS for the organisation are the group that is most vulnerable to phishing emails. When the HRM-department updates their policy and focusses on employees having short and temporary contracts, this introduces a security risk. To overcome this, new policies or information (e.g. from logging or an experiment) should be discussed with related departments. The related departments are e.g. HRM, FM and IT.

vi) *Put security on the agenda (literally).* Make it mandatory for all organised meetings to add the item 'Security' to the agenda. Informing people about the dangers and countermeasures associated with social engineering has a significant positive effect on reducing the threat for a short time, refer to Chapter 3 and 5. A single round of awareness training is insufficient as well as high levels of repetition of the same message (Stewart & Martin, 1994). An employee is at least once every two weeks in a meeting (i.e. 26 meetings in one year). This could result in 26 security awareness moments, which is more than the current situation for many organisations. There are several ways to put this item to practice; it is up to the chairman of the meeting to decide how to do this in particular. For example; share an experience with those in the meeting. There is always someone who experienced or noticed something that is related to this topic. Sharing makes the others informed too. A second suggestion is to watch a 1-minute video that discusses an attack and the countermeasures provided by the organisation or on the Internet (e.g. the 'Cyber Security Minute' or the 'Red Team USA' channels on `youtube.com`). Those who are involved are the chairman of meeting since they should put the items on the agenda. Furthermore, there is the Campus Security, FM and IT-department that can assist in the production or selection of relevant material.

Table 7.1: Overview of policies and responsibilities.

POLICY	WHO	ROLE
i) Secure communication	IT	System implementation
	Finance, HRM, IT Dean, Rector[1]	Send communications
ii) Security testing	IRB	Ethical considerations
	Dean	Approval and oversight
iii) Responsibility	HRM	Informing the employees
	CSec[2], FM	Making and testing policies
iv) Risk assessment	FM, FM, IT	Performing the assessments
v) Multi-department	FM, HRM, IT	Communication of policies & findings
vi) Security agenda	CSec, FM, IT	Provide the 'learning' materials
	Chairmen	Put item on agenda

[1] Rector is also referred to as Chancellor, President or Principal;
[2] CSec = Campus Security;

7.6 Future work

We have learned that F2F and telephone social engineering can be countered. Further research is needed to fully understand how to efficiently reduce the threat. There are two suggestions presented. First, in this thesis the effect of an intervention is discussed in Chapters 3 and 5. The effect of the interventions was only present in the short term. Therefore a suitable follow-up study could involve the use of 'quick' booster interventions to counter the decay effect. This was already tested in the context of Cardiopulmonary Resuscitation (CPR) skill retention (Sutton et al., 2011).

Second, all the experiments in this thesis were a snapshot of the organisation. The results tell something of the state at one particular moment. An interesting future work would be to do a longitudinal study where subjects are tested multiple times over a period. It would particularly interesting to see how being 'victimized' in such a test influences the behaviour in future tests.

Bibliography

Author's Publications

Bullée, J. H., Montoya, L., Junger, M., & Hartel, P. H. (2016). Telephone-based social engineering attacks: An experiment testing the success and time decay of an intervention. In A. Mathur & A. Roychoudhury (Eds.), *Proceedings of the inaugural singapore cyber security r&d conference (sg-crc 2016), Singapore, Singapore* (Vol. 14, pp. 107–114). Amsterdam: IOS Press. doi: 10.3233/978-1-61499-617-0-107

Bullée, J. H., Montoya, L., Junger, M., & Hartel, P. H. (2017a). *Physical location of smart key activators - a building security penetration test.* (Manuscript submitted for publication)

Bullée, J. H., Montoya, L., Junger, M., & Hartel, P. H. (2017b). *Social engineering - een experimentele benadering.* (Manuscript submitted for publication)

Bullée, J. H., Montoya, L., Junger, M., & Hartel, P. H. (in press). Spear phishing in organisations explained. *Information and Computer Security*, *xx*(x), xx–xx. doi: 10.1108/ICS-03-2017-0009

Bullée, J. H., Montoya, L., Pieters, W., Junger, M., & Hartel, P. H. (2015a). The persuasion and security awareness experiment: reducing the success of social engineering attacks. *Journal of Experimental Criminology*, *11*(1), 97–115. doi: 10.1007/s11292-014-9222-7

Bullée, J. H., Montoya, L., Pieters, W., Junger, M., & Hartel, P. H. (2015b). Regression nodes: Extending attack trees with data from social sciences. In *Workshop on socio-technical aspects in security and trust (stast), Verona, Italy.* USA: IEEE Computer Society. doi: 10.1109/STAST.2015.11

Bullée, J. H., Montoya, L., Pieters, W., Junger, M., & Hartel, P. H. (2017). On the anatomy of social engineering attacks - a literature-based dissection of successful attacks. *Journal of Investigative Psychology and Offender Profiling*, 1–26. doi: 10.1002/jip.1482

Bibliography

Abraham, S., & Chengalur-Smith, I. (2010). An overview of social engineering malware: Trends, tactics, and implications. *Technology in Society*, *32*(3), 183 - 196. doi: 10.1016/j.techsoc.2010.07.001

Ajzen, I. (1988). *Attitudes, personality, and behavior (Mapping social pschology series)*. Dorsey Press.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. doi: 10.1016/0749-5978(91)90020-T

Anandpara, V., Dingman, A., Jakobsson, M., Liu, D., & Roinestad, H. (2007). Phishing iq tests measure fear, not ability. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *4886 LNCS*, 362-366.

Anderson, R. (2008). *Security Engineering: A Guide to Building Dependable Distributed Systems*. Wiley.

Ashworth, A., Hill, C. M., Karmiloff-Smith, A., & Dimitriou, D. (2014). Sleep enhances memory consolidation in children. *Journal of Sleep Research*, *23*(3), 304–310. doi: 10.1111/jsr.12119

Assange, J. (2011). *Julian assange: The unauthorised autobiography*. Edinburgh: Canongate Books.

Bagnato, A., Kordy, B., Meland, P. H., & Schweitzer, P. (2012). Attribute decoration of attack-defense trees. *Int. J. Secur. Softw. Eng.*, *3*(2), 1–35. doi: 10.4018/jsse.2012040101

Ball, S., Eckel, C. C., & Heracleous, M. (2010). Risk aversion and physical prowess: Prediction, choice and bias. *Journal of Risk and Uncertainty*, *41*(3), 167–193. doi: 10.1007/s11166-010-9105-x

Bandura, A. (1986). *Social foundations of thought and action* (First Printing ed.). Prentice Hall.

Barlow, J. (1998). Knowledge in patients with rheumatoid arthritis: a longer term follow- up of a randomized controlled study of patient education leaflets. *Rheumatology*, *37*(4), 373–376. doi: 10.1093/rheumatology/37.4.373

Barsky, R. B., Juster, F. T., Kimball, M. S., & Shapiro, M. D. (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. *The Quarterly Journal of Economics*, *112*(2), 537-579. doi: 10.1162/003355397555280

Beauregard, E., Proulx, J., Rossmo, K., Leclerc, B., & Allaire, J.-F. (2007). Script analysis of the hunting process of serial sex offenders. *Criminal Justice and Behavior*, *34*(8), 1069-1084. doi: 10.1177/0093854807300851

Beauregard, E., Rebocho, M. F., & Rossmo, D. K. (2010). Target selection patterns in rape. *Journal of Investigative Psychology and Offender Profiling*, *7*(2), 137–152. doi: 10.1002/jip.117

Belmont Report. (1979). *The belmont report: ethical principles and guidelines for the protection of human subjects of research*. The Commission.

Bickman, L. (1974). The social power of a uniform1. *Journal of Applied Social Psychology*, *4*(1), 47–61. doi: 10.1111/j.1559-1816.1974.tb02599.x

Bistarelli, S., Fioravanti, F., & Peretti, P. (2006). Defense trees for economic evaluation of security investments. In *First international conference on availability, reliability and security (ares'06)* (p. 416-423). doi: 10.1109/ARES.2006.46

Blass, T. (1999). The Milgram Paradigm After 35 Years: Some Things We Now Know About Obedience to Authority1. *Journal of Applied Social Psychology*, *29*(5), 955–978. doi: 10.1111/j.1559-1816.1999.tb00134.x

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214-234. doi: 10.1207/s15327957pspr1003_2

Bond, R., & Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using asch's (1952b, 1956) line judgment task. *Psychological Bulletin*, *119*(1), 111 - 137. doi: 10.1037/0033-2909.119.1.111

Bosworth, S., Kabay, M., & Whyne, E. (2014). *Computer security handbook* (6th ed.). New York: Wiley.

Brooke, P. J., & Paige, R. F. (2003). Fault trees for security system design and analysis. *Computers & Security*, *22*(3), 256 - 264. doi: 10.1016/S0167-4048(03)00313-4

Broomfield, R. (1996). A quasi-experimental research to investigate the retention of basic cardiopulmonary resuscitation skills and knowledge by qualified nurses following a course in professional development. *Journal of Advanced Nursing*, *23*(5), 1016–1023. doi: 10.1111/j.1365-2648.1996.tb00084.x

Broos, A. (2005). Gender and information and communication technologies (ict) anxiety: Male self-assurance and female hesitation. *CyberPsychology & Behavior*, *8*(1), 21–31. doi: 10.1089/cpb.2005.8.21

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, *6*(3), 203–242. doi: 10.1111/j.1468-2885.1996.tb00127.x

Cahill, L., Prins, B., Weber, M., & McGaugh, J. L. (1994, 20). β-adrenergic activation and memory for emotional events. *Nature*, *371*(6499), 702–704. doi: 10.1038/371702a0

Camtepe, S., & Yener, B. (2007). Modeling and detection of complex attacks. In *Security and privacy in communications networks and the workshops, 2007. securecomm 2007. third international conference on* (p. 234-243). doi: 10.1109/SECCOM.2007.4550338

Carlson, K. A. (2011). The impact of humor on memory: Is the humor effect about humor? *Humor - International Journal of Humor Research*, *24*(1). doi: 10.1515/humr.2011.002

Carré, P. C., Roche, N., Neukirch, F., Radeau, T., Perez, T., Terrioux, P., … Huchon, G. (2008). The Effect of an Information Leaflet upon Knowledge and Awareness of COPD in Potential Sufferers. *Respiration*, *76*(1), 53–60. doi: 10.1159/000115947

Casner, S., Heraldez, D., & Jones, K. (2006). Retention of aeronautical knowledge. *International Journal of Applied Aviation Studies*, *6*(1), 71-98.

Chan, M., Woon, I., & Kankanhalli, A. (2005). Perceptions of information security in the workplace: Linking information security climate to compliant behavior. *Journal of Information Privacy and Security*, *1*(3), 18-41. doi: 10.1080/15536548.2005.10855772

Chikudate, N. (2009). If human errors are assumed as crimes in a safety culture: A lifeworld analysis of a rail crash. *Human Relations*, *62*(9), 1267-1287. doi: 10.1177/0018726709335543

Chiu, Y.-N., Leclerc, B., & Townsley, M. (2011). Crime script analysis of drug manufacturing in clandestine laboratories: Implications for prevention. *British Journal of Criminology*, *51*(2), 355-374. doi: 10.1093/bjc/azr005

Cialdini, R. (2009). *Influence*. New York: HarperCollins.

Cialdini, R., Vincent, J. E., Lewis, S. K., Catalan, J., Wheeler, D., & Darby, B. L. (1975). Reciprocal concessions procedure for inducing compliance: The door-in-the-face technique. *Journal of Personality and Social Psychology*, *31*(2), 206 - 215. doi: 10.1037/h0076284

Clarke, R. (1997). *Situational crime prevention: Successful case studies*. Lynne Rienner Publishers.

Code of Federal Regulations. (2005). *Title 45: Public Welfare, Department of Health and Human Services, Part 46: Protection of Human Subjects*. U.S. Government Printing Office.

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, *44*(4), 588-608. doi: 10.2307/2094589

Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin, 116*(3), 457 - 475. doi: 10.1037/0033-2909.116.3.457

Cornish, D. B. (1994). The procedural analysis of offending and its relevance for situational prevention. *Crime prevention studies, 3*, 151–196.

Cornish, D. B., & Clarke, R. V. (1987). Understanding crime displacement: an application of rational choice theory. *Criminology, 25*(4), 933–948. doi: 10.1111/j.1745-9125.1987.tb00826.x

Cornish, D. B., & Clarke, R. V. (2003). Opportunities, precipitators and criminal decisions: A reply to wortley's critique of situational crime prevention. *Crime prevention studies, 16*, 41–96.

Cornish, D. B., & Clarke, R. V. (2014). *The reasoning criminal: Rational choice perspectives on offending*. Transaction Publishers.

Craik, F. I. M., & Blankstein, K. R. (1975). Psychophysiology and human memory. In R. in Psychophysiology (Ed.), (p. 388-417). London, UK: John Wiley & Sons.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior, 11*(6), 671–684. doi: 10.1016/S0022-5371(72)80001-X

Crossler, R. E., Johnston, A. C., Lowry, P. B., Hu, Q., Warkentin, M., & Baskerville, R. (2013). Future directions for behavioral information security research. *Computers & Security, 32*, 90 - 101. doi: 10.1016/j.cose.2012.09.010

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of consulting psychology, 24*(4), 349. doi: 10.1037/h0047358

Dang, H. (2008). The origins of socal engineering. *McAffee Security Journal, 1*(1), 4–8.

Deevy, M., Lucich, S., & Beals, M. (2012). *Scams, schemes & swindles a review of consumer financial fraud research* (Tech. Rep.). Financial Fraud Research Centre.

Dimkov, T., Pieters, W., & Hartel, P. H. (2009). *Two methodologies for physical penetration testing using social engineering* (Technical Report No. TR-CTIT-09-48). Enschede.

Doob, A. N., & Gross, A. E. (1968). Status of Frustrator as an Inhibitor of Horn-Honking Responses. *The Journal of Social Psychology, 76*(2), 213–218. doi: 10.1080/00224545.1968.9933615

Drewnowski, A., & Murdock, B. B. (1980). The role of auditory features in memory span for words. *Journal of Experimental Psychology: Human Learning and Memory, 6*(3), 319 - 332. doi: 10.1037/0278-7393.6.3.319

Dreyfus, S., & Assange, J. (2012). *Underground: Tales of hacking, madness and obsession on the electronic frontier*. Edinburgh: Canongate Books.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology* (No. 3). Teachers College, Columbia University.

Edmondson, A. C. (1996). Learning from mistakes is easier said than done: Group and organizational influences on the detection and correction of human error. *The Journal of Applied Behavioral Science*, *32*(1), 5-28. doi: 10.1177/0021886396321001

Ershoff, D. H., Mullen, P. D., & Quinn, V. P. (1989). A randomized trial of a serialized self-help smoking cessation program for pregnant women in an HMO. *American Journal of Public Health*, *79*(2), 182–187. doi: 10.2105/AJPH.79.2.182

Esgate, A., Groome, D., & Baker, K. (2005). *An introduction to applied cognitive psychology*. Psychology Press.

Favila, S. E., & Kuhl, B. A. (2014). Stimulating memory consolidation. *Nature Neuroscience*, *17*(2), 151–152. doi: 10.1038/nn.3638

Feeley, T. H., Anker, A. E., & Aloe, A. M. (2012). The door-in-the-face persuasive message strategy: A meta-analysis of the first 35 years. *Communication Monographs*, *79*(3), 316-343. doi: 10.1080/03637751.2012.697631

Ferguson, A. J. (2005). Fostering E-Mail Security Awareness: The West Point Carronade. *EDUCASE Quart.*, *1*, 54–57.

Festinger, L. (1957). *A Theory of Cognitive Dissonance.* Stanford University Press.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using r*. London: SAGE Publications.

Finn, P., & Jakobsson, M. (2007). Designing ethical phishing experiments. *IEEE Technology and Society Magazine*, *26*(1), 46-58. doi: 10.1109/MTAS.2007.335565

Flight, I., Wilson, C., & McGillivray, J. (2012). Turning intention into behaviour: The effect of providing cues to action on participation rates for colorectal cancer screening. *Colorectal Cancer-From Prevention to Patient Care. Shanghai: InTech*, 67–86. doi: 10.5772/27620

Flores, W. R., Holm, H., Nohlberg, M., & Ekstedt, M. (2015). Investigating personal determinants of phishing and the effect of national culture. *Information & Computer Security*, *23*(2), 178-199. doi: 10.1108/ICS-05-2014-0029

Fovino, I. N., Masera, M., & Cian, A. D. (2009). Integrating cyber attacks within fault trees. *Reliability Engineering & System Safety*, *94*(9), 1394 - 1402. ({ESREL} 2007, the 18th European Safety and Reliability Conference) doi: 10.1016/j.ress.2009.02.020

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology, 4*(2), 195 - 202. doi: 10.1037/h0023552

Furnell, S. (2007). Phishing: can we spot the signs? *Computer Fraud and Security, 2007*(3), 10-15. doi: 10.1016/S1361-3723(07)70035-0

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models* (Vol. Analytical methods for social research). New York: Cambridge University Press.

Ghaderi, F., Adl, A., & Ranjbar, Z. (2013). Effect of a leaflet given to parents on knowledge of tooth avulsion. *European journal of paediatric dentistry : official journal of European Academy of Paediatric Dentistry, 14*(1), 13-16.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European Review of Social Psychology, 2*(1), 83-115. doi: 10.1080/14792779143000033

Gisquet-Verrier, P., & Riccio, D. C. (2012). Memory reactivation effects independent of reconsolidation. *Learning & memory (Cold Spring Harbor, N.Y.), 19*(9), 401–9. doi: 10.1101/lm.026054.112

Glanz, K., Rimer, B. K., & National Cancer Institute, U. (1997). *Theory at a glance: a guide for health promotion practice*. U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, National Cancer Institute.

Glenberg, A. M. (1984). A retrieval account of the long-term modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(1), 16 - 31. doi: 10.1037/0278-7393.10.1.16

Grabosky, P. N. (2001). Virtual criminality: Old wine in new bottles? *Social and Legal Studies, 10*(2), 243-249.

Gragg, D. (2003). A multi-level defense against social engineering. *SANS Reading Room, March, 13*.

Graves, R. (1992). *The greek myths*. Penguin Books.

Greene, W. (2011). *Econometric analysis*. Pearson Education.

Greening, T. (1996). Ask and ye shall receive: A study in “social engineering”. *SIGSAC Rev., 14*(2), 8–14. doi: 10.1145/228292.228295

Greenspan, S. (2008). *Annals of Gullibility: Why We Get Duped and How to Avoid It*. Westport: Praeger.

Grewal, D., & Kavanoor, S. (1997). Comparative versus noncomparative advertising: A meta-analysis. *Journal of Marketing, 61*(4), 1. doi: 10.2307/1252083

Griskevicius, V., Goldstein, N. J., Mortensen, C. R., Sundie, J. M., Cialdini, R. B., & Kenrick, D. T. (2009). Fear and loving in las vegas: Evolution, emotion, and persuasion. *Journal of Marketing Research (JMR), 46*(3), 384 - 395. doi: 10.1509/jmkr.46.3.384

Gül, S. K. (2009). An evaluation of the rational choice theory in criminology. *Girne American University Journal of Social and Applied Science, 4*(8), 36–44.

Gulas, C. S., & Weinberger, M. G. (2006). *Humor in Advertising: A Comprehensive Analysis.* M.E. Sharpe, Incorporated.

Gupta, M., Agrawal, S., & Garg, N. (2011). A survey on social engineering and the art of deception. *International Journal of Innovations in Engineering and Technology, 1*(1), 31-35.

Hadnagy, C., & Wilson, P. (2010). *Social engineering: The art of human hacking.* New York: Wiley.

Halek, M., & Eisenhauer, J. G. (2001). Demography of risk aversion. *The Journal of Risk and Insurance, 68*(1), 1-24. doi: 10.2307/2678130

Happ, C., Melzer, A., & Steffgen, G. (2016). Trick with treat – Reciprocity increases the willingness to communicate personal data. *Computers in Human Behavior, 61*, 372–377. doi: 10.1016/j.chb.2016.03.026

Harley, D., Grooten, M., Burn, S., & Johnston, C. (2012). My pc has 32,539 errors: how telephone support scams really work. *22nd Virus Bulletin International Conference (VB2012).*

Harris, L. (2007). *Cliffsap psychology.* New York: John Wiley & Sons.

Harris, P., Middleton, W., & Joiner, R. (2000). The typical student as an in-group member: eliminating optimistic bias by reducing social distance. *European Journal of Social Psychology, 30*(2), 235–253. doi: 10.1002/(SICI)1099-0992

Hart, A. R., Barone, T. L., Gay, S. P., Inglis, A., Griffin, L., Tallon, C. A., & Mayberry, J. F. (1997). The effect on compliance of a health education leaflet in colorectal cancer screening in general practice in central England. *Journal of Epidemiology & Community Health, 51*(2), 187–191. doi: 10.1136/jech.51.2.187

Hawkey, G. M., & Hawkey, C. J. (1989). Effect of information leaflets on knowledge in patients with gastrointestinal diseases. *Gut, 30*(11), 1641–1646. doi: 10.1136/gut.30.11.1641

Heckerman, D. (1995). *A tutorial on learning with bayesian networks.* (Technical Report No. MSR-TR-95-06). Microsoft Research.

Hendrickson, S. M. L., Goldsmith, T. E., & Johnson, P. J. (2006). Retention of airline pilots' knowledge and skill. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(17), 1973-1976. doi: 10.1177/154193120605001755

Hersch, J. (1996). Smoking, seat belts, and other risky consumer decisions: Differences by gender and race. *Managerial and Decision Economics, 17*(5), 471–481. doi: 10.1002/(SICI)1099-1468(199609)17:5<471::AID-MDE789>3.0.CO;2-W

Hofling, C., Brotzman, E., Dalrymple, S., Graves, N., & Pierce, C. (1966). An experimental study in nurse-physician relationships. *The Journal of nervous and mental disease, 143*(2), 171.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related attitudes*. sage.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*. Sage Publications.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and Organizations: Software of the Mind, Third Edition*. Mcgraw-hill.

Holm, H., Flores, W. R., Nohlberg, M., & Ekstedt, M. (2014). An empirical investigation of the effect of target-related information in phishing attacks. In *2014 ieee 18th international enterprise distributed object computing conference workshops and demonstrations* (p. 357-363). doi: 10.1109/EDOCW.2014.59

Hong, J. H. (2012). The state of phishing attacks. *Commun. ACM, 55*(1), 74–81. doi: 10.1145/2063176.2063197

Hong, K., Kelley, C. M., Tembe, R., Murphy-Hill, E., & Mayhorn, C. B. (2013). Keeping up with the joneses: Assessing phishing susceptibility in an email task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*(1), 1012-1016. doi: 10.1177/1541931213571226

Huber, M., Kowalski, S., Nohlberg, M., & Tjoa, S. (2009). Towards automating social engineering using social networking sites. In *Computational science and engineering, 2009. cse '09. international conference on* (Vol. 3, p. 117-124). Vancouver, BC, Canada: IEEE. doi: 10.1109/CSE.2009.205

Humphris, G. M., Duncalf, M., Holt, D., & Field, E. (1999). The experimental evaluation of an oral cancer information leaflet. *Oral Oncology, 35*(6), 575–582. doi: 10.1016/S1368-8375(99)00040-8

Humphris, G. M., Ireland, R. S., & Field, E. A. (2001). Randomised trial of the psychological effect of information about oral cancer in primary care settings. *Oral Oncology, 37*(7), 548–552. doi: 10.1016/S1368-8375(01)00017-3

ISACA. (2015). *State of cybersecurity: Implications for 2015* (Tech. Rep.). ISACA and RSA Conference.

Jafari Kelarijani, S. E., Heidarian, A. R., Jamshidi, R., & Khorshidi, M. (2014). Length of service and commitment of nurses in hospitals of social security organization (sso) in tehran. *Caspian Journal of Internal Medicine, 5*(2), 94–98.

Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Commun. ACM*, *50*(10), 94–100. doi: 10.1145/1290958.1290968

Janczewski, L., & Colarik, A. (2008). *Cyber warfare and cyber terrorism*. Hershey, PA: Information Science Reference.

Junger, M., Montoya, L., & Overink, F.-J. (2017). Priming and warnings are not effective to prevent social engineering attacks. *Computers in Human Behavior*, *66*, 75 - 87. doi: 10.1016/j.chb.2016.09.012

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kaspersky. (2015). *Carbanak apt - the great bank robbery* (Tech. Rep.). Kaspersky Lab.

Kearney, W., & Kruger, H. (2014). Considering the influence of human trust in practical social engineering exercises. In *2014 information security for south africa* (p. 1-6). doi: 10.1109/ISSA.2014.6950509

Khand, P. (2009). System level security modeling using attack trees. In *Computer, control and communication, 2009. ic4 2009. 2nd international conference on* (p. 1-6). doi: 10.1109/IC4.2009.4909245

Kordy, B., Mauw, S., Radomirović, S., & Schweitzer, P. (2011). Foundations of Attack—Defense trees. In P. Degano, S. Etalle, & J. Guttman (Eds.), *Formal aspects of security and trust* (Vol. 6561, p. 80-95). Springer Berlin Heidelberg. doi: 10.1007/978-3-642-19751-2_6

Kordy, B., Piètre-Cambacédès, L., & Schweitzer, P. (2014). Dag-based attack and defense modeling: Don't miss the forest for the attack trees. *Computer Science Review*, *13-14*, 1–38. doi: 10.1016/j.cosrev.2014.07.001

Krawczyk, A., Lau, E., Perez, S., Delisle, V., Amsel, R., & Rosberger, Z. (2012). How to inform: comparing written and video education interventions to increase human papillomavirus knowledge and vaccination intentions in young adults. *Journal of American college health : J of ACH*, *60*(4), 316–22. doi: 10.1080/07448481.2011.615355

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2010). Teaching johnny not to fall for phish. *ACM Trans. Internet Technol.*, *10*(2), 7:1–7:31. doi: 10.1145/1754393.1754396

Lancaster, T., & Stead, L. F. (2005). Self-help interventions for smoking cessation. *Cochrane Database of Systematic Reviews*, *3*(3), CD001118. doi: 10.1002/14651858.CD001118

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, *33*(1), 159–174. doi: 10.2307/2529310

Lastdrager, E. E. (2014). Achieving a consensual definition of phishing based on a systematic review of the literature. *Crime Science*, *3*(1), 1–10. doi: 10.1186/s40163-014-0009-y

Lavorgna, A. (2014). Wildlife trafficking in the internet age. *Crime Science*, *3*(1), 5. doi: 10.1186/s40163-014-0005-2

Lee, C., Pillutla, M., & Law, K. S. (2000). Power-distance, gender and organizational justice. *Journal of Management*, *26*(4), 685-704. doi: 10.1177/014920630002600405

Lee, J., & Soberon-Ferrer, H. (1997). Consumer vulnerability to fraud: Influencing factors. *Journal of Consumer Affairs*, *31*(1), 70–89. doi: 10.1111/j.1745-6606.1997.tb00827.x

Lefkowitz, M., Blake, R. R., & Mouton, J. S. (1955). Status factors in pedestrian violation of traffic signals. *The Journal of Abnormal and Social Psychology*, *51*(3), 704–706. doi: 10.1037/h0042000

Lewis, A. (2012). *The cambridge handbook of psychology and economic behaviour*. Cambridge: Cambridge University Press.

Lien, N. H. (2001). Elaboration likelihood model in consumer research: A review. *Proceedings of the National Science Council*, *11*(4), 301–310.

Luo, R. X., Brody, R., Seazzu, A., & Burd, S. (2011). Social engineering: The neglected human factor for information security management. *Information Resources Management Journal*, *24*(3), 1-8. doi: 10.4018/irmj.2011070101

Lv, W., & Li, W. (2011). Space based information system security risk evaluation based on improved attack trees. In *Multimedia information networking and security (mines), 2011 third international conference on* (p. 480-483). doi: 10.1109/MINES.2011.94

Madden, C. (2006). Undergraduate nursing students' acquisition and retention of CPR knowledge and skills. *Nurse Education Today*, *26*(3), 218 - 227. doi: 10.1016/j.nedt.2005.10.003

Mann, I. (2008). *Hacking the human: Social engineering techniques and security countermeasures*. Aldershot: Gower.

Marconato, G. V., Kaaniche, M., & Nicomette, V. (2012). A Vulnerability Life Cycle-Based Security Modeling and Evaluation Approach. *The Computer Journal*, *56*(4), 422–439. doi: 10.1093/comjnl/bxs112

Mauw, S., & Oostdijk, M. (2006). Foundations of attack trees. In D. Won & S. Kim (Eds.), *Information security and cryptology - icisc 2005* (Vol. 3935, p. 186-198). Springer Berlin Heidelberg. doi: 10.1007/11734727_17

McCornack, S. A., & Parks, M. R. (1986). Deception detection and relationship development: The other side of trust. *Annals of the International Communication Association, 9*(1), 377-389. doi: 10.1080/23808985.1986.11678616

McCusker, R. (2006). Transnational organised cyber crime: Distinguishing threat from reality. *Crime, Law and Social Change, 46*(4-5), 257-273.

Meyerowitz, B. E., & Chaiken, S. (1987). The effect of message framing on breast self-examination attitudes, intentions, and behavior. *Journal of Personality and Social Psychology, 52*(3), 500 - 510. doi: 10.1037/0022-3514.52.3.500

Miles, J., & Shevlin, M. (2001). *Applying Regression and Correlation: A Guide for Students and Researchers*. SAGE Publications.

Milgram, S. (1963). Behavioral Study of obedience. *The Journal of Abnormal and Social Psychology, 67*(4), 371–378. doi: 10.1037/h0040525

Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations, 18*(1), 57-76. doi: 10.1177/001872676501800105

Milgram, S. (1974). *Obedience to authority: an experimental view*. New York: Harper & Row.

Mitnick, K., & Simon, W. (2002). *The art of deception: Controlling the human element of security*. New York: Wiley.

Mitnick, K., Simon, W., & Wozniak, S. (2011). *Ghost in the wires: My adventures as the world's most wanted hacker*. New York: Little, Brown.

Mobley, W. H., Horner, S. O., & Hollingsworth, A. T. (1978). An evaluation of precursors of hospital employee turnover. *Journal of Applied psychology, 63*(4), 408.

Montoya, L., Junger, M., & Hartel, P. (2013). How "digital" is traditional crime? In *Intelligence and security informatics conference (eisic), 2013 european* (p. 31-37). doi: 10.1109/EISIC.2013.12

Mouton, F., Malan, M. M., Kimppa, K. K., & Venter, H. (2015). Necessity for ethics in social engineering research. *Computers & Security, 55*, 114 - 127. doi: 10.1016/j.cose.2015.09.001

Murgor, T. K. (2013). A comparison of technical and vocational acquired skills differences based on gender in tvet institutions, Uasin Gishu county, Kenya. *Journal of Education and Practice, 4*(22), 181–187.

Muscanell, N. L., Guadagno, R. E., & Murphy, S. (2014). Weapons of influence misused: A social influence analysis of why people fall prey to internet scams. *Social and Personality Psychology Compass, 8*(7), 388–396. doi: 10.1111/spc3.12115

Neve, R., & van der Hulst, R. (2008). *High-tech crime: Inventarisatie van literatuur over soorten criminaliteit en hun daders* (Tech. Rep. No. 978-90-5454-998-7). WODC.

O'Keefe, D. J., & Hale, S. L. (2001). An odds-ratio-based meta-analysis of research on the door-in-the-face influence strategy. *Communication Reports, 14*(1), 31–38. doi: 10.1080/08934210109367734

Orgill, G. L., Romney, G. W., Bailey, M. G., & Orgill, P. M. (2004). The urgency for effective user privacy-education to counter social engineering attacks on secure computer systems. In *Proceedings of the 5th conference on information technology education* (pp. 177–181). New York, NY, USA: ACM. doi: 10.1145/1029533.1029577

Oshagbemi, T. (2000). Is length of service related to the level of job satisfaction? *International Journal of Social Economics, 27*(3), 213-226. doi: 10.1108/03068290010286546

Pallant, J. (2010). *Spss Survival Manual: A step by step guide to data analysis using SPSS*. McGraw-Hill Education.

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The role of individual differences in the accuracy of confidence judgments. *The Journal of General Psychology, 129*(3), 257-299. doi: 10.1080/00221300209602099

Parsons, K., McCormac, A., Pattinson, M., Butavicius, M., & Jerram, C. (2013). Security and privacy protection in information processing systems: 28th ifip tc 11 international conference, sec 2013, auckland, new zealand, july 8-10, 2013. proceedings. In L. J. Janczewski, H. B. Wolfe, & S. Shenoi (Eds.), (pp. 366–378). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-39218-4_27

Pascual, A., & Guéguen, N. (2005). Foot-in-the-door and door-in-the-face: A comparative meta-analytic study 1. *Psychological Reports, 96*(1), 122–128. doi: 10.2466/pr0.96.1.122-128

Payne, J. G. (2010). The bradley effect: Mediated reality of race and politics in the 2008 us presidential election. *American Behavioral Scientist, 54*(4), 417–435. doi: 10.1177/0002764210381713

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology, 49*(12), 1373–1379. doi: 10.1016/S0895-4356(96)00236-3

Petrova, P. K., Cialdini, R. B., & Sills, S. J. (2007). Consistency-based compliance across cultures. *Journal of Experimental Social Psychology, 43*(1), 104 - 111. doi: 10.1016/j.jesp.2005.04.002

Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and Persuasion–classic and Contemporary Approaches*. W.C. Brown Company Publishers.

Petty, R. E., & Cacioppo, J. T. (1984). Source factors and the elaboration likelihood model of persuasion. *Advances in Consumer Research*, *11*(1), 668–672.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, *19*(C), 123-205. doi: 10.1016/S0065-2601(08)60214-2

Pfeffer, J. (1985). Organizational demography: Implications for management. *California Management Review*, *28*(1), 67 - 81. doi: 10.2307/41165170

Poulsen, K. (2011). *Kingpin: How one hacker took over the billion-dollar cybercrime underground*. New York: Crown/Archetype.

Prehn-Kristensen, A., Göder, R., Fischer, J., Wilhelm, I., Seeck-Hirschner, M., Aldenhoff, J., & Baving, L. (2011). Reduced sleep-associated consolidation of declarative memory in attention-deficit/hyperactivity disorder. *Sleep Medicine*, *12*(7), 672 - 679. doi: 10.1016/j.sleep.2010.10.010

Rader, E., Wash, R., & Brooks, B. (2012). Stories as informal lessons about security. In *Proceedings of the eighth symposium on usable privacy and security* (pp. 6:1–6:17). New York, NY, USA: ACM. doi: 10.1145/2335356.2335364

Reason, J. (1990). The contribution of latent human failures to the breakdown of complex systems. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *327*(1241), 475–484. doi: 10.1098/rstb.1990.0090

Rhee, H.-S., Kim, C., & Ryu, Y. U. (2009). Self-efficacy in information security: Its influence on end users' information security practice behavior. *Computers & Security*, *28*(8), 816 - 826. doi: 10.1016/j.cose.2009.05.008

Riley, W. B., & Chow, V. K. (1992). Asset allocation and individual risk aversion. *Financial Analysts Journal*, *48*(6), 32-37. doi: 10.2469/faj.v48.n6.32

Robb, K. A., Miles, A., Campbell, J., Evans, P., & Wardle, J. (2006). Can cancer risk information raise awareness without increasing anxiety? A randomized trial. *Preventive Medicine*, *43*(3), 187–190. doi: 10.1016/j.ypmed.2006.04.015

Rogers, R. W. (1975). A Protection Motivation Theory of Fear Appeals and Attitude Change1. *The Journal of Psychology*, *91*(1), 93–114. doi: 10.1080/00223980.1975.9915803

Rosenstock, I. M. (1974). Historical origins of the health belief model. *Health Education & Behavior*, *2*(4), 328-335. doi: 10.1177/109019817400200403

Ross, M., Grossmann, I., & Schryer, E. (2014). Contrary to psychological and popular opinion, there is no compelling evidence that older adults are disproportionately

victimized by consumer fraud. *Perspectives on Psychological Science, 9*(4), 427-442. doi: 10.1177/1745691614535935

Rowe, E., Akman, T., Smith, R. G., & Tomison, A. M. (2012). Organised crime and public sector corruption: A crime scripts analysis of tactical displacement risks. *Trends and Issues in Crime and Criminal Justice*(444), 1.

Roy, A., Kim, D. S., & Trivedi, K. S. (2010). Cyber security analysis using attack countermeasure trees. In *Proceedings of the sixth annual workshop on cyber security and information intelligence research* (pp. 28:1–28:4). New York, NY, USA: ACM. doi: 10.1145/1852666.1852698

Sarker, S. J., Crossman, A., & Chinmeteepituck, P. (2003). The relationships of age and length of service with job satisfaction: an examination of hotel employees in thailand. *Journal of Managerial Psychology, 18*(7), 745-758. doi: 10.1108/02683940310502421

Schmidt, S. R. (1994). Effects of humor on sentence memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 953. doi: 10.1037/0278-7393.20.4.953

Schneier, B. (1999). Attack trees. *Dr. Dobb's journal, 24*(12), 21–29.

Schoenmakers, Y., de Vries Robbé, E., & van Wijk, A. P. (2009). *Gouden bergen* (Vol. 48). Stapel & De Koning.

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J. (2010). Who falls for phish?: A demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 373–382). New York, NY, USA: ACM. doi: 10.1145/1753326.1753383

Shenhav, Y., & Haberfeld, Y. (1992). Organizational demography and inequality. *Social Forces, 71*(1), 123–143. doi: 10.2307/2579969

Shim, S. M., Seo, S. H., Lee, Y., Moon, G. I., Kim, M. S., & Park, J. H. (2011). Consumers'knowledge and safety perceptions of food additives: Evaluation on the effectiveness of transmitting information on preservatives. *Food Control, 22*(7), 1054–1060. doi: 10.1016/j.foodcont.2011.01.001

Smith, E., & Kosslyn, S. (2008). *Cognitive psychology: Mind and brain.* Pearson Prentice Hall.

Smith, P. B., & Schwartz, S. H. (1997). Values. In J. W. Berry, M. H. Segall, & C. Kagitçibasi (Eds.), *Handbook of cross-cultural psychology: Social behavior and applications* (Vol. 3, p. 77-118).

Stewart, D. W., & Martin, I. M. (1994). Intended and unintended consequences of warning messages: A review and synthesis of empirical research. *Journal of Public Policy & Marketing*, 1–19.

Stubbings, S., Robb, K., Waller, J., Ramirez, A., Austoker, J., Macleod, U., … Wardle, J. (2000). Development of a measurement tool to assess public awareness of cancer. *Br J Cancer*, *101*(S2), S13–S17. doi: 10.1038/sj.bjc.6605385

Sutton, R. M., Niles, D., Meaney, P. A., Aplenc, R., French, B., Abella, B. S., … Nadkarni, V. (2011). Low-dose, high-frequency cpr training improves skill retention of in-hospital pediatric providers. *Pediatrics*, *128*(1), e145-e151. doi: 10.1542/peds.2010-2105

Szymanski, D. (2001). Modality and offering effects in sales presentations for a good versus a service. *Journal of the Academy of Marketing Science*, *29*(2), 179-189. doi: 10.1177/03079459994542

Tanford, S., & Penrod, S. (1984). Social influence model: A formal integration of research on majority and minority influence processes. *Psychological Bulletin*, *95*(2), 189 - 225. doi: 10.1037/0033-2909.95.2.189

Tembe, R., Hong, K. W., Murphy-Hill, E., Mayhorn, C. B., & Kelley, C. M. (2013). American and indian conceptualizations of phishing. In *2013 third workshop on socio-technical aspects in security and trust* (p. 37-45). doi: 10.1109/STAST.2013.10

Thompson, L., & Chainey, S. (2011). Profiling illegal waste activity: Using crime scripts as a data collection and analytical strategy. *European Journal on Criminal Policy and Research*, *17*(3), 179-201. doi: 10.1007/s10610-011-9146-y

Titus, R. M., Heinzelmann, F., & Boyle, J. M. (1995). Victimization of persons by fraud. *Crime & Delinquency*, *41*(1), 54-72. doi: 10.1177/0011128795041001004

Tremblay, P., Talon, B., & Hurley, D. (2001). Body switching and related adaptations in the resale of stolen vehicles. script elaborations and aggregate crime learning curves. *British Journal of Criminology*, *41*(4), 561–579. doi: 10.1093/bjc/41.4.561

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131.

Twitchell, D. P. (2009). Social engineering and its countermeasures. In *Handbook of research on social and organizational liabilities in information security* (pp. 228–242). Hershey, PA: IGI-Global. doi: 10.4018/978-1-60566-132-2.ch014

Van Dijk, J., & Nijenhuis, N. (1979). Ja zeggen, nee doen? een onderzoek naar de overeenkomst tussen verbale attitudes en feitelijk gedrag bij angstgevoelens tav criminaliteit. *Tijdschrift voor criminologie*, *21*(6), 257-273.

Van Wyk, J., & Benson, M. L. (1997). Fraud victimization: Risky business or just bad luck? *American Journal of Criminal Justice*, *21*(2), 163–179. doi: 10.1007/BF02887448

Vesely, W. E., Goldberg, F. F., Roberts, N. H., & Haasl, D. F. (1981). *Fault Tree Handbook*. Washington, DC: U.S. Nuclear Regulatory Commission.

Vrij, A., Granhag, P. A., & Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest, 11*(3), 89-121. doi: 10.1177/1529100610390861

Waldrop, M. (2016). How to hack the hackers: The human side of cybercrime. *Nature, 533*, 164.

Wang, J., Whitley, J. N., Phan, R. C.-W., & Parish, D. J. (2011). Unified parametrizable attack tree. *International Journal for Information Security Research, 1*(1), 20–26. doi: 10.20533/ijisr.2042.4639.2011.0003

Waters, L., Roach, D., & Waters, C. W. (1976). Estimates of future tenure, satisfaction, and biographical variables as predictors of termination. *Personnel Psychology, 29*(1), 57–60. doi: 10.1111/j.1744-6570.1976.tb00401.x

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of personality and social psychology, 39*(5), 806. doi: 10.1037/0022-3514.39.5.806

Whittingham, R. (2004). *The blame machine: Why human error causes accidents*. London: Taylor & Francis.

Winkler, I. S., & Dealy, B. (1995). Information security technology?...don't rely on it: A case study in social engineering. In *Proceedings of the 5th conference on usenix unix security symposium - volume 5* (pp. 1–1). Berkeley, CA, USA: USENIX Association.

Winstok, Z. (2013). Partner violence as a rational choice. In *Partner violence* (p. 47-60). Springer New York. doi: 10.1007/978-1-4614-4568-5_3

Wright, R., Jensen, M., Thatcher, J., Dinger, M., & Marett, K. (2014). Influence techniques in phishing attacks: An examination of vulnerability and resistance. *Information Systems Research, 25*(2), 385-400. doi: 10.1287/isre.2014.0522

Yager, R. R. (2006). Owa trees and their role in security modeling using attack trees. *Information Sciences, 176*(20), 2933 - 2959. doi: 10.1016/j.ins.2005.08.004

Yoo, B., Donthu, N., & Lenartowicz, T. (2011). Measuring hofstede's five dimensions of cultural values at the individual level: Development and validation of cvscale. *Journal of International Consumer Marketing, 23*(3-4), 193-210. doi: 10.1080/08961530.2011.578059

Zhao, B., & Olivera, F. (2006). Error reporting in organizations. *Academy of Management Review, 31*(4), 1012-1030. doi: 10.5465/AMR.2006.22528167

Zonouz, S., Khurana, H., Sanders, W., & Yardley, T. (2009). Rre: A game-theoretic intrusion response and recovery engine. In *Dependable systems networks, 2009. dsn '09. ieee/ifip international conference on* (p. 439-448). doi: 10.1109/DSN.2009.5270307

Other Sources

Arthur, C. (2010). *Virus phone scam being run from call centres in india* [Newspaper Article]. Retrieved from http://www.theguardian.com/world/2010/jul/18/phone-scam-india-call-centres

Chmielewski, D. (2015). *Balabit csi report.* Retrieved 05-aug-2016, from https://www.balabit.com/news/press/social-engineering-leads-the-top-10-list-of-most-popular-hacking-methods-balabit-survey-results-from-black-hat-usa

Cross, J. (2011). *Social Engineering is Often Overlooked.* Retrieved from http://www.immense.net/social-engineering-planning/

European Social Survey Education Netu. (2013). *Adding interaction terms to ols regression models.* Retrieved 27-jun-2016, from http://essedunet.nsd.uib.no/cms/topics/multilevel/ch1/5.html

FNV. (2016). *Onzekerheid, werkdruk en veiligheidsrisico's - een verkennend onderzoek naar de veiligheids- beleving van passagemedewerkers op schiphol* (No. 61604). Retrieved from https://www.fnv.nl/over-fnv/pers/persberichten/persarchief/2016/november/FNVonderzoek-bagage-en-incheck-Schiphol-Onveilig-door-te-weinig-vaste-mensen-/

Hight, S. D. (2005). *The importance of a security, education, training and awareness program.* Retrieved 21-oct-2015, from http://www.infosecwriters.com/Papers/SHight_SETA.pdf

Kennedy, D. (2011). *There's something "human" to social engineering.* Retrieved from http://magazine.thehackernews.com/article-1.html

King, R. (2011). *Emc's rsa security breach may cost bank customers $100 million.* Retrieved 05-aug-2016, from http://www.bloomberg.com/news/articles/2011-06-08/emc-s-rsa-security-breach-may-cost-bank-customers-100-million

McGoogan, C. (2016). *Whatsapp users targeted with £100 sainsbury's scam - how to protect yourself.* Retrieved 22-nov-2016, from http://www.telegraph.co.uk/technology/2016/10/25/whatsapp-users-targeted-with-100-sainsburys-scam---how-to-protec/

Proofpoint. (2016). *The human factor 2016.* Retrieved from https://www.proofpoint.com/us/human-factor-2016

PWC. (2014). *Us cybercrime: Rising risks, reduced readiness.* Retrieved 09-nov-2016, from http://www.pwc.com/us/en/increasing-it-effectiveness/publications/assets/2014-us-state-of-cybercrime.pdf

Rouse, M. (2006). *Definition Social Engineering.* TechTarget. Retrieved 23-oct-2013, from http://www.searchsecurity.techtarget.com/definition/social-engineering

Savage, M. (2012). *The rsa breach: One year later.* Retrieved 04-sep-2016, from http://searchsecurity.techtarget.com/magazineContent/The-RSA-breach-One-year-later

Schellevis, J. (2011). *Grote amerikaanse bedrijven vatbaar voor social engineering.* Retrieved 27-dec-2013, from http://tweakers.net/nieuws/77755/grote-amerikaanse-bedrijven-vatbaar-voor-social-engineering.html

Schneier, B. (2000). *Secrets & lies: Digital security in a networked world* (1st ed.). New York, NY, USA: John Wiley & Sons, Inc.

Schneier, B. (2005). *Flaw in Winkhaus Blue Chip Lock.* Retrieved 12-nov-2013, from https://www.schneier.com/blog/archives/2005/03/flaw_in_winkhau.html

Sparshott, M. (2014). *The psychology of phishing.* Retrieved 26-may-2016, from https://www.helpnetsecurity.com/2014/07/23/the-psychology-of-phishing/

The Federal Bureau of Investigation. (2013). *Internet Social Networking Risks* (Vol. 2013) (No. 4 October). U.S. Department of Justice. Retrieved 23-oct-2013, from http://www.fbi.gov/about-us/investigate/counterintelligence/internet-social-networking-risks

The SANS Institute. (2012). *Cyber security newsletter (social engineering - hacking your mind).* Retrieved 23-aug-2016, from https://www.uab.edu/it/home/images/Module02-SocialEngineering-Newsletter.pdf

The TREsPASS Project, D1.3.3. (2015). *Dynamic features of socio-technical security models.* (Deliverable D1.3.3)

Thomson, I. (2017). *74 countries hit by nsa-powered wannacrypt ransomware backdoor: Emergency fixes emitted by microsoft for winxp+.* Retrieved 25-may-2017, from https://www.theregister.co.uk/2017/05/13/wannacrypt_ransomware_worm/

Usa v mitnick. (1996). (Indictment, CR 96-881, 145 F.3d 1342)

Usa v mitnick. (1998). (No. 97-50365)

Acronyms

ADTree Attack Defence Tree

CPR Cardiopulmonary Resuscitation

DitF Door-in-the-Face

ELM Elaboration Likelihood Model

EPV Events Per Variable

F2F Face-to-Face

FitD Foot-in-the-Door

FM Facility Management

HRM Human Resource Management

IDS Intrusion Detection Systems

IRB Institutional Review Board

IT Information Technology

MO *Modus Operandi*

PAT Problem Analysis Triangle

PDI Power Distance

PII Personally Identifiable Information

RAT Routine Activity Theory

RCT Rational Choice Theory

SE Social Engineering

SMP Secure Messaging Portal

USD United States Dollar

UT University of Twente

VIF Variance Inflation Factor

YoS Years of Service

Colophon

This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Arno Pro, designed by Robert Slimbach in the style of book types from the Aldine Press in Venice, and issued by Adobe in 2007. A template, which can be used to format a PhD thesis with this look and feel, has been released under the permissive MIT (X11) license, and can be found online at github.com/suchow/ or from the author at suchow@post.harvard.edu.