Investigations into Social Engineering Evidence for Security Research

Sven Uebelacker

TUHH
Hamburg University of Technology

SVA

AWS Literatur
Der Verlag von Alles wird schön e.V.

Imprint

Publisher: awsLiteratur — Der Verlag des Kulturvereins Alles wird schön e.V.
https://alles-wird-schoen-e-v.de/
Second printing, May 2024; printed in Germany
ISBN: 987-3-947051-31-1
DOI (this document): 10.15480/882.9595
DOI (doctoral thesis): 10.15480/882.4933
ORCID: 0000-0001-9228-8248
https://research.uebelacker.net/

License

Funding

European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 318003 and UK's Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/M020320/1 (see also Acknowledgement in Chapter Preface)

Tools and Template

LaTeX book style "The Legrand Orange Book" version 2.4 from Mathias Legrand with modifications by Vel under Creative Commons license BY-NC-SA 3.0. Further modifications were made by Sven Uebelacker and Sofia Morais.

Media

Front matter picture "The Fisherman" taken by Roland W. Reed in 1908 (public domain). A Native American of the Ojibwe tribe went spear fishing. Image was adjusted by Sofia Morais. Source: https://commons.wikimedia.org/wiki/File:The_Fisherman,_Roland_W._Reed,_1908_Img018.jpg

Back matter picture taken by Alfred Hart between 1865–1869 (public domain). A Native American looked down at the Transcontinental Railroad near Sacramento, California. Image was adjusted by Sofia Morais. Source: Library of Congress, https://www.loc.gov/resource/stereo.1s00629/

Chapter images are derived images from the two above, created by Sofia Morais.

Investigations into Social Engineering Evidence for Security Research

Vom Promotionsausschuss der Technischen Universität Hamburg

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

Sven Übelacker

aus

Delmenhorst

2023

1. Gutachter: Prof. Dr. Dieter Gollmann
2. Gutachter: Prof. Dr.-Ing. Felix Freiling

Datum der mündlichen Prüfung: 24.06.2022

Abstract

When technical protection mechanisms are too inflexible and security decisions are passed to users, users become part of the attack surface of a socio-technical system. This thesis contributes to the science of security regarding Social Engineering (SE), where users are the core enabler of successful attacks. Various sources were consulted to identify, analyse, and understand SE, enabling the development of a multidisciplinary knowledge base and appropriate safeguards. A suitable definition comprising five SE indicators was developed to examine anecdotes whether they express SE. Court documents (phishing), Lego modelling (cloud), and a novel SE poetry slam concept served as sources for in-depth analyses.

Abstrakt

Wo technische Schutzmaßnahmen zu unflexibel sind und Sicherheitsentscheidungen den Menschen überlassen werden, werden Menschen Teil der sozio-technischen Angriffsfläche. Es wird zur Science of Security bzgl. Social Engineering (SE) beigetragen. Verschiedene Quellen wurden herangezogen, um SE zu identifizieren, zu analysieren und zu verstehen. Dies ermöglicht die Entwicklung einer multidisziplinären Wissensdatenbank und geeigneter Schutzmaßnahmen. Eine Definition inkl. fünf SE-Indikatoren wurde entwickelt, mithilfe derer SE identifiziert werden kann. Gerichtsurteile (Phishing), Legomodellierung (Cloud) sowie ein neuartiges SE Poetry Slam-Konzept dienten für tiefergreifende Untersuchungen.

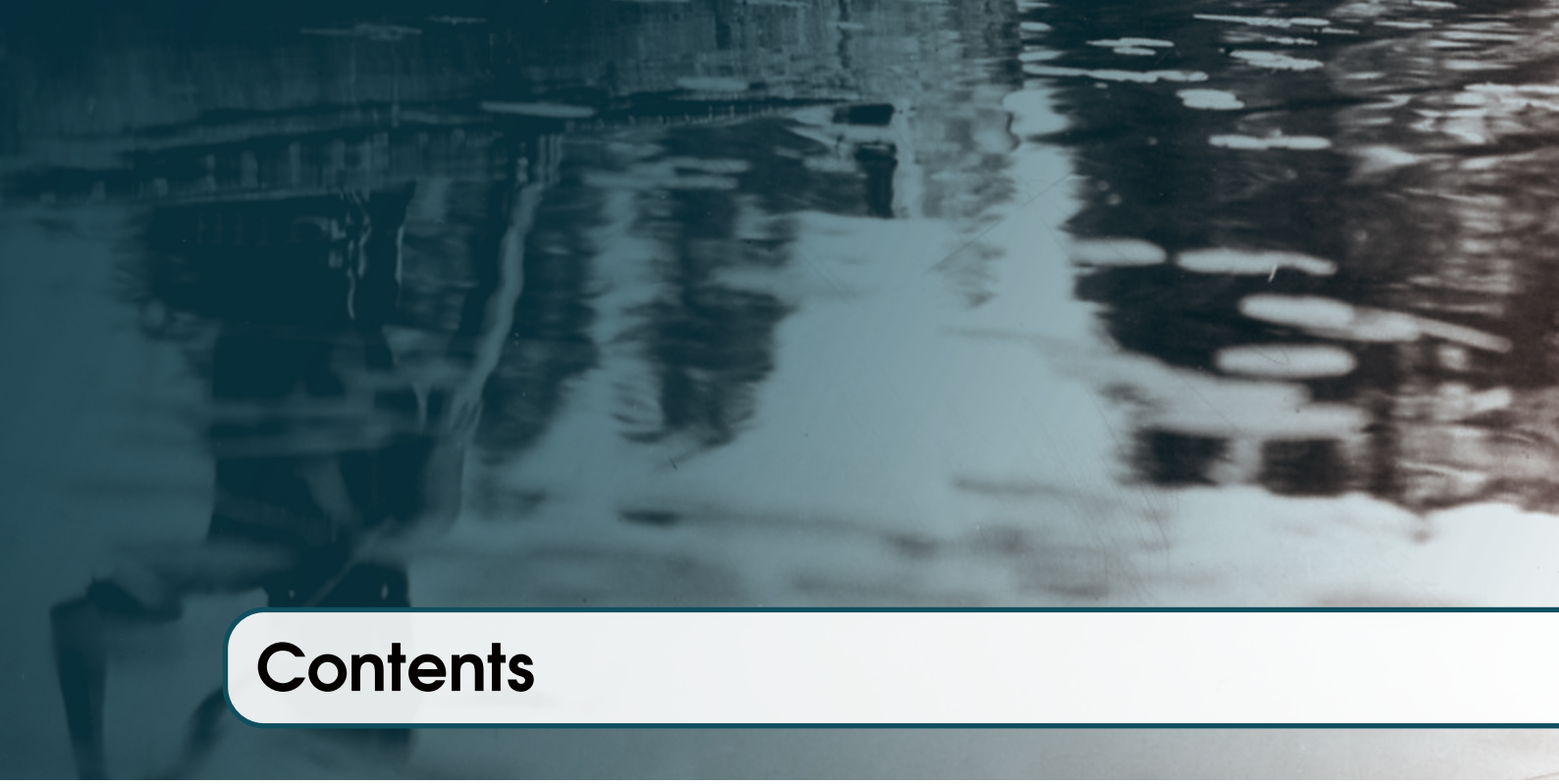
Summary

Human interaction with security mechanisms and its resulting outcome is omnipresent in computing practice. End-users become then part of the attack surface of a socio-technical system if forced to make security decisions. Creating and executing attacks against that 'human firewall' is often called *Social Engineering (SE)*. SE research is interdisciplinary research. Besides computer science, it reaches into behavioural economics, humanities, psychology, and cultural sciences and can refer to insights from marketing and sales. This thesis combines insights about persuasion principles, personality traits, and cultural values of national or organisational cultures to understand how they may affect the targeted persons' susceptibility. Narrative psychology serves as a means of transporting SE situations to the target audience in form of anecdotes.

Researchers need an empirical basis for identifying, analysing, and understanding SE to enable the development of appropriate safeguards, such as effective trainings. Due to security and privacy concerns, data of real life SE events for quantitative analyses are hard to gather. SE experiments contain an additional ethical aspect when subjects are exposed to deceptive SE techniques that try to mimic real life malicious intent.

In this thesis three sources of SE evidence are examined qualitatively: *court documents* about phishing express real events, are well documented for their original purpose, and are often publicly available. Court documents underlie an iterative truth finding process which can add further SE information. An uncommon SE attack was found where the attacker applied spear phishing to acquire carbon emission certificates. A *poetry slam* style event for ethical hacker gatherings was conceptualised and conducted. It lets ethical hackers tell their anecdotes about SE with optional pseudonymisation. Recorded sessions can be examined after the event to identify SE. In one anecdote attackers used a puppy as a distraction to enter company premises and access office computers. Although fictitious anecdotes may be presented, they offer insights influencing future experimental designs for further analyses. *Lego modelling* was applied in the FP7 EU research project TRE_{S}PASS. Participants of different professions can come up with an imagined attack based on a given setting. The Lego bricks offer an easy and quick start to dive into a haptic attack modelling as Lego bricks are well-known and playful to use. Such a Lego modelling session was reconducted, crafting attacks in a cloud scenario setting.

A suitable SE definition for a science of security, usable in computer science and various disciplines, was created that can foster a common, multidisciplinary knowledge base. This knowledge base lays the foundation stone for a science of security for SE. The SE definition comprises five SE indicators, applicable for the identification of SE in examined evidence. The SE indicators are independent from a changing threat landscape and served well to identify SE. Court documents on phishing cases became a useful source, complemented by contributions from ethical hacker conferences and modelling sessions. These SE insights contribute to the knowledge base upon which further research can be based, e.g., by designing ethical SE experiments in a more controlled environment or for longitudinal, multicultural studies.



Contents

Preface

The End is Nigh

Will it be more joy to bring this thesis to an end and move on to new pastures? Oscar Wilde wrote: "In this world there are only two tragedies. One is not getting what one wants, and the other is getting it."[1] Or was it the excitement in slipping into the researcher's role? If I would have known the exact scope and outcome of my research, what had been the point of doing it?[2] I was all agog with curiosity — and hoping in the end, not to have to leave immediately for Nepal with an unfinished thesis to live as a goat.[3] At some point in writing this thesis, I had to stop searching for new publications — even though my curiosity insisted on reading a lot more. I needed to be told the famous Ernest Hemingway quote "fuck literature"[4] and finish your thesis. Well, I tried and almost succeeded stopping to read more while prolonging its finalisation.

It all started when I realised and asked myself why matured and hitherto existing digital and physical security measures become almost useless when end-users are forced to make security decisions. I saw a lot of this during my previous work in computer centres. Why all the effort to optimise technical measures when a simple phishing e-mail or a USB stick with a malicious payload could bypass them? I was intrigued to investigate human factors in information security, especially the Social Engineering part. I commenced in asking why so many views on Social Engineering differ. To ground my research I needed

[1] Oscar Wilde: "Lady Windermere's Fan – A Play About a Good Woman" (1893)

[2] Pablo Picasso: "Si l'on sait exactement ce que l'on va faire, à quoi bon le faire?" ("If you know exactly what you are going to do, what is the point of doing it?")

[3] Rowan Atkinson in Blackadder (1987): "I am therefore leaving immediately for Nepal, where I intend to live as a goat."

[4] Letter (1924) to Ezra Pound; published in Ernest Hemingway: Selected Letters 1917–1961 (1981) edited by Carlos Baker, p. 113

a precise definition to also identify evidential data about the modus operandi of attacks. Understanding how Social Engineering attacks were conducted lead to the follow-up questions, why are some people more susceptible than others and what can we do about it.

Social Engineering research is interdisciplinary research in cybersecurity — non-exclusive for computer science. Such research requires cross-disciplinary expertise and a mutual understanding of terminologies and scientific methodologies as well as the interpretation of 'evidence' to gain insights. Consulting fellow researchers of non-computer science disciplines becomes an intrinsic requisite. One challenge was therefore to approach this interdisciplinary understanding what Social Engineering is about when interlocked with security research. I hope that my insights and conclusions might come useful for fellow researchers — whether to build upon or to refute them.

Acknowledgement

This thesis would not have been successful without the support of family, friends, and fellow researchers. Without Alexander I would not have visited Felix Freiling at the RWTH Aachen and, thanks to that, started to envision a doctorate; without Roland I would not have contacted Dieter Gollmann for joining the TRE$_S$PASS project as a doctoral candidate.

Heartfelt thanks go to Dieter Gollmann whose continuous support and extraordinary security knowledge provided me the crucial in-time reviews and necessary resources. It truly amazes me that he recalls names, papers, and conferences immediately from memory — even decades later. He offered me the perfect opportunity to participate in three Dagstuhl seminars where I was able to exchange ideas with renowned international researchers.

I am grateful that I had Florian as a sparring partner during the doctoral process. We both had our doctoral examinations just days apart. Memories, like writing together for a week in the Algarve, will remain. Thank you, Christiane and Gabriele, for your guidance as coaches and friends in the doctoral process. There are a lot of fellow researchers to mention, e.g., the partners of the TRE$_S$PASS project as well as my former colleagues from the Cambridge Cybercrime Centre. Thank you, Antonia, for your general reviews, Petra for perusing the legal part of my court documents research, and Jan-Willem for reviewing the human factors part. For the Social Engineering Poetry Slams, I was lucky to work together with Youssef, Anna, and kolAflash. And not to forget, kudos to my students. Last, but not least, without the support of Anika during the times of a challenging work-life-research balance, there would not have been existing this thesis.

Funding

This thesis received funding from the following grants.

2013–2016 Hamburg University of Technology, Germany.

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 318003 (TRE_SPASS – Technology-supported Risk Estimation by Predictive Assessment of Socio-technical Security). This publication reflects only the author's views and the European Union is not liable for any use that may be made of the information contained herein.

Thanks to the TRE_SPASS project, I was able to attend three Dagstuhl seminars (co-funded by the Leibniz Association and its funding partners). I am grateful that I was given this opportunity.

- December 2012, "Organizational Processes for Supporting Sustainable Security"
  https://www.dagstuhl.de/12501/
- November/December 2014, "Socio-Technical Security Metrics"
  https://www.dagstuhl.de/14491/
- November 2016, "Assessing ICT Security Risks in Socio-Technical Systems"
  https://www.dagstuhl.de/16461/

2016–2017 Cambridge Cybercrime Centre, University of Cambridge, UK.

Project "Interdisciplinary Centre for Finding, Understanding and Countering Crime in the Cloud" funded by UK's Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/M020320/1.

1. Introduction

Security in computer science and the security measures implemented often focus on *technological* controls with respect to the classic CIA triad (Confidentiality, Integrity, and Availability). However, Matt Bishop [Bis03] asserts that "the heart of any security system is people". Regarding security engineering Ross Anderson [And08] states that it "requires *cross-disciplinary* expertise, ranging from cryptography and computer security through hardware tamper-resistance and formal methods to a knowledge of economics, applied psychology, organizations and the law". For end-users, security is both a feeling as well as a reality, says Bruce Schneier [Sch08b]. Sometimes this feeling creates a 'security theatre' to feel more secure, but the reality tells a different story [Sch08b]. The security reality does not offer "any direct financial reward" [Bis03]. For instance, "it limits losses, but also requires the expenditure of resources that could be used elsewhere" [Bis03]. From an economic point of view, rejecting security advice is a rational decision because of a "poor cost-benefit tradeoff" [Her09]: direct costs of attacks are hidden while creating greater indirect costs.

A gap between the real and the perceived security exposure by an end-user may result in favour of palliative security measures [Sch08b]. The challenge remains to align the security theatre to the security reality by raising the feeling of security in the desired direction [Sch08b]. In an organisational context, employees may create their own security measures based on their understanding of security risks as a compromise to stay productive [KPS15]. This so-called "shadow security" [KPS15] expresses unofficial security workarounds from which an organisation can learn to improve and align existing security policies. In a private setting where no organisational security policy exists, end-users rely on their own or a friend's risk assessment when using security-relevant services, for instance, to determine whether an e-mail is trying to phish them for their e-banking credentials.

"Security goals are usually not primary, but secondary goals for the user" [Ben+15], such as sending an e-mail encrypted (primary goal: communication) and confidentiality (secondary goal). Often humans are referred to as the "weakest link" [SBW01; Sch00] in security and human intervention can usually bypass technological controls [Bis03]. However, designers of security mechanisms have to acknowledge that users are not the enemy [AS99]. Implemented security mechanisms must serve end-users (see user-centred design) to support them. In an ideal state, security mechanisms would be a built-in and not an add-on software feature, thus, opposing the "unfortunately pervasive 'penetrate-and-patch' approach to security" [VM02]. Regarding user interaction, these built-in security mechanisms should guide end-users to use any service securely without the need of making security decisions. End-users become part of the security system when interacting with security-relevant mechanisms. Again, in an ideal state, end-users would make the right security decisions if unavoidable or security systems would tolerate bad security decisions by minimising the impact. But security mechanisms in reality still confuse and force many end-users to perform security-relevant tasks, incapable of satisfying the security needs, see why Johnny can't encrypt using Pretty Good Privacy (PGP) in 1999 [WT99]. Still today Johnny's task has improved only slightly when facing security trade-offs, e.g., by encouraging opportunistic encryption with pretty Easy privacy[1] (p≡p). The combination of usability and security to usable security is one approach to address this issue [Ben+15]. A more realistic positive outcome of security measures is to reduce the dissatisfaction rather than achieving satisfaction of end-users [Ben+15].

To conclude, human interaction with security mechanisms and its resulting outcome is omnipresent in computing practice. It is crucial to understand and address human factors in cybersecurity regarding the growing ubiquity of digital services and devices that need human interaction — whether in an organisational or private setting. Security mechanisms must always be understood in the context of the tasks end-users are performing. As long as end-users are part of the security system and forced to perform security-related tasks — as 'basic' as memorising, recalling and entering passwords —, they become part of a socio-technical attack surface. Hence, the end-user as a human element is introduced, which requires completely different safeguards compared to technical ones. This human attack vector offers an attacker additional attack options such as tricking an end-user into performing actions benefiting the attacker. Creating and executing such attacks is often called Social Engineering (SE).

1.1 Social Engineering (SE)

Social Engineering (SE) is defined as the act of maliciously manipulating an unaware targeted person to comply with a request through an attacker's intentional communication using deceptive techniques (see Definition 3.17). Only the targeted person can make SE succeed. SE can bypass physical and digital security measures [Sch08b]. Phishing as one deceptive technique is well-known in computer science. The reaction of a targeted person determines whether the phishing attack was successful. A special publication on digital identities by the National Institute of Standards and Technology (NIST) stated on passwords: "keystroke logging, phishing, and social engineering attacks are equally

[1] https://www.pep.security/en/

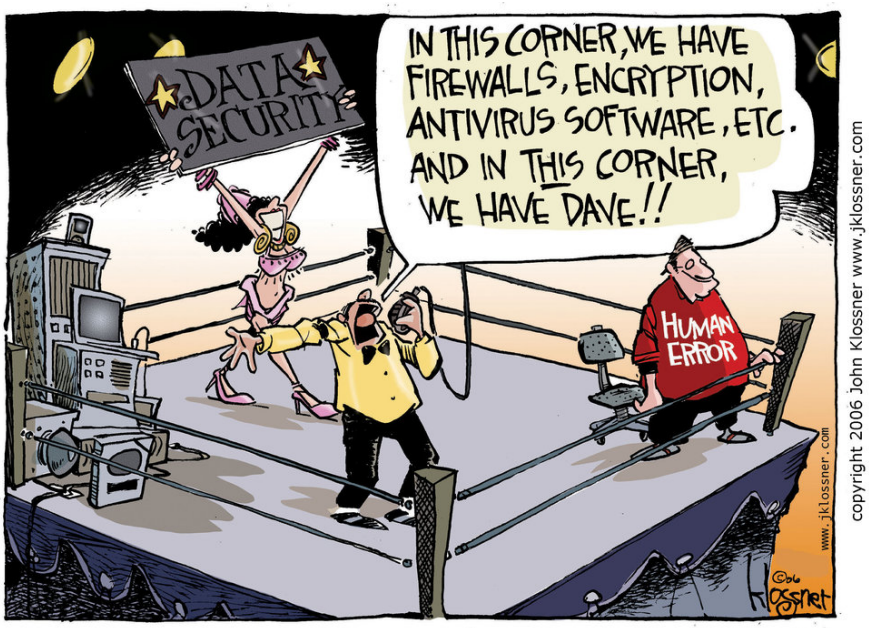


Figure 1.1: Human Nature: Data Security and Dave [Klo06]

effective on lengthy, complex passwords as simple ones." [GGF17, Appendix A: Strength of Memorized Secrets] That is, security measures like lengthy, complex passwords or multi-factor authentication help against typical brute-force attacks; against SE attacks they are often useless though. SE does not exist solely in e-mail communication and more general within information technology, although some SE definitions restrict it to require technology. Attacks with SE exist in the non-digital world and attacks without SE exist in the digital domain. SE comprises malicious attacks such as face-to-face grandparent scams without using any technology (German: Enkeltrick). Hence, prior to the digital age of well-known Nigerian prince scams (419 scam), SE was already performed: for instance, in previous centuries scams like the 'Spanish prisoner' (16^{th} century) were conducted mostly in person or the 'Lettre de Jérusalem' (18^{th} and 19^{th} century) circulated via postal services. They can be categorised as advance-fee scams and they still exist today. Back then, the attackers tried to convince targeted persons to send money to bribe prison guards to free a wealthy person who would reward the targeted person in return. With "technique propagation" [Sch00] sending such scam messages en masse through new mediums becomes easier, cheaper, faster, and/or more automatable for attackers, e.g., via facsimiles (inheritance scam, see Figure C.5.4), e-mail (General Data Protection Regulation (GDPR) scam, Figure 4.3), social networks or instant messaging. Nowadays, technology, such as automated scraping social networks for Personally Identifiable Information (PII), exists to craft personalised messages for attacks automatically, e.g., for spear phishing campaigns. That is, when in pure real-life situations attackers are able to spot valuable assets (jewelry), 'read' the targeted persons and react accordingly (con artists), technology-based scams involving PII are advancing in the same direction. Where SE overlaps with the cyberspace, it concerns cybersecurity research.

Information security started as a "mono-disciplinary direction" in computer science as Leukfeldt [Leu17] stated. Information security research became cyberspace research,

covering human factors and Socio-Technical Systems (STS)[2] [Leu17]. Without the insights from humanities or social sciences (criminologists, ethicists etc.), scientific achievements in cybersecurity concerning human factors would hardly be possible [Leu17]. Spring and Illari [SI18] stated, summarising Flechais et al. [FRS05] and Anderson and Moore [AM06], that "adequate security evaluations" are only possible if STS are examined holistically. SE occurs in STS. Anderson required "cross-disciplinary expertise" in security engineering [And08] as mentioned above. Schneier referred to the following disciplines that concern the psychology of security: behavioural economics, psychology of decision-making, risk perception, and neuroscience [Sch08b]. SE itself needs a human enabler to be able to succeed: the targeted person. That is, for instance, phishing research must not ignore the targeted persons. This core SE enabler together with security research makes it clear: SE research is interdisciplinary research in cybersecurity — non-exclusive for computer science. Such research requires a mutual understanding of terminologies and scientific methodologies as well as the interpretation of 'evidence' to gain insights. Consulting fellow researchers of non-computer science disciplines becomes an intrinsic requisite.

1.1.1 The Human Element

Human factors in other disciplines are observed and analysed over a longer period of time than the relatively new discipline computer science. The classical Milgram shock experiment [Mil65] in 1965 showed how subjects administered electric shocks to actors who pretended to be the test subjects. Some of the real subjects did not stop to even choose lethal shock levels in order to follow the experimenter's instructions. The Milgram shock experiment is an example of perceived authority and the resulting obedience. On a level of cognitive biases or psychological triggers, effects exist that influence the human thinking, perception, decision making or behaviour. For instance, the Pavlovian conditioning is well-known. While the bell in Pavlov's experiment triggered the dog's anticipatory salivation, the simple word 'Pavlov' may trigger people who know the experiment into thinking of a dog[3]. That is, the association of words or the wording of a sentence may recall memories or trigger thinking processes that may result in corresponding actions or belief systems. A corresponding action could be that if a good is presented as scarce, it may nudge a person into trying to acquire it and avoid missing out (loss aversion, Section 2.3.1). The use and creation of specific terms may result in desired associations influencing the belief system, e.g., 'Großer Lauschangriff' (English: big surveillance attack) could be connotated negatively and 'Datenreichtum'[4] (English: data riches/richness; could also relate to 'big data' which was not translated to German) as antonym for 'data minimisation' (Art. 5 GDPR on personal data) positively. Words are just one way that may influence recipients.

[2] The term STS originates from social sciences and is nowadays commonly used in risk management and information security. It describes a system where social elements (people) interact with technical ones (machines, computers). In a cyber security context, the domains social/organisational, technical/physical, and digital/virtual are discussed.

[3] "the true genius of classical pavlovian conditioning is that every time i hear 'pavlov' i automatically think of a dog" [Sun18]

[4] German BigBrotherAward 2016 in the category 'Neusprech' [HB16]. Neusprech translates to 'newspeak' created by George Orwell for his dystopian book 'NINETEEN EIGHTY-FOUR' [Orw95] about a fictional totalitarian superstate.

In general, each of the three levels of unique mental programming (human nature, cultural background, and personality) contains factors affecting human perception and behaviour (Chapter 2). These aspects apply certainly to SE as well where the human element (targeted person) determines whether the attack succeeds. Common SE attacks include phishing which would not succeed without tricking an end-user into performing an action in the attacker's interest.

SE in the digital domain can be as simple as leaving malware infected USB sticks in the parking area of an organisation and waiting that any employee plugs it into an organisation's computer (Road Apple Attack, Section 4.4.1). The same applies to phishing: phishing e-mails sent en masse may lure some targeted persons into entering e-banking credentials (Section 4.4.3); spear phishing to steal carbon emission certificates seems more complicated to execute and more uncommon though [VG B, 10 K 333.10] (Anecdote 7.3). As SE is not bound to the digital domain solely, it can also happen via other media, e.g., the Spanish prisoner scam mentioned earlier (in-person) or the burglary at an FBI office [Med14] in 1970 (Anecdote 1.1; leaving a note at a door).

> **Anecdote 1.1 — Social Engineering Attack at FBI Office.** "As burglars, they used some unusual techniques, ones Davidon enjoyed recalling years later, such as what some of them did in 1970 at a draft board office in Delaware. During their casing, they had noticed that the interior door that opened to the draft board office was always locked. There was no padlock to replace, as they had done at a draft board raid in Philadelphia a few months earlier, and no one in the group was able to pick the lock. The break-in technique they settled on at that office must be unique in the annals of burglary. Several hours before the burglary was to take place, one of them wrote a note and tacked it to the door they wanted to enter: 'Please don't lock this door tonight.' Sure enough, when the burglars arrived that night, someone had obediently left the door unlocked. The burglars entered the office with ease, stole the Selective Service records, and left. They were so pleased with themselves that one of them proposed leaving a thank-you note on the door. More cautious minds prevailed. Miss Manners be damned, they did not leave a note." [Med14][5]

1.2 Research Problems & Questions

1.2.1 Suitable Definition of SE for Interdisciplinary Security Research

Research in an interdisciplinary field such as SE requires suitable definitions. It should be applicable in other disciplines to foster an interdisciplinary understanding. A knowledge transfer from computer science to other disciplines and vice-versa would be possible, resulting in a knowledge base on SE. Without an interdisciplinary approach this research would reside in the computer science discipline only, harming a holistic view on SE. Various SE definitions exist with the need to examine whether predominant ones are *well enough defined* and also *suitable* for interdisciplinary research including this thesis. A definition must avoid being too broad ('everything is SE') as well as too narrow ('only

[5] Medsger [Med14] as cited in Schneier [Sch14]

spear phishing targeting e-banking credentials is SE').

Research Question 1.1 Which existing SE definition is suitable for interdisciplinary research (not too broad and not too narrow)?

1.2.2 Data Collection and Identification of SE in Various Sources

Evidence that can be identified as expressing SE can foster the understanding of SE. The variety of sources may range from scientific literature and experiments over court documents to news articles and hearsay. The quality, quantity, veracity, and level of detail would differ widely. Thus, criteria are needed to reliably classify evidence as SE. A suitable SE definition (Research Question 1.1) could be broken down into a few obligatory SE indicators.

Research Question 1.2 How to determine whether a story of any source expresses SE?

1.2.3 Social Engineering Susceptibility of Targeted Persons

Attackers attack target persons using SE. The reaction of a targeted person can lead to a successful SE attack if that person was susceptible to the attempt. The reasons for the susceptibility are often manifold. Researchers have to look into various disciplines to grasp this challenge. This thesis focuses rather on the targeted person than on the attacker. With more insights about why some targeted persons are susceptible and why some are more susceptible than others, the Research Question 1.3 will be discussed. This can reveal also how targeted persons fall for specific SE techniques.

Research Question 1.3 Is it possible to characterise which targeted persons are particularly susceptible to particular SE techniques?

1.2.4 Transport SE Situation to the Desired Audience

After an understanding of why and how targeted persons fall for SE attacks, the question follows what can lead to targeted persons becoming more resilient against SE. Various approaches are discussed in the research literature such as interventions, awareness campaigns, trainings or education [AS18; Bul+15; Cap+14; Har13; JMO17; Sch+14; Ueb13a]. The format to conduct SE on targeted persons in such approaches must be effective and easy to comprehend. Research Question 1.4 presents the hypothesis whether the anecdotal presentation of SE events are supportive, e.g., in organisational security policies.

Research Question 1.4 Can anecdotes transport SE situations to the recipients comprehensibly?

1.3 Contributions

This thesis contributes to SE research threefold.

1.3.1 Defining and Identifying Social Engineering for Security Research

A definition for SE is provided that suits interdisciplinary security research. The definition aims to place itself between too narrow and too broad definitions while simultaneously being applicable to various disciplines, notably computer science. Additional SE indicators are developed dissecting the created SE definition into five interrelated components. As longitudinal, interdisciplinary studies regarding cybersecurity do not exist [Leu17], these SE indicators support interdisciplinary research. They can be used to identify SE in longitudinal studies of fellow researchers. The SE indicators are dual-use: they were initially used to find a suitable SE definition. But they can also identify whether an anecdote expresses SE. This is especially useful for the second contribution:

1.3.2 Sources of Information for Social Engineering Research

To understand SE, information about SE is needed. Empirical data sources can be found in various places. Sometimes it becomes challenging to determine whether they are genuine or fictitious. For instance, if a former attacker writes a book stating how some SE attacks worked, researchers have to trust that author and cannot easily verify the veracity in detail via other sources. The same applies to interviews of active criminals. However, ideas originating from fictitious attacks can lead to further investigations, e.g., by conducting experiments to examine whether the attack would be feasible. Experiments become then a reproducible source based on scientific methods. Thus, researchers can gain insights about SE from real or fictitious events — no imaginable SE anecdote will be overlooked. The developed SE indicators help identify SE in both cases.

Three sources are examined in particular in this thesis: court documents about phishing as one case of SE were consulted and analysed. They express real events, are well documented for their original purpose, and are often publicly available. Court documents underlie an iterative truth finding process which can add further SE information. Lego modelling as a second elaborated source was applied in the FP7 EU research project TRE$_S$PASS[6]. Participants of different professions can come up with an imagined attack based on a given setting. The Lego bricks offer an easy and quick start to dive into a haptic attack modelling as Lego bricks are well-known and playful to use. A Lego modelling session was conducted with Hamburg University of Technology (TUHH) students in a seminar created by the author. Finally, a poetry slam style event for ethical hacker gatherings was conceptualised and conducted. The Social Engineering Poetry Slam (SEPS) let ethical hackers tell their anecdotes about SE with optional pseudonymisation, based on real or fictitious events. Recorded sessions can be examined after the event. Researchers cannot ascertain which participant presented a real attack.

1.3.3 Multidisciplinary Overview of Principles, SE Techniques and Susceptibility Factors

To understand how SE affects the susceptibility of targeted persons, a theoretical overview of selected human factors throughout various disciplines, such as psychology, is presented. Firstly, security-relevant general human factors are categorised under the three levels of mental programming, ranging from universal human nature over group specific cultural

[6] Technology-supported Risk Estimation by Predictive Assessment of Socio-technical Security (TRE$_S$PASS)

background to individual personality. Secondly, focusing on SE, well-known factors are discussed, adjoined by an overview of SE techniques such as phishing. Applicable principles, for instance, from Cialdini designed for marketing, describe how SE techniques may succeed in luring a targeted person to comply with an attacker's request. All parts are enriched by insights from SE anecdotes of various sources, e.g., originating from court documents. Finally, a refined Social Engineering Personality Framework (SEPF) based on a literature review relates the susceptibility towards Cialdini's principles to the personality traits of the Five-Factor Model (FFM). One aspect about SE susceptibility concerns why some stories the attacker presents are more successful than others whether regarding the content or the presentation thereof. Hence, some stories may appear more credible to specific targeted persons.

1.4 Scope

Two personas are necessary in SE: the roles of attacker and targeted person. This thesis focuses on the targeted person while aspects of the attacker are considered just to shed light on principles and techniques concerning the targeted person's susceptibility. Therefore, no deep-dives into darknet forums and no interviews with criminals were conducted to examine the attacker's side more profoundly. The Research Questions cover this accordingly. Furthermore, a group of targeted persons can be chosen for experiments more representatively for any desired scientific interest. With the usual limitation of experimental designs, besides qualitative insights also quantitative insights can be gained. Whereas analyses of historic data, e.g., in court documents, offer more qualitative options for understanding of the modus operandi because the entirety of all occurred SE attacks (dark figure of attacks) are being quite obscured. An annual (quantitative) figure about the financial or reputational damage of SE can be extrapolated only vaguely.

This thesis did not develop and conduct, but rather relies on existing SE experiments. Different SE sources are identified and examined qualitatively with focus on the modus operandi and the targeted person's susceptibility. As the targeted persons are a single enabler for SE to succeed, they possess the 'kill switch' to detect, mitigate, and report any SE attack. Targeted persons can be addressed via education, trainings or awareness campaigns whether in schools, organisations or in advertisements for the public. Hence, the understanding of the reaction of targeted persons is of utmost concern.

1.5 Publications

The author published the following documents in chronological order, some of which are listed under the author's Open Researcher & Contributor ID (ORCID) 0000-0001-9228-8248.

"Security-Aware Organisational Cultures as a Starting Point for Mitigating Socio-Technical Risks" by Sven Uebelacker [Ueb13b] was presented in the RiskKom workshop at the GI INFORMATIK 2013 conference. It followed the author's ideas behind the 2002 diploma thesis about "IT-Sicherheit, Unternehmenskulturen und wirtschaftsbedrohende Kriminalität" [Übe02] regarding organisational culture as well as Hofstede's human mental programming. The insights were applied to STS and are incorporated in this thesis.

The 2014 paper "The Social Engineering Personality Framework" by Sven Uebelacker and Susanne Quiel [UQ14] was published in the Workshop on Socio-Technical Aspects in Security and Trust (STAST; IEEE Computer Society). The psychological insights are covered partly in this thesis and are improved. The author contributed as main author by introducing coping strategies per personality trait and complemented the contents of a preceding bachelor thesis. The bachelor thesis of Susanne Quiel [Qui13] about the "Social Engineering in the Context of Cialdini's Psychology of Persuasion and Personality Traits" was supervised by the author.

Zinaida Benenson, Gabriele Lenzini, Daniela Oliveira, Simon Parkin, and Sven Uebelacker published "Maybe Poor Johnny Really Cannot Encrypt – The Case for a Complexity Theory for Usable Security" [Ben+15] at the New Security Paradigms Workshop (NSPW) in 2015. The authors contributed collectively.

A research of all phishing cases in the German juris database by Sven Uebelacker resulted partly in the 2019 paper about "Phishing in höchstgerichtlicher Judikatur" [SU19] in the Rechtstatsachen conference, complemented by the analysis of Austrian court documents. The insights are covered partly in this thesis and are improved. The authors contributed collectively. A previous pilot study in a 2015 supervised bachelor thesis by Ngoc-Minh Michal Pham with the title "Court Rulings as Evidence for Social Engineering Research" [Pha15] provided the preliminary approach for analyses of court documents regarding phishing.

The poster "Privacy-Respecting End-User Reporting of Suspicious E-Mails Using X-ARF" [UM17] by Sven Uebelacker and Adrian Metzner was presented at the 10th International Conference on IT Security Incident Management & IT Forensics (IMF2017; GI FG SIDAR). It was one result of the master thesis "End user reporting of suspicious E-mails using X-ARF" [Met17] by Adrian Metzner, supervised by Sven Uebelacker. Preliminary work on the eXtensible Abuse Report Format (X-ARF) format was done by Sven Uebelacker in previous years [BÜ11; HUV12].

1.6 Naming Convention: Attacker and Targeted Person

Various terms exist to describe roles or personas in presumed SE attacks. Anecdotes are analysed whether they express SE according to Definition 3.17 and its SE indicators (Section 3.1). That is, anecdotes may express SE or not. To avoid introducing additional terms that may be interpreted differently, the same terms will be used. Therefore, the broad terms of *attacker* and *targeted person* will be found throughout the thesis where applicable. Both roles are 'actors' — they act or react. The SE indicators, the SE definition used, and the analyses of anecdotes will use these terms exclusively. As this is an interdisciplinary thesis, the following naming convention is proposed for all disciplines.

1.6.1 Definition of Attacker

The *attacker* role can be found in other sources as perpetrator, social engineer, criminal, offender, fraudster, phisher, agent, scammer or hustler/shill [SW09]. These terms are not entirely synonymous because some express just a subset of what an attacker is. In the SE Indicator 3.4 (Section 3.1), an attacker initiates an attack intentionally with a malicious goal in mind. To clarify: the attacker causes an attack which is an intentional incident, but not an accident (unintentional incident) [Hol07, Def. 2.1]. Figure 1.2 breaks down

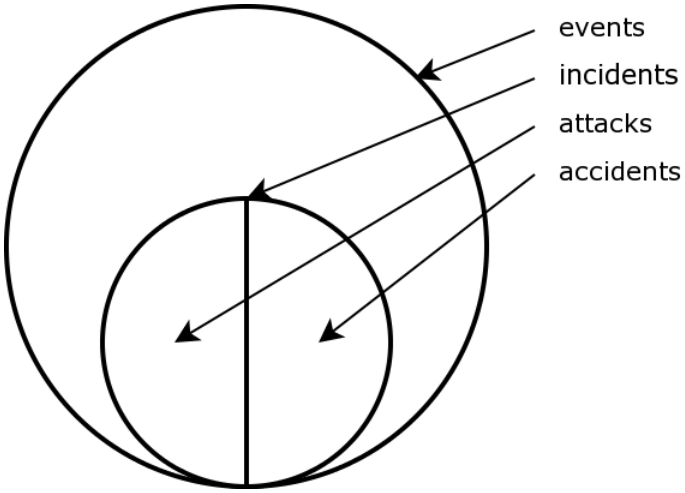


Figure 1.2: Relationship of events, incidents, attacks, and accidents by Holst [Hol07]

this representation which was refined from various definitions in Holst [Hol07]. When focusing on digital identities solely like the NIST Special Publication 800-63-3 [GGF17], an attacker is someone who wants to compromise a *system* maliciously (Definition 1.1). For SE this restriction would be too narrow; SE assets and attackers' goals are not solely of digital nature. The broader Definition 1.2 will be used here instead.

Definition 1.1 — Attacker (GGF17). "A party, including an insider, who acts with malicious intent to compromise a system." [GGF17]

Definition 1.2 — Attacker. An *attacker* is a party who acts with malicious intent.

1.6.2 Definition of Targeted Person and Social Engineering Victim

Some sources use 'target' or 'victim' to describe the *targeted person*. However, target alone may be misunderstood in an interdisciplinary context. For instance, in computer science it can describe a set of decision requests to be evaluated as one element in an eXtensible Access Control Markup Language (XACML) policy [OAS10]. Whereas Friestad and Wright [FW94] used it as a synonym for a targeted person in the Persuasion Knowledge Model (PKM) (Section 4.1). The often used word 'victim' is similar to 'targeted person', but describes only targeted persons who fell for an attack. Hence, a person, who is targeted by an attacker and does not fall for the attack, is not a victim. Unsuccessful SE attacks raise the question why they failed. The targeted person may have detected it because of an awareness training or situational awareness. It is important to understand why some natural persons are more susceptible or more resilient than others. Furthermore, a persons becomes a victim only *after* a successful attack. Hence, the term *targeted person* will include attacks that failed or where the outcome is still unknown. Targeted persons of interest for an attacker can be, e.g., new employees [MS02, p. 61–64] or entry-level employees [MS02, p. 195–208]. The following definitions will be used in this thesis to shed light on victimisation among organisations and beyond:

Definition 1.3 — Targeted Person. A *targeted person* is a natural person who was, is or will be attacked intentionally by an attacker with malicious intent.

Definition 1.4 — Social Engineering Victim. A targeted person becomes a *Social Engineering victim* iff the SE attack targeting that person succeeds benefiting the attacker's intentions.

Because gender plays its part in the susceptibility to SE (Section 4.3.2), gender-identifying words should be avoided for attackers and targeted persons where necessary. That is, (fictional) anecdotes, where gender does not influence the narrative, can omit it. Some sentences may sound unusual because of rephrasing or the plural use of targeted person (his/her passphrase ↪ their passphrases).

1.7 Of Anecdotes in Research

The reconstruction of documented cases of cybercrime activities and the retracing of their analyses can be challenging and impede thorough research. If such cases can be cited, one can reference them, but can often not validate their veracity or completeness entirely. Completeness[7] in the sense that reporters can leave out crucial information because of their own biases (unintentional) or agendas (intentional). The terms '**Anecdote**' (with unique numbering) and equally 'anecdotal information' will be used in this thesis wherever such stories appear — even for those of the author. The reader will not always be able to validate their authenticity. However, anecdotes provide an insight into what could have happened or appeared to have happened for the story teller. The Cambridge Dictionary describes an anecdote as "a short, often funny story, especially about something someone has done" (English) or "a short, often amusing story about an event, usually involving a particular person" (American) [Cam18a]. The reference to a 'person' or 'someone' will fit into the later presented SE indicators (Section 3.1). An anecdote can be seen as an adapted 'case study'. With slight adaptation, case study research methods may be consulted and applied.

Anecdotal information constitutes one of the weakest forms of scientific evidence: it could comprise story-telling, fictional, non-representative or non-reproducible cases with futile falsifiability, and somehow cherry-picked and biased by the reporter. To subsume all found stories under one level of evidence, the evidence category with the lowest scientific claim deemed helpful to avoid missing possible anecdotes. However, it may also flood the data collection as the above set of sources suggests. With respect to evidence-based research (Chapter 5), structured observations include besides case studies also experiments to generate a general knowledge base [SMP17]. In anecdotes, experimental stories can also be told. The environment of SE *experiments* can be controlled most of the time and genuine insights can be created. Anecdotes of SE *in the wild* though could challenge researchers whether they are based on facts or are fiction. Research in narrative psychology (Section 1.8) shows that stories resemble the mode of thinking [AS15]. The reader needs

[7] Completeness expresses the "degree to which values are present in a data collection" [BS06].

less mental effort to immerse into a topic. Anecdotes therefore support a more comprehensive understanding of a described situation. One challenge remains to formulate anecdotes in a way that is understood by the intended recipients correctly. For instance, the sentence "I hate Mondays more than Garfield" can raise the question "why do you hate Garfield at all?" [Ada15]. Cross-linguistic research [GW04] hint to an achievable way to address different languages by translating a Natural Semantic Metalanguage (NSM) to *comprehensible* anecdotes. This metalanguage can support creating comprehensible anecdotes because the language patterns recipients speak influence the way of their thinking [Luc01] (Sapir–Whorf hypothesis).

SE anecdotes here must cover human action or interaction which can lead to a successful SE attack. Sometimes insights based on literature lack the proof of authenticity of each detail. Mitnick and Simon [MS02] are often cited (good!) as a well-known source of deception and SE examples, but it is almost impossible to verify their details. Court documents offer a trustworthier source of the underlying truth finding process, but show only partly the bright field of incidents that had occurred. More on court documents as one SE evidence source will follow in Chapter 7. After further examination they may or may not express SE (Section 3.1 about SE indicators). 'Anecdotes' will be called 'SE anecdotes' if SE was identified.

For instance, Anecdote 1.2 happened to the author, but the authenticity cannot be corroborated by fellow researchers:

> **Anecdote 1.2 — To follow or not to follow security policies.** In some situations it is hard to decide whether to follow or ignore an organisational policy — and even worse if this policy is implicit. Around 2006 I just finished my daily work and wanted to leave a building of TUHH. After 6pm the doors were locked, entrance was only possible with a key. A formally dressed man wanted to get through while I was leaving. I politely denied him access because the University's decision and policy was to grant access to authorised personnel only. He said he had an appointment with the president in that building (4^{th} floor). Again I had to apologise that I could not know that and therefore cannot let him through. His response was to call someone on his mobile phone, then give it to me. The called person identified herself as the president's secretary and told me to let him in. Still I did not know if this was a fake or genuine call and had to refuse. Many people would be polite, gullible, helpful or just ignorant and let anyone in. In my role as deputy of the IT security officer, I was well aware of consequences. Also the ignorance of the law (policy) is no excuse, but expert knowledge can lead to harsher judgements. I had to deny access. Cases in Hamburg describe that strangers with malicious intent try to tailgate or ring at an organisation's door and behave pushy to enter.

It shows that on one hand many organisations cultivate team spirit, mutual respect, trust, politeness and helpfulness. On the other hand, security policies and risk awareness could undermine such efforts when denying some requests is recommended, thus be alienating likewise. In hindsight possible persuasion principles (more in Section 4.5.1) can be identified: the formal dress as an authority indicator can lead to compliance, supported by pushy behaviour. The same applies for mentioning the importance of meeting a person in

higher position. Handing over one's personal mobile phone to an unknown recipient seems to reciprocate trust and the called person could be a shill impersonating an authority figure ('Authority by Hierarchy', Section 4.5.2). For a helpful and polite person an inner conflict arises between granting access to an immediate contact or following (unaligned) policies.

1.8 Narrative Psychology as Means for SE Research

SE awareness campaigns aim to reduce the targeted persons' susceptibility to SE attacks. Evaluation of the distinct requirements of which training episode needs to be concentrated on, helps to find proper material. SE anecdotes form a communication channel between a narrator and an audience. They can be applied in other areas, too. To explain the foundation of that approach, Research Question 1.4 will be discussed here.

To examine how anecdotes form our way of thinking or vice-versa, a dive into the field of psychology is necessary, in particular narrative psychology. Bruner [Bru90][8] interpreted narrative psychology as that persons act on the basis of their "beliefs, desire, and moral commitment". He identified that actions are determined mainly by the changing draft of the autobiography in one's mind. These autobiographies are imagined stories influencing decision-making. Story telling is a mode of thinking [AS15]. In accordance with Crossley [Cro02, Chapter 1] experiencing time and temporality as well as finding the 'order of meaning' constitutes the human consciousness. What gives a narrative meaning is based on activity (time, sequence of events) to become meaningful, understood, and interpreted. Sarbin (1986)[9] applies the term narrative coterminously with story. In this thesis, the term anecdote as described in Section 1.7 will be used instead of 'story' after this elaboration on narrative psychology.

1.8.1 Applicability of Stories

Telling stories of SE attacks seems promising for reaching an audience. Stories resemble the mode of thinking like an 'isotonic' source or an easy-to-swallow pill — less mental effort needed to immerge into a topic. For instance, recipients do not need to create own stories based on their interpretation of organisational policies if narrative elements exist. A policy stating 'Do not plug-in an USB stick of unknown origin into any of the company's computers' is a valid rule, but lacks the insight of 'why?' and 'what could happen?' if violated. Examples in anecdotes can explain how policies are meant to be understood.

The frequently updated *IT-Grundschutz Catalogues* [BSI15] of the German Federal Office for Information Security (BSI) provide "recommendations for standard security safeguards". The catalogues are a widely accepted and grounded source of security threats like SE (Definition 3.16) and countermeasures. The provided so-called examples substantiate each topic by conveying situations mainly in third-person perspective (examples: Anecdote 1.3, Anecdote 4.9, and Anecdote 4.11), but are, generally speaking, anecdotal information (Section 1.7). This offers the reader a better understanding of each threat as it resembles the reader's mode of thinking.

The challenge for policy makers remains to foster a mutual understanding with recipients

[8] as cited in Akerlof and Shiller [AS15]
[9] as cited in Crossley [Cro02]

concerning the big picture on information security. Policies are mainly designed top-down and need checking if sufficiently comprehensible for the intended audience. Stories supplement organisational policies and enable a more flexible way to express the meaning of each policy. That is, stories can be adapted to serve various departments and professions or even be adjusted with respect to linguistic differences and cultural background. While a policy can be out of context for or misinterpreted by the recipient, stories establish context. Hopefully, recipients memorise the intentions as easily as recall and adapt them in real-life situations. To enable recipients to improve recalling stories later, the narrator should favour vivid over pallid stories due to the availability bias [Sch08b; TK74] (Section 2.3.2).

> **Anecdote 1.3 — T 5.142: Spreading malicious software via mobile data media.**
> "At a convention, a visitor wanted copies of the slides of the presentation just made and asked the lecturer if he could make copies of the slides available. The lecturer gave the visitor the USB stick containing the presentation slides. When the visitor inserted the USB stick into his laptop to copy the slides, a malicious program on the USB stick installed itself without the visitor noticing it." [BSI13][10]

Interlocutors are essential for self-interpretation, i.e., a relation to other actors is needed for understanding one's own self (Taylor (1989)[9]). This "interchange of speakers" can be found in SE stories because of the direct or indirect communication of attackers and targeted persons (SE definition 3.17). Thus, SE stories as a narrative source contain at least two actors: attacker and targeted person (Section 1.6). Policies contain a higher level of abstraction than stories would. Policies in general are further away from narrating 'social' interaction and may draw away its audience from its intentions. Based on Taylor's (1989)[9] research on the formation of moral and ethics, adapted stories can be seen as one approach envisioning what is interpreted as the intended 'good' for individuals. Hence, well-crafted SE anecdotes support policies to avoid misinterpretation and to communicate the intention likewise. As narrative psychology shows, stories are of great value for communicating a topic — in this thesis SE anecdotes. However, a normative approach is essential for comparable, comprehensible and varying anecdotes with little bias. Vague SE stories can be refined. For security-relevant decision-making, Rader et al. [RWB12] showed that "security stories" are shared from family and friends as "informal lessons" of incidents. They influence the security behaviour and thinking of the recipients even more if they originate from knowledgeable persons or entities. While these security stories are (re-)told, the documented SE anecdotes in this thesis share one of the intentions, namely to learn from them impacting the decision-making of targeted persons.

To address Research Question 1.4, stories such as SE anecdotes express the mode of thinking of the recipients. They present not only a feasible, but furthermore a promising way of communication compared to unintelligible organisational policies. Because of the nature of SE situations, SE anecdotes as a subset of story-telling must contain human interaction, a basic facet of narrative psychology.

[10] Example 2 of IT-Grundschutz Catalogues, threat T 5.142 [BSI13] / G 5.142 [BSI15, draft]

1.9 Outline

The thesis is structured in two parts: Part I starts with an overview of selected general human factors with a focus on SE (Chapter 2) in preparation for Chapter 4. Chapter 3 searches for a suitable SE definition by creating five SE indicators and a metric, following Research Question 1.1. Because of the dual-use character of the SE indicators, Research Question 1.2 can be answered, too. The last chapter of Part I shows additional susceptibility factors, SE techniques and principles to give an understanding of how SE works. It provides an overview of why some targeted persons fall more likely for SE grounded on Chapter 2's general human factors (Research Question 1.3). There (Chapter 4), the SE indicators are applied to the SE anecdotes exemplarily for the first time. The chapter ends with a proposed mapping of Cialdini's principles and personality traits. This SEPF specialises on the susceptibility research effort exploring the targeted persons' personality traits in scientific literature (Research Question 1.3).

Part II shows in Chapter 5 how a science of security can be pursued, addressing common criticism and based on a previous digression into the philosophy of science. The chapter also discusses the creation of an interdisciplinary, shareable, and evidence-focused knowledge base which should foster a mutual understanding of SE aspects. To not overlook any hints on SE, a wide range of SE sources are presented in Chapter 6 under the umbrella of evidence-focused research. Again, the SE indicators of Chapter 3 help to separate SE from non-SE anecdotes. The last chapters of Part II cover analyses conducted by the author: Chapter 7 examines court documents (historic data) regarding the SE technique phishing. Chapter 8 documents a Lego modelling session in a cloud scenario setting. Finally, the novel concept of a Social Engineering Poetry Slam (SEPS) is described and a conducted session transcribed and analysed (Chapter 9). The thesis closes with the Conclusion in Chapter 10.

Defining Social Engineering

2. It's About People

When information security involves or is affected by human behaviour, these human factors (Definition 2.1) must not be ignored. Without proper assessment, information security will be impacted negatively like choosing inappropriate security trade-offs [Sch08b]. Humans are a complex research object as researchers are intrinsically part of that species. But information security is, of course, not the only field affected by human factors. As Figure 1.2 categorised, humans can also produce or influence the outcome of events, some of which be viewed as incidents. That is, human-related incidents can comprise accidents or attacks — depending on the perspective. Reason [Rea90] elaborated on human error related to unsafe acts regarding psychological aspects (Figure 2.1). He split unsafe acts into groups of unintended (slip: attentional failures; lapse: memory failures) and intended actions (mistake: rule-based and knowledge-based mistakes; violation: routine violation, exceptional violations, acts of sabotage). The category 'violation' was separated from basic error types (slip, lapse, mistake), meaning that in unintended actions and intended actions leading to mistakes[1] the subjects acted without malicious intent. Hence, accidents would fall under the human error types. "Human error happens because we do not notice the sequence of events leading up to the visible error until (if ever) it is too late." [Thi16] Reason [Rea08] refined his view in 2008 on human error by focusing his interest more on human resilience to errors than on the original error [Thi16]. Human error affects security and safety likewise. This matches the intentions of this thesis: categorise what expresses Social Engineering (SE), identify and collect SE evidence as well as understand why some targeted persons are more susceptible than others. For sure, an attacker initiates an attack and cannot be left out of the equation; however, this thesis focuses on the targeted person. Excluded are intended actions violating knowingly security precautions such as sabotage by a disgruntled employee.

[1] not to be confused with intended mistakes

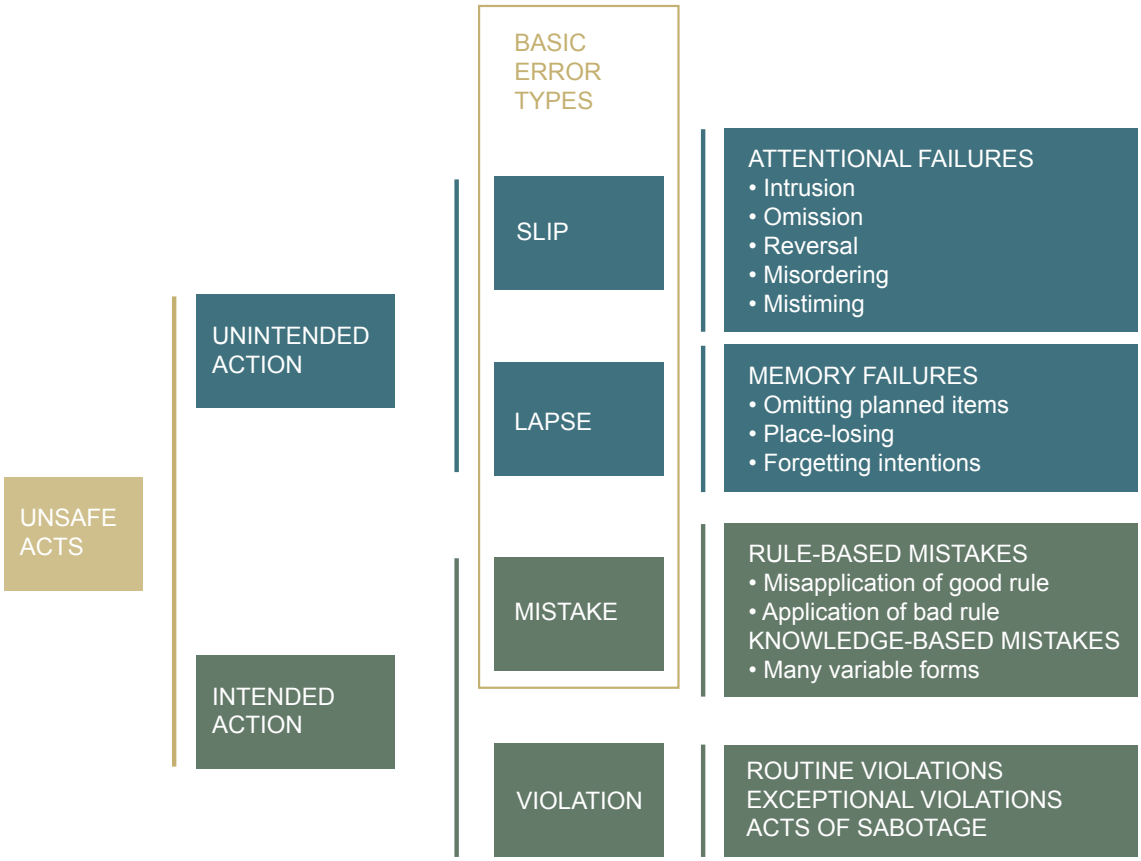


Figure 2.1: Human error classification of unsafe acts as defined by Reason [Rea90] (original version in Figure C.2.2)

Definition 2.1 — Human Factor (Fag07). "The mere fact that human beings can potentially make mistakes that might affect a given situation." [Fag07]

This thesis is about people *and* cybersecurity with a focus on SE. Before elaborating on SE, a summary of human factors that might influence the susceptibility of targeted persons towards SE attacks will be presented in this chapter. Although human factors are assumed to cover work-related behaviour [Thi16], the author likes to examine it more generally in this chapter. All three levels of human mental programming (Figure 2.2) according to Hofstede [Hof01] are consulted, starting with human nature, followed by cultural background, then personality. Hofstede theorised that all humans have an underlying universally predominant human nature that is inherited and not learned (Sections 2.1–2.6). The cultural level is learned from and passed to other members in a group as a collective phenomenon (Section 2.7). Finally, what makes a subject unique in its characteristics is the personality level, specific to each individual (Section 2.8). Where exactly the borders between personality and culture as well as culture and human nature reside, is disputed in social sciences [Hof01]. For instance, a statistical link exists between personality traits and culture [HM04].

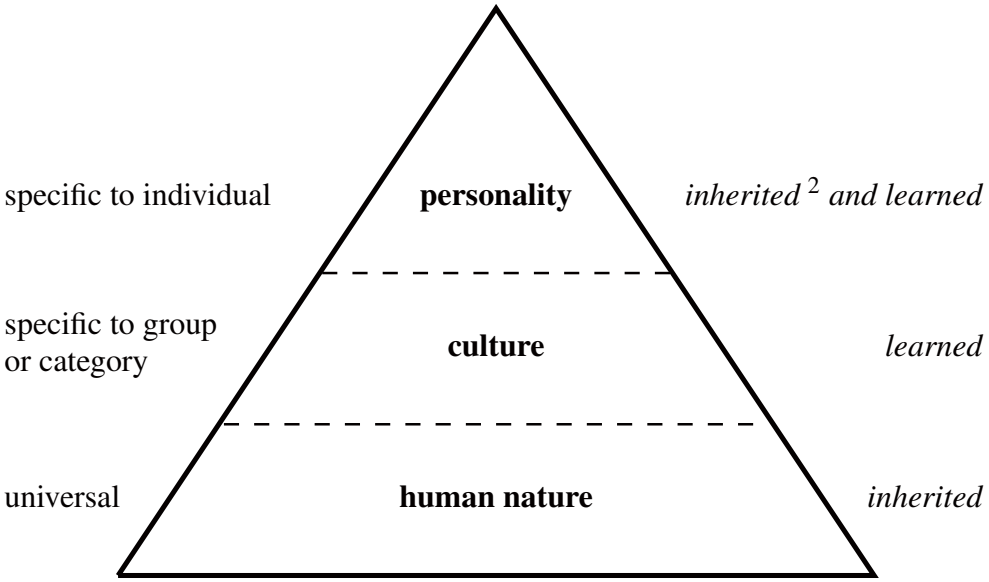


Figure 2.2: Three levels of uniqueness in human mental programming [Hof01] as depicted in Uebelacker [Ueb13b] and Übelacker [Übe02] (German Figure C.1.1)

2.1 Human Memory

Schacter [Sch99] identified seven shortcomings ("sins") of human memory: transience, absent-mindedness, blocking, misattribution, suggestibility, bias, and persistence — sorted into categories of forgetting, distortions, and intrusive recollections. Some of which can be found in Reason's human error classification (Figure 2.1). The following description originates from Benenson et al. [Ben+15] applied to security:

Transience (forgetting[3])
"forgetting passwords, steps of authentication routines, meaning of warnings, key points of awareness training"

Absent-mindedness (forgetting)
"clicking on a link in an email without paying attention; forgetting laptops or smartphones in public places; forgetting an authentication token at home or in some other place (and not being able to recollect where it is)"

Blocking (forgetting)
"not being able to recollect passwords or PINs that are usually well remembered, but just slipped the memory in a concrete situation, for example under stress or observation"

[2] The English version of Hofstede [Hof01] used 'inherited' twice. The German diagram is more precise: personality is 'erlebt und erlernt', human nature is 'ererbt'. The German version is attached in the appendix as Figure C.1.1

[3] How good memories can be recalled depends also on the vividness of its stories [TK74] as mentioned in Narrative Psychology (Section 1.8) and in Section 2.3.2 on the availability bias.

Misattribution (distortions)
"falsely recollecting seeing a person on the premises of the company although actually the person was seen in some other context (e.g., in a shop or on an exhibition), and thus letting them to tailgate; email scam that includes the (necessarily incorrect) customer ID of the victim in an attempt to make the victim 'recognise' their ID, believing in the information implanted by social engineers in spear phishing e-mails"

Suggestibility (distortions)
"agreeing to the course of events as suggested by a social engineer ('we met last week in this meeting, where X said this and Y did this, do you remember?') and thus complying with his/her request for information or action; email scams that mention past communication in the email subject, such as 'Re: request No. 23019' "

Bias (distortions)
"falsely remembering after an awareness training that a fraudulent email can only come from an 'unknown' person and thus not being suspicious of an email with a spoofed known sender"

Persistence (intrusive recollections)
"remembering an old password or PIN instead of a new one; remembering that a former colleague is still working for the company (although the knowledge that it is not so is actually present) and talking to him/her about confidential things."

2.1.1 Transactive Memory

Another interesting aspect reveals the *transactive memory* of a 'group mind': Wegner [Weg87] stated that a transactive memory system comprises a set of individual memory systems joined with interpersonal communication. Individual memory can be found stored externally in a group memory, e.g., group members may recognise knowledge experts for specific fields and consult them. However, the knowledge of knowledge experts in a group also adds to the complexity of an individual's memory system [Weg87]. If a member perceives an e-mail as suspicious, maybe phishing, a knowledge expert can be asked for an expert opinion. The group and individual memory combined may correlate to higher resilience against security threats. Furthermore, research in group theory examining cross-functional teams may focus on such resilience.

2.2 Attribution Theory

Attribution processes describe the "systematic ascription of causes and effects in situations of failure or success" [JG13]. They are important for analysing and understanding human emotions, motivation, and behaviour. Attribution theory examines a subjects' explanation of specific events: on one hand the *causation* of 'why' something happened in a specific way; on the other hand the amount of *control* subjects think they had over the outcome. The following attributional dimensions exist according to Janneck and Guczka [JG13], appended with statements illustrated by the author:

Locus
determines who or what the subject blames for a failure. *Internal* attribution categorises a self-inflicted cause, whereas *external* attribution claims external causes: 'I was susceptible to this phishing attack' vs. 'the IT department did not protect me sufficiently against phishing'.

Stability
differs between a subject's *temporal* explanation whether it was a one-time cause or not, e.g., 'this time I failed to detect the scam' vs. 'I always fall victim to scams'.

Controllability
describes how much subjects think they had influence on a specific outcome. 'Without any time pressure to get my work done, I would have detected the scam' vs. 'the recent scam techniques are too sophisticated, I was not be able to identify this scam'.

Globality
covers how much the explanation of the event is believed to be generalisable: 'I succumbed to this specific scam' vs. 'I will always fall victim to all types of attacks'.

Attribution styles are considered stable over time, partially expressing a subject's self-concept. They represent attributional patterns that re-appear in different events and persist in various contexts. A subject with internal locus who believes a situation is uncontrollable might feel helpless and resign. If such perceived helpless situations pervade varying constellations, the subject may end up with a behaviour called 'learned helplessness'. Attribution theory may shed light on understanding behaviour and decision-making of targeted persons as the example statements suggest.

2.2.1 Attribution Biases

Attribution biases express one group of cognitive biases — other types of cognitive biases will follow later in this chapter. The *negativity bias*, another attribution bias, will be covered under prospect theory, Section 2.3.1. A few attribution biases relevant to this thesis are described here:

Defensive attributions[4] occur when subjects assign responsibility of failure to other people instead of themselves. The subject assumes that the other side would have been able to control the situation, but failed. Simultaneously, this belief protects subjects from blaming themselves, cf. attributional dimensions of 'Locus'. This bias may influence the subject's ability to learn from mistakes. Defensive attributions can partially relate to cognitive dissonance.

The *actor-observer bias* or asymmetry shows similarities to defensive attributions when interpreting the behaviour of other people (observer) or of oneself (actor): subjects tend to view their own behaviour as influenced more by situational factors. Observed behaviour of

[4] Definition: "Defensive attributions are explanations of behaviors that serve to defend an individual's preferred beliefs about self, others, and the world." [BV07]

others are more likely to be interpreted as the observed subject's dispositions. That is in a SE context, a subject may assume that it would have less control over mitigating SE due to situational factors (attributional dimension 'Controllability'). On the other hand, the subject may attribute more control to the observed targeted person, see also optimism and control bias (Section 2.6).

2.2.2 Pollyanna Principle

Unsurprisingly, when there is a negativity (conscious) bias (Section 2.3.1), a *positivity bias* also exists. It covers the subconscious bias that subjects tend to recall positively attributed, optimistic memories better. The synonymously used *Pollyanna Principle* can be found, for instance, in a universal bias towards the positive in human communication, as a study across ten languages showed [Dod+15]. It can correlate to gullibility when communication is involved (Section 4.3.1). Human communication in SE is an inextricable part of the later presented SE indicators (SE Indicator 3.2 in Section 3.1).

2.3 Prospect Theory

In the last century the disciplines of economics and psychology pursued different paths to become more scientific [Cam99]. Economists focused more on mathematical modelling inspired by physics. Whereas psychology followed experimental traditions similar to natural sciences. The result was and still is the application of divergent methods. That may be the reason why Adam Smith's "The Theory of Moral Sentiments" (1759) was mainly ignored by both disciplines [Cam99]. Smith created the *impartial spectator*[5], who observes the economic behaviour of one individual without interaction [ACL05]. It covered, for instance, intertemporal choice where short-term gratification conflicts with long-term costs. Furthermore, individual preferences such as loss aversion, overconfidence (Section 2.5), and altruism can be identified in Smith's work [ACL05].

2.3.1 Loss Aversion

When researchers began to reunify economics and psychology, behavioural economics emerged [Cam99]. Cognitive psychologists studied, e.g., economic decision making: how economic decisions are influenced by human biases, be it emotional, social or cognitive [Sch08b]. Bernoulli's Expected Utility (EU) theory assumes that subjects assess risks and occurrence probabilities *rationally* and integrate these insights into their decisions, e.g., whether to purchase an insurance [Cam99]. However, it ignores psychological aspects of the decision maker: loss aversion motivates subjects more than gains of the same amount [Wes08]. In 1979 Tversky and Kahneman [Kah11] coined this effect under *prospect theory*. For instance, the decision on similar trade-offs (losing/gaining $500 vs. a 50% chance of losing/gaining $1,000) showed that subjects preferred gaining $500 (84%) over a 50% chance of gaining $1,000. When deciding on loss, subjects chose a 50% chance of losing $1,000 (70%) over the certain loss of $500. Depending on loss or gain probabili-

[5] impartial spectator: "moral hector who, looking over the shoulder of the economic man, scrutinizes every move he makes" (Grampp, 1948, as cited in Ashraf et al. [ACL05])

ties were over- or underweighted.[6] That is, depending on how a choice on a trade-off is presented, decision outcomes may be influenced by the *framing effect* [Sch08b]. In general, evidence shows that low probabilities are overweighted for losses and gains [Cam99]. One reason may come from evolutionary strategies to survive the day: a safe, smaller gain is more likely to be chosen over risking any larger gain as well as risking larger losses over accepting smaller losses [Sch08b]. The *negativity bias* as one attribution bias (Section 2.2.1) strongly relates to that assumption. Research shows that the human brain handles losses and gains in different regions, most likely to be processed differently as well [ACL05]. Experiments were conducted that suggest loss aversion behaviour is 2.0–2.5 times more motivating than a potential gain.[7] While this behaviour is categorised under universal human nature, cultural nuances in uncertainty avoidance might impact the magnitude of the behaviour (Section 2.7.1). Anecdote 2.1 shows a type of financial loss aversion when buying a house with time being scarce. Scarcity was also involved the following year when many buyers had already bought a new home (fewer buyers, bad for sellers) as well as the real estate market became smaller (fewer real estates, bad for buyers). Concerning SE, maliciously targeting a person with an expected loss will motivate more than with an expected gain.

> **Anecdote 2.1 — Eigenheimzulage.** The German government let expire the 'tax credit for first-time home buyers' (Eigenheimzulage[8]) for newly bought homes end of 2005. The real estate market was on fire towards the end; many people wanted to get the real estate subsidiary, hoping to spend less in total. But higher demand meant higher prices and potential buyers were under time pressure. Buyers wanted to avert losing this subsidiary.
>
> When I bought a home in 2006, the market was extremly quiet and as expected the prices dropped. The previous owner mentioned that another buyer was interested and would pay more than me. I kept calm and said she can call me if the other buyer backed out. She called me a few days later. Finally, I bought that home at a price in total less than it would have been the year before with the tax credit substracted.

2.3.2 Availability Bias

Plous[9] categorised this cognitive bias in three general terms:

(i) if an event is more *available* to a subject, it will be perceived as more frequent or probable as it really is.
(ii) the *vividness* of an information determines how good it can be recalled later while becoming more convincing.
(iii) if an information is more *salient* it can seem to be more causal.

Tversky and Kahneman illustrated availability as a person may "assess the risk of heart attack among middle-aged people by recalling such occurrences among one's acquain-

[6] Over- or underweighting *given* probabilities differs from the effect of over- or underestimating (missing) probabilities.
[7] Tversky/Kahneman (1981) as cited in Schneier [Sch08b]
[8] EigZulG: https://www.gesetze-im-internet.de/eigzulg/

tances" [TK74]. This type of assessment of risk frequency or risk probability is based on heuristics (mental shortcuts, Section 2.4). Vivid memories are better recalled than pale ones and impact judgements in future situations [Sch08b]. Regarding cybersecurity, abstract security policies may hinder their adaptation to real-life events. Moreover, repetitive news coverage of certain events makes them more available. Subjects may draw false conclusions that such events are more or less likely to occur, e.g., in reportings of crimes when the actual number in most fields has in fact declined over decades. Security decisions may not fit the really existing threats.

2.4 Dual Process Theory

As prospect theory shows, human decisions are not entirely based on rational, logical thinking. The dual process theory relates to this insight. Tversky and Kahneman [Kah11] defined the *system 1* (also known as peripheral processing or type 1) as responsible for decisions based on heuristics. It takes less mental effort and is generally faster because the brain processes information more subconsciously and without rational thought. Evolutionary, the reaction to sighting a potential predator is more vital to survival than evaluating whether it is dangerous. A false-positive decision (fleeing from a not hungry predator or from a look-alike) is better than a false-negative one. System 1 is often associated with 'rule of thumb', shortcuts, stereotypes, and biases [Sch08b]. It enables humans to make "close-to-optimal answers quickly with limited cognitive capabilities" [Sch08b]. Cialdini [Cia07] states that people need heuristics to handle the amount of decisions to make. The human brain is able to process all decisions *rationally* (see system 2).

System 2 (central/systematic information processing or type 2) handles the more rational decisions. It uses higher mental effort demanding more attention and decides primarily based on conscious thinking and not on heuristics. Researchers assume that system 1 is evolutionary older than system 2. System 2 is able to "respond to threats that loom in an unseen future [and it] is still in beta testing".[10] Both systems describe the typical human decision making process.

When designing security-related procedures or (counter-)measures, designers need to be aware of the dual process theory. That is, usability plays a major role in developing applications with security impact. The combination of usability and security resulted in the field of *usable security*. Often unusable security processes get rejected by the end-users, e.g., creating PGP encrypted e-mails [WT99]. A security procedure that is intuitively usable (using system 1 heuristics) should be preferred. This way the end-user can make security decisions and their corresponding security trade-offs intuitively [Sch08b]. The same applies to safety-critical warnings: loose wheel nut indicators on lorries show whether a nut became loose [Mas15]. The arrows are aligned circular if nuts are tightened correctly. The heuristic system does not need much effort to recognise this pattern. However, if the circle is broken, system 1 detects it and system 2 can start to evaluate the situation.

Recent creativity research [Ole+17] links the dual process model to the personality traits of the Five-Factor Model (FFM): artistic creativity as one aspect of the openness to experience

[9] Scott Plous (1993) as cited in Schneier [Sch08b]
[10] Psychologist Daniel Gilbert (2006) as cited in Schneier [Sch08b]

trait is more associated to system 1's features like pattern detection. System 2 features seem to be more likely used in scientific creativity (intellect as the other aspect in the openness to experience trait), more on personality traits in Section 2.8. The dual process model was also adapted to persuasion research, e.g., by Guadagno and Cialdini [GC05], see Section 4.5.1 for Cialdini's persuasion principles.

2.4.1 Cognitive Capacities

"In cognitive psychology, cognitive load is the total amount of mental effort used by the human working memory." [Ben+15] If a security task becomes infeasible, that is, beyond one's cognitive capacity, it may be possible to substitute it with a task of the same security level or to support subjects with training, mnemonic techniques or complementary security technologies [Ben+15]. However, it is important to focus on the level of security of the original problem solution: the attribute substitution bias (Section 2.6) does the opposite and substitutes intuitively (system 1) a high effort problem for an easier one, likely resulting in a wrong solution. Consciously (system 2), if subjects perceive a security task too complex or unworkable, they might look for their own workaround, eventually sidestepping designed security controls [BSW08]. It is crucial to understand the cognitive capacity involving security tasks to minimise subconscious biases (system 1) as well as to prevent the creation of undesired workarounds. If these security workarounds exist without the knowledge of security policy makers, they become uncontrollable and may harm the security landscape. The concept of "shadow security" suggests that an organisation can learn from such security workarounds if known, transforming them to desired workarounds [KPS15].

Miller [Mil56] hypothesised about the limitations of the working memory in human information processing. One information bit is referred to as sufficient to decide between two alternatives. Based on empirical data from Irwin Pollack's 'channel capacity' research (1953), Miller [Mil56] presumed that the information transmission has an upper limit of 2.5 bits (SD 0.6) in our nervous system for judgements, i.e., 7 alternatives ± 2 ("Magical Number Seven" [Mil56]). The exact number of bits deviates depending on whether people are novices or experts, younger or older.[11] In the majority of tasks, security is the secondary goal for end-users, e.g., the primary goal is communication (sending an e-mail), the secondary goal is confidentiality (encrypting it) if confidentiality was not required [Ben+15]. An upper cognitive bound may influence the applicability and effectiveness of security policies and mechanisms. *Usable security* can support end-users by lowering the cognitive effort on behaving 'securely'. Benenson et al. [Ben+15] theorised a complexity framework for usable security. Their approach consisted of including cognitive capacities (inside/beyond) into complexity categories of security (secure/insecure) and usability (fully/partially/maybe/unknown usable; fully/partially/unknown unusable). Figure 2.3 depicts the two dimensions of security and usability. When starting with S (Socio-Technical Systems (STS)) different directions to reach S' can be taken to improve security, usability or even both. Also, security designers may identify which trade-offs are acceptable if usability or security need to be degraded. For now, end-users may face the burden to perform security tasks beyond their cognitive capacities.

[11] Wickens et al. (2013) as cited in Benenson et al. [Ben+15]

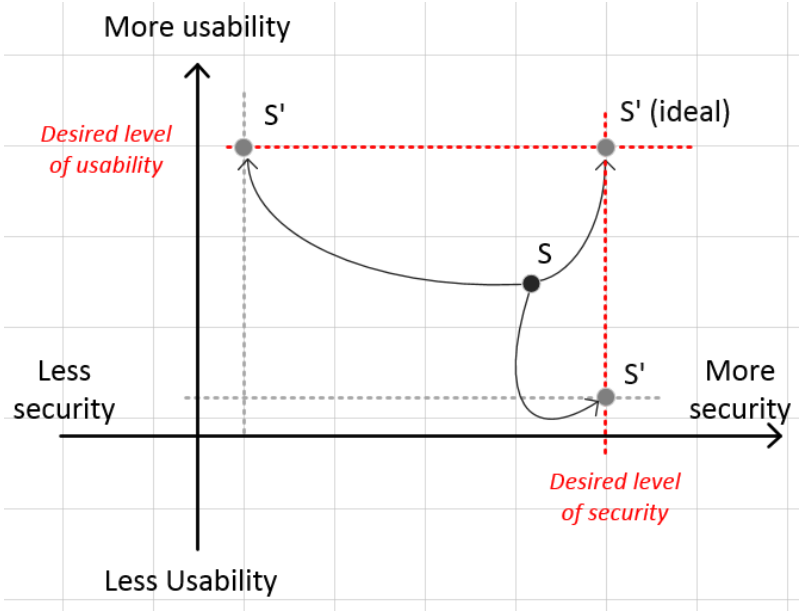


Figure 2.3: Improvement directions for usability and/or security in Benenson et al. [Ben+15]

2.5 Overconfidence

Adam Smith already elaborated on the effects of overconfidence in the 18th century [ACL05]. Moore and Schatz reviewed the literature on overconfidence and broadly defined overconfidence as a subject has "greater confidence than reality justifies" [MS17a]. Since Smith's work, three types of overconfidence were explored [MS17a]:

(i) *overestimation*,
(ii) *overplacement*, and
(iii) *overprecision*.

The first one concerns the self-confidence that subjects think they are 'better' than in reality. Overplacement means that subjects believe they are better compared to others. If a subject is certain to know a topic but reality contradicts this view, the subject behaves overprecise. Moore and Schatz [MS17a] showed Donald J Trump's behaviour as examples: he claimed his worth is higher than credible sources revealed (overestimation). Trump claimed he won the "largest electoral victory since Ronald Reagan" compared to other presidents (overplacement). He was certain that "thousands of Arab Americans in New Jersey publicly celebrated the fall of the World Trade Center on September 11, 2001", but no evidence confirmed this statement (overprecision).

No matter which type of overconfidence is involved, all may impact the decision making concerning cybersecurity. Subjects may overestimate their ability to detect attacks and *feel* secure, Section 8.2.2, but not be as secure in reality. They can be overconfident about their knowledge to be more secure than others (see also optimism bias in Section 2.6). Or they may express a level of certainty towards a topic, for instance, dealing once with a unique phishing e-mail allegedly promotes them to knowledge experts on phishing.

2.5.1 Dunning-Kruger Effect

One specific aspect of overconfidence related to the work environment can be found in the *Dunning-Kruger Effect* [KD99]. This cognitive bias lets subjects overestimate their competence in a task while lacking the metacognitive skills to self-reflect. They are "unskilled and unaware of it" [KD99]. Kruger and Dunning [KD99] see a "dual burden" for people less competent in one topic, but not realising it:

(i) false conclusions lead to errors (unskilled);
(ii) due to incompetence, they are not realising it (unaware).

Study participants in the bottom competence quartile showed "deficient metacognitive skills" [KD99]. They were also struggling more to assess the competence and performance of others compared to participants in other quartiles. Kruger and Dunning [KD99] compared this effect to the medical condition of anosognosia, a deficient self-awareness of existent disabilities. The German philosopher Marquard [Mar74] coined the related term "*Inkompetenzkompensationskompetenz*" (incompetence compensation competence). Interestingly, training can improve metacognitive skills for the bottom quartile resulting in "less inflated self-assessments" [KD99] and acknowledgement of previous incompetence. Metacognitive therapy elements of mental health prevention become more applied in recent work-related coaching practices [Kor+16], such as the evidence-based trainings with metacognitive techniques of the Hamburg based addisca gGmbH.[12]

Related to competence in the workplace, the *Peter principle* [PH69] focuses on promotion in management where a person is promoted based on competence as long as the person reaches the final position. There, the person is incompetent (level of incompetence) and stays. Although initially meant satirically, Adams [Ada08] defined the *Dilbert Principle*. It states that incompetent employees get promoted to keep them away from productive, competent employees. Peter and Hull [PH69] called this "percussive sublimation" (pseudo promotion) as one apparent exception to their Peter principle.

2.6 Other Cognitive Biases

The term cognitive bias is widely used in the literature throughout disciplines. In the previous sections some cognitive biases and their underlying theories were discussed that the author deems relevant to cybersecurity: defensive attribution, actor-observer, and positivity bias (Pollyanna principle) in the attribution biases Section 2.2.1; loss aversion with the negativity bias and availability bias under prospect theory (Section 2.3); the dual-process theory and cognitive capacities in Section 2.4; the Dunning-Kruger effect in the overconfidence Section 2.5. Additional relevant cognitive biases can be found here.

Optimism Bias

Subjects are convinced that certain negative risks in the future are more likely to concern other people than themselves. Subjects are more optimistic towards their activities than to the same activities of others [Sch08b], for instance, in activities

[12] https://www.addisca.org/

harming one's health like smoking. Such bias is strengthened by the logical fallacy that dead (optimistic) subjects cannot refute this optimistic view anymore (also known as *Survivorship Bias*). Weinstein [Wei80] showed under the synonymous term *unrealistic optimism* that subjects considered their chances to be afflicted by negative events below average and positive ones above average. He concluded that the "degree of desirability, perceived probability, personal experience, perceived controllability, and stereotype salience would influence the amount of optimistic bias" [Wei80]. Subjects may accept more likely security risks that they assume others are more prone to. If credit card fraud via phishing happens only to others, subjects can become more susceptible.

Control Bias

or illusions of control express the tendency that subjects more likely accept risks over which they think they have control [Sch08b]. They overestimate the real, personal influence of the perceived controllability of events based on personal factors. These illusions are likely to be found in situations where "personal involvement, familiarity, foreknowledge of the desired outcome, and a focus on success" are present [Tho99]. For Schneier [Sch08b] the optimism bias manifests itself in the control bias. Subjects may feel like having more control over dealing with phishing e-mails. They could knowingly follow a phishing link in the illusion of knowing how to control such deceptive websites. They may simultaneously underestimate the malicious behaviour such as drive-by downloads of malware or cryptocurrency miners. Attribution theory (see controllability dimension, Section 2.2) analyses the causation and control of *former* events. Here, the illusion of control of future events is of interest.

Confirmation Bias

is a selection bias to find evidence that confirms a subject's belief. This 'evidence' may come from better recalling confirming information or filtering unsuitable explanations. Although the scientific method supports a falsifiability approach[13], subjects try to confirm rather than falsify their believes, even if the same information is consulted by different subjects [Thi16]. Subjects may stay in their confirmation bubble when looking for security-related decisions. For instance, Apple users may be convinced that MacOS X is protected against any malware because they assume that 'a Mac is not a PC'. Hence, they do not need to take any security precautions. Apple users may suppress contradicting reports and cherry-pick confirming ones. Strand Consult [Str09] researched about iPhone users and found that they try to defend their view on the superiority of iPhones at all costs — they even related this behaviour to the *Stockholm Syndrome*. The confirmation bias consolidates and increases a preexisting *cognitive dissonance* instead of resolving this internal inconsistency.

Attribute Substitution

If solving a problem requires a lot of effort, subjects may substitute it intuitively with an easier heuristic problem instead of thinking of the original one [Thi16]. This cognitive bias called attribute substitution can result in a wrong solution to the original problem. Kahneman [Kah11] exemplified the effect as that beautiful people

[13] Philosophy of Science (Section 5.1) will discuss Popper's falsifiability claim.

are connotated more positively. Visual effects or appearances outweigh properties that are not easy to think about. An attractive attacker outshines drawbacks [Thi16]. For instance, Ryan and Mauch [RM10] created a fake profile of an attractive woman on several social networks and contacted targeted persons. They obtained in only 28 days various sensitive information (Section 4.5.1 about the 'Robin Sage' experiment). The bias *affect heuristic* resembles attribute substitution: it addresses the heuristic influence (system 1) of emotions such as "an overall good feeling toward a situation leads to a lower risk perception" [Sch08b]. Vice versa, bad feelings result in a higher risk perception.

2.7 Cultural Background

The previous sections mainly discussed the universal human nature, the first level of Hofstede's unique mental programming [Hof01], cf. Figure 2.2. Section 2.8 will elaborate on the personality aspect (third level) which is specific to each individual. The second level comprises culture as a learned part specific to a group of individuals [Hof01] and "not built on a genetic or inherited ground" [Ueb13b]. Culture creates a sense of collective identity, also known as *collective programming of the mind* [Hof01]. It shows the uniqueness of human groups, similar to personality which determines the uniqueness of a person [Sho18]. Evolutionary, "human customs and tribal rituals commonly give great emphasis to kinship; ancestor worship is widespread, family obligations and loyalties dominate much of life." [Daw06] Cultural norms are not isolated from human nature and personality. The Five-Factor Model (FFM) personality traits 'Extraversion' and 'Agreeableness' (Section 2.8.1) seem sensitive to the cultural context [Rol02]. The cultural background of subjects can help to design SE anecdotes (Anecdotes, Section 1.7) more comprehensible and as vivid as possible (Narrative Psychology, Section 1.8). That is, security policies and their example anecdotes need to be aligned with the cultural background. The previously mentioned Natural Semantic Metalanguage (NSM) in cross-linguistic research [GW04] can support this effort by using cultural scripts resulting in culturally comprehensible anecdotes.

As an exemplified approach for the cultural background, Hofstede's national culture dimensions [Hof14a] and published data were consulted:

2.7.1 National Cultures

The Hofstede Center [Hof14a] developed various approaches for cultural dimensions. Since published in the 1980's, this cultural values framework and its variants have been widely used in empirical research, ranging from psychology to business studies [KLG06]. Although culture is not limited to or unique inside national borders and cultural regions should be preferred, these national dimensions are deemed useful for researchers. The initial data were collected in 72 countries (N=88,000) [KLG06]. The public data per country originate from the Hofstede Center's website [Hof14a]. The dimensions cover indices for *Power Distance*, *Individualism*, *Masculinity*, *Uncertainty Avoidance*, and *Long-Term Orientation*. In an enhanced model from 2010 a sixth dimension appears, called *Indulgence vs. Restraint* (IVR), for which sufficient data for the diagram were not available

in 2013. The one used here expresses the five dimensions as depicted in Figure 2.4 for the countries Denmark, Germany, and Portugal [Ueb13b].[14]

Power Distance (PDI)
determines how much a subject accepts and expects unequally distributed power among members of institutions and organisations.

Individualism vs. Collectivism (IDV)
shows "the degree of interdependence a society maintains among its members." according to Hofstede Center [Hof14a]

Masculinity vs. Femininity (MAS)[15]
describes what subjects motivate and think what is important to achieve: to achieve the best financially and being assertive identifies as "masculine", liking what you do and caring for others is seen as "feminine".

Uncertainty Avoidance (UAI)
categorises national cultures in how subjects feel threatened, for instance, in "uncertain and ambiguous situations" [Hof14a]; to avoid such situations cultural groups may prefer "establishing more formal rules, not tolerating deviant ideas and behaviours"[16]

Long-Term vs. Short-Term Orientation (LTO)
(also known as 'Confucian dynamism') pictures how much a national culture focuses on future-oriented (persistence, thrift) or past- and present-oriented values (respect for tradition, fulfilling social obligations).

Figure 2.4 shows the huge uncertainty avoidance gap between Denmark and Portugal. Hofstede Center [Hof14a] described Portugal that on average they might feel an "emotional need for rules", see security as important, but also might resist innovation. Such rules may be expressed by the amount of security policies demanded. However, more security policies also result in the effort to keep them up-to-date and align them horizontally and vertically (high-level policies refined into corresponding low-level ones) [Dim12]. In comparison, Denmark scores lower and it may be assumed that misaligned or inconsistent security policies matter less [Ueb13b]. The behaviour influenced by the cultural dimension of uncertainty avoidance resembles the loss aversion behaviour and the resulting decision-making [TK74] of the universal human nature (Section 2.3.1).

The Hofstede Center dimensions were applied in the cybersecurity field: Shojaie [Sho18] researched the implementation and adaptation of Information Security Management Sys-

[14] The rationale behind selecting these countries was the location of partners in the TRE_SPASS project.
[15] Hofstede Center [Hof14a] has connotated one national culture dimension with the nowadays disputable terms of feminine and masculine.
[16] Geert Hofstede (1980b) as cited in Kirkman et al. [KLG06]

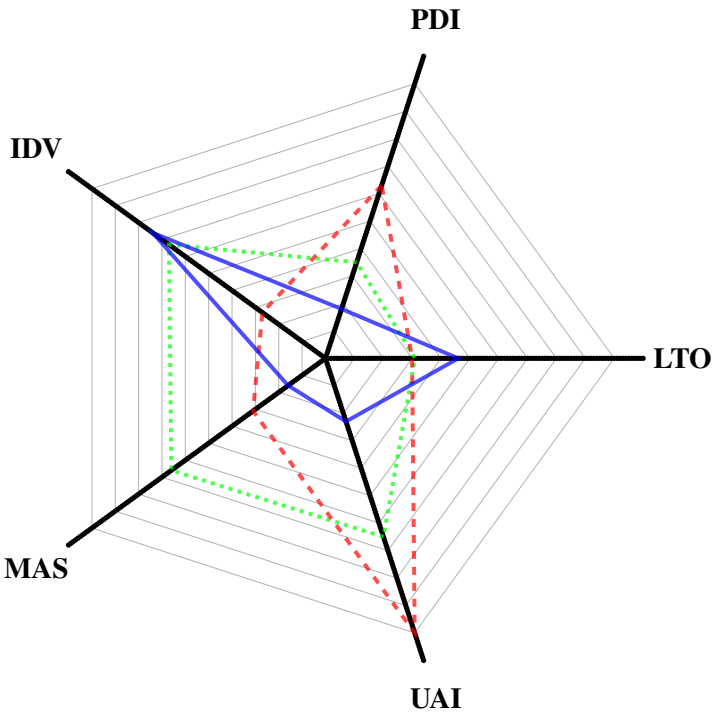


Figure 2.4: National culture comparison of Denmark [blue line], Germany [green dotted], Portugal [red dashed] [Ueb13b]; data retrieved from Hofstede Center [Hof14a]

tems (ISMS) (ISO/IEC 27001) in different cultures quantitatively. Shojaie et al. [SFS15] chose national characteristics split into cultural, political and economic characteristics complemented by organisational and personal aspects. For the cultural background they chose Hofstede Center's national dimensions of uncertainty avoidance (UAI), power distance (PDI), and individualism (IDV) comparing them to the amount of ISO/IEC 27001 implementations. Countries with a higher level of uncertainty avoidance, appear to implement more ISMS based on ISO/IEC 27001. Such correlations can enlighten the motives behind various security measures — sensible or not —, such as ISMS implementations.

2.8 Personality

"Personality is an individual's typical way of feeling, thinking, and acting. Given that personality is typical, it is fairly stable over time." [BV07] Personality is unique to each subject and sticks out from the human nature and culturally learned aspects as previously mentioned in the human mental programming from Hofstede [Hof01]. One characteristic of personality is that it stays relatively stable regarding feelings, thoughts, and behavioural patterns [MJ92]. Human personality research is an advancing field in psychology starting with in-depth clinical examinations. Shortened analyses are applied in other disciplines such as smartphone usage behaviour hinting to personality traits [CBG13; Mon+13]. Observing eye movement of subjects can identify a subject's personality [Hop+18]. Such observations (usage behaviour, eye movement) can be conducted without the subjects noticing it. A personality assessment of an unaware targeted person may show to which type of SE the targeted person is more susceptible. This can support either researchers

who want to correlate to personality traits in deceptive experiments or attackers who want to find a more promising attack vector. Personality-observation correlations are possible because personality and *social behaviour* (social psychology) influence each other [BV07]: knowledge of one's personality enables predictions about social behaviour which is defined as "a person's feelings, thoughts, or actions as he or she relates to other people" [BV07]. Extensive research exists about describing work-related personality, e.g., in the Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP) [Hos13; HP04], or commercial approaches to evaluate the personality of job applicants prior to an interview, such as the Predictive Index (PI) survey[17]. In the latter applicants choose adjectives about their behaviour from two perspectives: how they assess themselves and how they think others would assess them. This directs to the *Johari window* research of self-awareness ('Selbstwahrnehmung') and external perception ('Fremdwahrnehmung') to comprehend the relationship of both. In an organisational context these personality surveys and assessments may hint to suitable awareness trainings.

Ubiquitous and available technologies such as social networks foster technology-driven approaches as mentioned above. If automated personality assessments become more and more applied and the subjects are left in the dark, an ethical discussion must arrive. Shropshire et al. [Shr+06] applied one personality model to IT security. They chose personality assessments over attitudes towards technology for the following two reasons. They are relevant for similar research in cybersecurity if persistent assessments are needed.

(i) personality characteristics persist over time and enable longer term predictions ("Stability", also mentioned in the personality definition [BV07]);
(ii) personality is always measurable even when the subject is not aware of any technology ("Presence").

The International Personality Item Pool (IPIP) is a "scientific collaboration for the development of advanced measures of personality and other individual differences"[18]. In December 2020 they presented 3,320 personality items assigned to 463 scales. Validated questionnaires based on these scales exist with different granularity. There are various personality models developed over time, describing two models found in the IPIP here: the *Five-Factor Model (FFM)* (five dimensions) will be covered in Section 2.8.1, used later in the Social Engineering Personality Framework (SEPF), Section 4.6. The FFM was more prevalent in the author's scientific community and therefore chosen. To mention another model for complementary, future approaches, the *HEXACO* model developed by Ashton and Lee [AL07] deems useful. The HEXACO abbreviation stands for each of the six domains: **H**onesty-Humility, **E**motionality, e**X**traversion, **A**greeableness, **C**onscientiousness, **O**penness to Experience. In comparison to the below described FFM, the relatively recent HEXACO framework adds a moral domain (Honesty-Humility) containing scales of sincerity, fairness, greed avoidance, and modesty [AL07]. Ashton and Lee state that their model may also explain "reciprocal and kin altruism and the patterns of sex differences" [AL07] which the FFM may lack. It was offered, e.g., entities (humans, SSH probes) in a survey who logged into Secure Shell (SSH) honeypots ("Experiment 2: Surveying system trespassers") and who may have seen the limesurvey link in a welcome

[17] https://www.predictiveindex.com/
[18] https://ipip.ori.org/

message [Vet20, Section 3.4]. The moral domain may come in handy when investigating presumably malicious behaviour. Other psychometric scales like the Dark Triad (DT) can complement such efforts [Vet20]. Both personality models share the view of personality traits (domains, dimensions) in which a subject's personality might range diametrically in each dimension. They developed these traits as statistically independent as possible. These traits can be fine-grained into subtraits. The major personality evaluation found in the author's research disciplines is a subject's self-assessment via questionnaires. The HEXACO framework advertises, for instance, the revised HEXACO-PI-R questionnaire[19] published with self-reported as well as observer-reported forms.

2.8.1 The Five-Factor Model of Personality Traits

In the beginning of personality research and before it was called Five-Factor Model (FFM), over 18,000 personality-related terms were identified by Allport and Odbert (1936) [Shr+06]. This set was later refined and validated multiple times into five factors. The FFM is often also called the *Big 5* (B5) or *OCEAN*, analogue to the HEXACO abbreviation. McCrae and John [MJ92] published their FFM with the basic five dimensions in 1992. They provided adjectives and facet scales originating from observed and self-assessed analyses. The majority of self-assessments is conducted via questionnaires. Hirsh et al. [HKB12] reviewed the literature regarding persuasion and the *motivational system* behind each trait, later used in the Social Engineering Personality Framework (SEPF) [UQ14] (Section 4.6).

Openness,
sometimes referred to as 'Openness to Experience', "encompasses as a preference for creativity, flexibility, fantasy as well as an appreciation of new experiences and different ideas and beliefs" [UQ14].

Adjectives[20]	artistic, curious, imaginative, insightful, original, wide interests
Scales[21]	fantasy, aesthetics, feelings, actions, ideas, values
Motivation[22]	creativity, innovation, intellectual stimulation

Conscientiousness
focuses "on competence, self-discipline, self-control, persistence, and dutifulness as well as following standards and rules" [UQ14].

Adjectives[20]	efficient, organised, planful, reliable, responsible, thorough
Scales[21]	competence, order, dutifulness, achievement striving, self-discipline, deliberation
Motivation[22]	achievement, order, efficiency

[19] https://hexaco.org/hexaco-inventory
[20] Adjective Check List items created in a study by psychologists as *observers* [MJ92]
[21] *Self-reported* facet scales as found in the NEO-PI-R questionnaire [MJ92]
[22] The motivational system as summarised via literature review by Hirsh et al. [HKB12]

Extraversion

"comprises positive emotions, sociability, dominance, ambition, and excitement seeking" [UQ14].

Adjectives[20]	active, assertive, energetic, enthusiastic, outgoing, talkative
Scales[21]	warmth, gregariousness, assertiveness, activity, excitement seeking, positive emotions
Motivation[22]	rewards, social attention

Agreeableness

"includes compassion, cooperation, belief in the goodness of mankind, trustfulness, helpfulness, compliance, and straightforwardness" [UQ14].

Adjectives[20]	appreciative, forgiving, generous, kind, sympathetic, trusting
Scales[21]	trust, straightforwardness, altruism, compliance, modesty, tender-mindedness
Motivation[22]	communal goals, interpersonal harmony

Neuroticism

"describes the tendency to experience negative emotions, anxiety, pessimism, impulsiveness, vulnerability to stress, and personal insecurity" [UQ14].

Adjectives[20]	anxious, self-pitying, tense, touchy, unstable, worrying
Scales[21]	anxiety, hostility, depression, self-consciousness, impulsiveness, vulnerability
Motivation[22]	threats, uncertainty

Questionnaires for self-assessment exist such as the original NEO-PI, the revised NEO-PI-R (240 items), the updated NEO-PI-3, and the NEO-FFI-3 all refined by McCrae and John and also published as shortened versions. Hoppe et al. [Hop+18] used the NEO-FFI-3 (60 questions) for their eye movement and personality study. Depending on the goal of a questionnaire, the duration and level of detail must be balanced. When conducting SE experiments, a personality questionnaire of the test subjects may shed light on their susceptibility. A short questionnaire may persuade more subjects to answer it and avoid a 'questionnaire fatigue'. The very brief Ten Item Personality Inventory (TIPI) questionnaire [GRS03] asks two diametrical questions for each FFM trait. Metzner [Met17] used the TIPI (combined with other questionnaires) to evaluate the acceptance and wishes to report suspicious e-mails in an organisation. Another questionnaire (sample size of N=3,648) to understand the personality and cybercrime victimisation in the Netherlands, Van de Weijer and Leukfeldt [VL17] used a FFM scale (50 items) from the IPIP project. They showed that targeted persons with a high degree of the openness to experience trait are more likely to fall victim of cyber-enabled crimes. Shropshire et al. [Shr+06] were able to link conscientiousness and agreeableness to IT security compliant behaviour. Furthermore, cybersecurity training needs to be adjusted to different personality types which influences cybersecurity policy compliant behaviour (FFM study starting with 100 items, then reduced to 44; N=150) [WCM11]. Uebelacker [Ueb13b] mentioned that conscientious employees may also be more security conscientious with respect to security policy compliance and an organisational security culture due to self-discipline and occupational identification.

Concurrent validity of NEO-PI-R and the BIP showed that “both contributed significantly to the explanation of objective and subjective indicators of career success” [HSS06], i.e., the German BIP may become useful for research in security-related workplace behaviour. Summing up, personality traits have an impact on the security behaviour and susceptibility towards cybercrime.

3. Defining Social Engineering

The core element in this investigation is dealing with Social Engineering (SE). There is a need to determine which sources of information (anecdotes) express SE. Various SE definitions can be found. Some partially express the term in a way appropriate for this thesis. Shirey [RFC4949] even called their SE definition too vague and deprecated it (Definition 3.1). In discussions, the understanding of what comprises SE can range up to "everything is SE", that involves at least one targeted human being, i.e., every social interaction. Instead of vaguely deciding for each anecdote, this task was approached by setting criteria beforehand to find an appropriate definition.

> **Definition 3.1 — Social Engineering (Deprecated) (RFC4949).** "(D) Euphemism for non-technical or low-technology methods, often involving trickery or fraud, that are used to attack information systems. Example: phishing.
>
> Deprecated Term: IDOCs SHOULD NOT use this term; it is too vague. Instead, use a term that is specific with regard to the means of attack, e.g., blackmail, bribery, coercion, impersonation, intimidation, lying, or theft." [RFC4949]

The term "Social Engineering" is nothing new and not used in cybersecurity solely [And08]. The "Encyclopedia of Genetics, Genomics, Proteomics and Informatics" [Spr08] covers computer science, too. It describes SE from a genetical and societal view: genetics of individuals are excluded from that view, e.g., that genetics define each individual's (future) abilities in life. Only environmental elements like education, welfare, medical services etc. do form abilities. Hence, the active control of human behaviour must be institutionalised to *socially engineer* an individual to a functional member of the envisioned society. This utopian view ignores factors like individuality besides the importance of genetic predispositions [Spr08]. The Cambridge Dictionary sees SE very similar as "the artificial controlling or changing of the groups within society, usually according to particular political beliefs" [Cam18b]. These definitions do not fit in this thesis. The SE

Figure 3.1: Police mugshot of Wilhelm Voigt (1906)[3] and an equivalent uniform exhibited in the town hall of Berlin-Köpenick[4]

used in this thesis comes in various facets and does not constitute a new phenomenon at all. It just evolved to the digital realm with this new term. Some of the SE indicators defined later in Section 3.1 like deceptive, human interaction with malicious intent, can be found in historical events and famous fairy tales.

> **Anecdote 3.1 — The Captain of Köpenick (Zuc67).** Shoemaker Friedrich Wilhelm Voigt tricked people in 1906 by wearing a Prussian Guards officer's uniform. He told soldiers he acted in the name of the Emperor and was even able to command the local police. Questioning authority was not accepted in Prussian culture. He claimed that he suspected fraudulent bookkeeping of the city's treasure and had the mayor and treasurer arrested. Finally, he was able to 'confiscate' the town's treasure of around 3500 marks and left. [Zuc67]

"The Captain of Köpenick"[5] [Zuc67], a famous book and movie in Germany, is based on a real event that happened in 1906: a shoemaker impersonated a Prussian Guards officer by wearing the typical uniform, see Figure 3.1. He exploited the people's trust in uniforms and was able to obtain the municipal treasure (Anecdote 3.1). The shoemaker used the authority

[3] source: https://commons.wikimedia.org/wiki/File:Wilhelmvoigt.JPG (public domain, accessed on 2019-04-08)

[4] source: photo taken by User:Membeth (public domain, accessed on 2019-04-08), https://commons.wikimedia.org/wiki/File:Hauptmann_von_Koepenick_-_Uniform.jpg

[5] German: "Der Hauptmann von Köpenick" [Zuc67]; see also Wikipedia article [Wik19]

of hierarchy principle (Section 4.5.1 and Section 4.5.2) and successfully made others obey. This and similar stories[6] tell us that SE attacks are not novel attacks. Section 3.1.2 will explain why this incident was categorised as SE.

This chapter follows Research Question 1.1 to find a suitable SE definition for this interdisciplinary research. By initially describing indicators identifying SE in Section 3.1, existing SE definitions will be evaluated (Section 3.2) whether they are well defined in research literature.

3.1 Social Engineering Indicators

The following **Indicators** will be used to identify SE — all of which *must* be found in any attack incorporating SE. Further optional properties are listed in Section 3.1.1. The indicators here describe SE, the core element of a SE attack. The act of SE can be distinguished from the wider understanding of an attack involving SE. This was discussed for instance in the "Social Engineering Attack Framework" [MLV16]. The proposed SE indicators do not always exist isolated from each other. All reflect an obligatory component which can relate to another indicator. Thus, some aspects correlate, for instance, demanding an unaware targeted person (SE Indicator 3.3) and the use of deceptive techniques (SE Indicator 3.5). One sheds light on the targeted person, the other on the attacker's techniques; both revolve around deception and its detection.

The SE indicators developed here are intended, on one hand, to find suitable SE definitions (this chapter), on the other hand, to identify acts of SE in the wild. For the latter, the SE indicators foster structured observations of empiric sources to enable building a knowledge base (more in Section 5.3). Therefore, these indicators can create shareable knowledge and also help the discussion with fellow researchers of how SE can be confined. These SE indicators are the answer to Research Question 1.2.

In many definitions an attack starts and ends at one point in time, so does SE. SE begins when the attacker initiates a communication channel to the targeted person (SE Indicator 3.2). When the communication *and* the process to determine whether the targeted person fell victim terminates, the SE ends (SE Indicator 3.3). That is, the attacker can use the gained sensitive information later in an attack, but the act of SE is done. For instance, upon reading a phishing e-mail the SE starts. If the targeted person enters credit card details on a phishing website or ignores/deletes the message, SE finishes. Because a fixed *maximum* time frame is restricting and complicating the identification of SE, it will not be demanded.

Indicator 3.1 — Targeted Person: Human Enabler

Only by the action or reaction of at least one targeted person an attack can be successful. The targeted person (a *natural* person as in Definition 1.3) is therefore an obligatory enabler and is clearly a "human factor" (Definition 2.1). Information gathering like dumpster diving (Definition 3.2) or searching for all types of open source information

[6] While the story of the impostor of Köpenick is seen as a funny story, similar historical cases with lethal impact exist such as the "Standgericht Herold" (Anecdote 4.12) [Wös15]

can be pre-attack elements, but are not SE if no enabling action of a targeted person is involved.[7] Hence, this may be part of a SE attack, but not SE itself. Information gathering via shoulder surfing (Definition 3.3) does not clearly indicate whether it was enabled by deceiving the targeted person into inserting their credentials while standing behind. Research exists to even mitigate this using EEG-based authentication schemes in brain-computer interfaces [GMK19]. Although 'shoulder' may be understood as eavesdropping nearby and in-person, the reception of the display signals with photosensors would be possible in farther distances when CRT displays were more widely used [Kuh02]. That is, sensitive information on the screen could be acquired from a distance, hence, an interaction with the targeted person has most likely not happened if no prior intentional communication to display the information was initiated. Information about possible targeted persons can be used to prepare SE, e.g., to "identify potentially gullible individuals" [Bar14, Section 9.4] (gullibility in Section 4.3.1) or to collect personal information for more effective spear phishing attacks (Definition 4.4).

Definition 3.2 — Dumpster Diving (Bha07). "A technique adopted by social engineers that involves physically searching through trash in dumpsters in an attempt to retrieve useful information prior to launching a social engineering attack." [Bha07]

Definition 3.3 — Shoulder Surfing (Har13). "Viewing information in an unauthorized manner by looking over the shoulder of someone else." [Har13, p. 25, Key Terms]

Indicator 3.2 — Attacker: Intentional Communication

A Social Engineer initiates directly or indirectly a communication channel to the targeted person or persons. The communication is intentional and is not initiated by accident. This communicative interaction can consist of one step (unidirectional) where the attacker contacts the targeted person and by the intended reaction the person falls victim, e.g., a one-time phishing e-mail. It can comprise more steps (bidirectional) where the attacker is talking or writing to the targeted person, for instance, in romance scams (Definition 4.2). The attacker does not need to initiate each attack separately: any bulk e-mail phishing communication is started intentionally as well.

Mouton et al.'s SE ontology [Mou+14a] (Figure 3.2) can be used: a direct communication channel is *bidirectional* if the attacker and targeted person communicate in both ways; *unidirectional* where no answer of the targeted person is expected, except revealing the valuable information. Furthermore, *indirect* communication can occur where no actual interaction happens. The *road apple attack* [PDP13] expresses such a case: a USB flash drive left in publicly accessible location (parking lot) of an organisation can be carried by a random employee to the office space. Once plugged into an organisation's computer, malware may infect the system circumventing any firewall protection from the outside (Anecdote 4.4). Here, the communication is intentionally initiated (not by accident), but

[7] Human Intelligence (HUMINT) is defined by NATO Standardization Office [NATO19] as "Intelligence derived from information collected by human operators and primarily provided by human sources." However, it is unclear how actively human operators gather such information and if SE may be involved.

not directed towards a specific targeted person (indirect).

Regarding the domains of Socio-Technical Systemss it can be said that the communication happens in the digital (e.g., e-mail) or physical domain/medium (e.g., in person). As the communication is between attacker and targeted person in the social domain, one can say that SE Indicator 3.2 shows that SE involves the social domain plus at least one of the other two socio-technical domains.

Besides this bilateral constellation (attacker $\leftrightarrow$ targeted person), SE can occur in more complex situations. For instance, a phisher sends phishing e-mails via a paid spamming service to a targeted person. If successfully phished, the targeted person transfers money to a third party (money mule) who then moves the money to other accounts via MoneyGram. In this case, the phisher is the one initiating the communication to the targeted person, hence, calling the phisher 'attacker' here. The phisher manipulates and deceives. The money mule (Section 7.2.1) is often just the middle person for money laundering (Definition 3.4) and does not use deceptive techniques etc. The targeted person may fall victim because of the phisher's actions. All stakeholders would be part of a SE *attack*, but not of the sole act of SE. Moreover, not discussed here is the facet whether the money mule is intentionally laundering money or not, meaning being unaware of being exploited as well (and becoming another targeted person).

In general, SE does not exist without any direct or indirect human communication; concluding that each act of SE *must* have at least one attacker and at least one targeted person. Some definitions talk about an organisation as a SE target; organisations are made of people, hence, this is SE because at the end the attacker communicates with a natural (not legal) individual.

> **Definition 3.4 — Money Laundering (Wod07).** "This term is defined as an intentional committed offense with the purpose of concealing or disguising of the true origin, the nature, the disposition, or of the controlling rights of properties, which were acquired illegally." [Wod07]

Indicator 3.3 — Targeted Person: Unawareness

At the time SE reaches the targeted person, this person does not know anything of being targeted maliciously. The targeted person may hesitate regarding the attacker's request and may comply for various reasons willingly, but must be unaware of the attacker's true intentions during this process. That is, after the end of the SE act (see above), the targeted person may very well recognise it whether successful or not. Together with the attacker's technique to deceive (SE Indicator 3.5), the targeted person must be *unaware* to enable a *successful* SE attack (SE Indicator 3.1). Hence, this SE indicator shows the other side of the coin with focus on the targeted person. Therefore, blackmailing, bribery or coercion do not fit this SE indicator and are not SE because the attacker does not abuse the unawareness of the targeted person. If a targeted person becomes aware of a SE attempt before succumbing to it, it expresses a mitigated SE attack (and still is SE).

Indicator 3.4 — Attacker: Malicious Intent with Goal
At least one person *must* aim at a targeted person with malicious intent. This attacker means harm to the targeted person, some other person or some organisation. That is, a targeted person *must* be attacked (to enable SE, SE Indicator 3.1), but harm may be done to an organisation or other person and not the targeted person itself. Maliciousness is meant from the point of view of the targeted persons or their associated organisations. In the human error classification of Reason [Rea90] (Figure 2.1) the malicious intent resembles the intended violations category of unsafe acts.

Some definitions explicitly define possible goals, cf. Definition 3.6. Malicious intent implies that the attacker has a goal to achieve. These two components are intrinsically tied; for clarity reasons the author added the "goal" to the indicator name. A goal is anything of value to the attacker, where "value" does not mean monetary interest solely. Goals may cover accessing sensitive information, financial gain (Financially-Motivated Social Engineering (FMSE) [Ver19]) including retrieving valuable items, vandalism, hate crime, destroying reputation, attacker's publicity, attacker's entertainment or even terrorism [MS02, p. 10]. They can violate confidentiality, integrity and availability of data and services.

An attacker can target people to gain useful sensitive information or unauthorised access for instance. Useful information can be credentials like passphrases, but also confidential information enabling future attacks beyond SE. For instance, an attacker may make use of acquired sensitive information through successful SE by extorting money from an organisation. The attacker can offer not to leak the information.

With respect to targeted persons and referring to the classic paradigm for authentication mentioned in National Institute of Standards and Technology (NIST) Special Publication 800-63-3 [GGF17, Section 4.3.1 Authenticators]: an attacker may be interested in something the targeted person *knows* like a passphrase. Something a targeted person *possesses* ("has"), e.g., a key or security token. Or something a targeted person *is*, meaning things inseparable from the person such as biometrics, for instance, luring a targeted person into using the fingerprint for granting the attacker access. Thus, the attacker can manipulate the targeted person to reveal something in each of the three socio-technical domains.

For experimental research in SE, the role of an attacker may change into a scientist with hopefully ethical precautions. Scientists must mimic malicious intent as they will not fulfil the entire malice endeavour until the end depending on the performed SE. Mimicking may lead to a different outcome of the experiment and ethical considerations (Section 5.2.1) may limit the experiment options. Hence, insights of how to perform ethical SE experiments that do not divert the outcome are crucial, e.g., in SE penetration tests [Dim+10].

The goal of education, especially of children, should be to intentionally influence people about good and bad behaviour, even if they will not comprehend it's purpose entirely, viz. SE Indicator 3.3 (unawareness). Often, the parental explanation is coined "for your own good" [MS02] talking to their children. The author disagrees with Mitnick and Simon's statement that "we were all moulded by our parents: benevolent (and sometimes not so benevolent) social engineers" [MS02, pp. 10–11], see also Springer [Spr08]. SE *must* contain malicious intent. Parenting with good intentions — even if they turn out bad by accident — cannot be SE.

Intentional bad parenting and education however, can express SE if the other indicators are present. So is subliminal, targeted advertisement. It may fall under this indicator and become SE if the rest of the indicators, esp. SE Indicator 3.3 (unawareness), can be found, e.g., by using viral marketing techniques and social media influencers. They could count as malicious if the decision proposed by the advertiser is harming the targeted person as a side effect, e.g., financially or the wellbeing. In the 1920s, Edward Bernays staged a campaign promoting women to smoke in public, deceiving the audience that the employed actresses were part of the Women's Liberation Front (Anecdote 4.3). Another aspect are introducing stronger addictive elements in computer games that may target unaware players.[8]

Indicator 3.5 — Attacker: Deceptive Techniques
Each act of SE *must* comprise one or more techniques the attacker uses to target a person. These techniques use deception as one element (Section 4.4). The unawareness SE Indicator 3.3 reflects this point from the targeted person's perspective. Deceptive techniques consist of at least one of the persuasion principles (Section 4.5). Persuasion principles in these techniques can be applied iff the principle is used deceptively. Because many similar terms exist, the following will be synonyms for deceptive techniques: manipulation and tricking techniques. Concerning the specificity of later discussed SE definitions: whether a definition is describing these techniques in detail or not, will not count as "**p**artially" covered. The mere fact of mentioning one suffices.

While phishing (Section 4.4.3) is a deceptive technique to trick targeted persons, phishing e-mails can contain persuasion principles to convince the targeted person of their authenticity and truthfulness. For instance, very similar to the advance-fee scam technique, targeted persons may be asked to transfer bitcoins to a specific address to then receive more bitcoins in return (see Bitcoin Doubler scam [Smi20]). Principles may be 'Scarcity' (only a few transactions left), 'Authority by Knowledge' (recommendation from a self-proclaimed bitcoin expert) or other (Section 4.5).

3.1.1 Non-Obligatory Properties

SE may have more properties as follows, but they are not obligatory.

In 1994 Friestad and Wright [FW94] defined the Persuasion Knowledge Model focusing on the knowledge that an attacker ("Agent") and targeted person ("Target") possess. The attacker and targeted person know of a specific topic, about persuasion, and of each other which they can use for attacking or mitigation [FW94]. It may partly cover a SE attack in the sense of SE Indicator 3.2 (communication), but elaborates more on a *knowledge property* than on the communication level (Section 4.1).

To emphasise: organisations are made of people (SE Indicator 3.2). For the SE indicators people are targeted, even if some definitions declare organisations as targets. As a legal person, an organisation cannot be deceived or have awareness, but organisational policies

[8] A recent court case about the computer game 'Fortnite' covers a possible 'Gaming Disorder' (WHO ICD-11 6C51) caused by intentionally crafted game elements to make children addictive. The lawyers claim that the producer Epic Games invested excessively in psychological studies to make Fortnite more addictive [Ols19]. (see also 'Dark Patterns' [TAB30] in Section 4.4)

can be violated by targeted persons. These definitions are understood insofar that the attacker's goal is related to an organisation, thus, assets of the organisation are attacked. However, targeted persons belong to an organisation and the attacker targets its members using SE. Therefore, an organisation can become a non-obligatory property in a SE attempt.

Whether the attacker is an insider or not, does not bother us here, but can provide insights for mitigation purposes (Insiderness in Section 4.2). The targeted person can be anyone; hence, no restrictions to the SE indicators were added. Also, whether the targeted person is complying with an attacker's request by *knowingly* (knowing of the violation) or *unknowingly* violating policies, does not matter when also being unaware (SE Indicator 3.3). Thus, a property can exist that covers whether the targeted persons comply to SE knowingly or not. Regarding the medium used to conduct the SE, all possible ways without limitation to any socio-technical domain are accepted.

3.1.2 Back to the Köpenick Anecdote

Anecdote 3.1 described the attack of shoemaker Friedrich Wilhelm Voigt. The SE Indicator 3.1 enabled the attack because without the cooperation of others fooled by Voigt, the attack would not have been possible and hence successful. He communicated directly with the targeted persons (SE Indicator 3.2). His goal was to retrieve the municipal treasure illegally (SE Indicator 3.4). Others were unaware of the deception and trusted him (SE Indicator 3.3). He applied at least the persuasion principle called "Authority" (Section 4.5.1) based on hierarchy by impersonating a uniformed authority figure (SE Indicator 3.5). All five indicators are found in this anecdote, i.e., it constitutes SE.

3.2 Discussion of Various SE Definitions

In the following section various SE definitions will be discussed and compared based on the SE indicators. Table 3.3 will show an overview. The terms listed in Table 3.1 will be used to categorise how *specific* a definition expresses one indicator. The terms will merely describe how good a definition covers each indicator. If all indicators are entirely met, the definition is useful to identify SE evidence. The optimum is '**y**es' and still acceptable is '**i**mplicit'. To understand how much a definition *qualitatively* covers one indicator, this simple ordering may help:

$$\textbf{m}\text{issing} < \textbf{p}\text{artial} < \textbf{i}\text{mplicit} < \underline{\textbf{y}\text{es}} < \textbf{t}\text{oo general}$$

First of all, I will discuss Mouton et al.'s extensive work.

3.2.1 Mouton et al. (Mou+14a)

Mouton et al.'s work on SE comes very close to the propositions presented here. Their extensive work on SE is conducted thoroughly: they presented two Social Engineering Attack Detection Models [BMV10; MLV15], discussed normative ethics [MMV13], formulated an ontological model of the SE domain [Mou+14a], created a Social Engineering Attack Framework [Mou+14b] as well as Social Engineering Attack Examples, Templates, and Scenarios [MLV16]. They distinguish between the act of SE (Definition 3.5) and SE attacks (Definition 3.6) like proposed in this thesis. They phrased the definitions in their

Specificity	Description
missing	indicator is missing completely; not even implicitly mentioned
partial	only a subset of the indicator is described
implicit	implicitly identifying an indicator; not mentioned explicitly
yes	indicator is present entirely
too general	definition exceeds scope of indicator

Table 3.1: Specificity of each SE indicator per definition

ontology based on a survey of fifteen other definitions. Their ontological model as depicted in Figure 3.2 shows the components of SE attacks.

> **Definition 3.5 — Social Engineering (Mou+14a).** "The science of using social interaction as a means to persuade an individual or an organisation to comply with a specific request from an attacker where either the social interaction, the persuasion or the request involves a computer-related entity." [Mou+14a]

Their definition of SE (Definition 3.5 in Mouton et al. [Mou+14a]) needs a few clarifications. The involvement of a "computer-related entity" restricts the identification of SE attempts too much. Therefore, no such restriction were applied to the SE indicators, cf. SE Indicator 3.2 (socio-technical domains in intentional communication) and Section 3.1.1 (medium in non-obligatory properties). The grandparent scam is well-known in some countries (Germany: 'Enkeltrick', Austria: 'Neffentrick'). It can involve a "computer-related" entity like a telephone (Anecdote 4.1, Anecdote 4.10), but may also occur in person at the front door. The attacker persuades the targeted person to hand-over valuables for an alleged relative in an emergency. Or the valuables will be kept safe by the police because an alleged robbery is expected. The attacker is unknown to the mostly elderly person, but is able to convince and deceive the targeted person of borrowing cash 'temporarily'.

Regarding SE Indicator 3.1, "social interaction" comprises a human enabler. But how can an attacker persuade an organisation? For SE Indicator 3.2 (communication) and SE Indicator 3.4 (malicious intent) an interaction between attacker and targeted person can be demanded. Organisations can be targeted through individuals because they are legal persons run by real individuals. Hence, their definition can be understood as complying with SE Indicator 3.1 (human enabler). Clearly a communication is initiated intentionally (SE Indicator 3.2). They are using the term "attacker" which can be interpreted as malicious intent (SE Indicator 3.4); however, malicious intent is not mentioned explicitly, but can be assumed. It is not clear whether the targeted person is aware of the deceit, marking SE Indicator 3.3 as partially found. Deceptive Techniques (SE Indicator 3.5) are present because the targeted person is complying with an attacker's request.

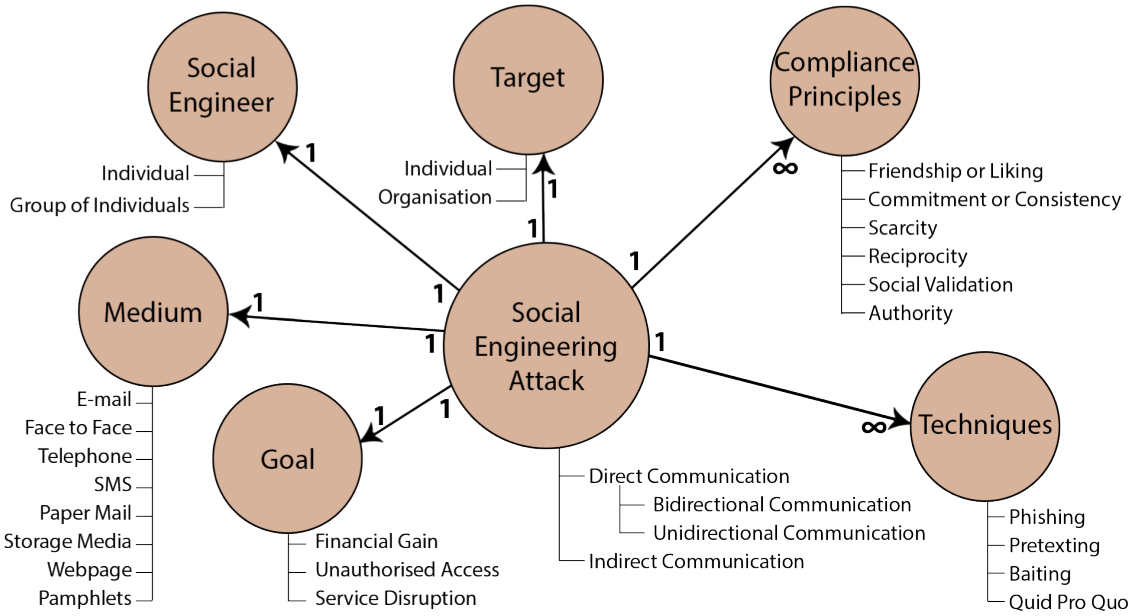


Figure 3.2: Ontological model of SE attacks [Mou+14a]

Ontological Model of Social Engineering Attacks (Mou+14a)

In their SE *attack* definition (Definition 3.6), the intentional communication SE Indicator 3.2 is presented in more detail. Other components are mentioned and better viewed in their ontological model diagram (Figure 3.2).

> **Definition 3.6 — Social Engineering Attack (Mou+14a).** "A Social Engineering attack employs either direct communication or indirect communication, and has a social engineer, a target, a medium, a goal, one or more compliance principles and one or more techniques." [Mou+14a]

To observe how the five SE indicators can be found in this attack model, each component will be checked. Although their definition handles a SE *attack* and not just the act itself, it shows which components are interdependent.

A **Social Engineer** as an attacker must be present as well as the **Target** (targeted person). The human enabler (SE Indicator 3.1) from Mouton et al.'s Definition 3.5 is part of the model because it is an obligatory component (**Target**). However, the diagram specifies the **Target** as either an individual (natural person) or an organisation (legal person). This non-human **Target** exists in their ontology, but would not express SE according to the SE Indicator 3.1 (natural person, Definition 1.3) postulated for this thesis. Using their notation as in **Techniques**, the amount of **Social Engineers** should have a value of infinity ∞ to stay consistent, not the proposed "Group of Individuals". The term **Target** resembles this: although the goal can be to retrieve an asset which resides in an organisation, the "Individual" is finally the target being deceived. Direct or indirect communication (not in a diagram circle, relationship unknown) is used between those two entities (SE Indicator 3.2). The **Medium** as part of the communication channel is here limited to one medium (1:1 relationship). This is plausible if more than one medium means another SE attack. The set

of media should be an extendable list of examples.

Their extendable list of **Compliance Principles** matches the proposed persuasion principles (Section 4.5). These principles are intertwined inextricably with the **Techniques** (SE Indicator 3.5). The list can be seen as an extendable set of examples. The fact of a targeted person unaware of an attack (SE Indicator 3.3) can be interpreted as part of their **Compliance Principles**. "The goal of an attack can be financial gain, unauthorised access or service disruption." [MLV16] If this **Goal** list is extendable, it would fit except that it should use the infinity symbol ∞ for more than one goal. That is, a **Goal** can be a "Financial Gain" (FMSE [Ver19]), but simultaneously "Service Disruption" if the organisation runs into financial problems paying their bills and cannot provide their services. A goal does not need to be unique, but it can clearly assumed as malicious intent (SE Indicator 3.4).

Finally, all of the proposed SE indicators exist in Mouton et al.'s ontological model, although the model seems to need clarifications, such as how an organisation (legal person) and not an individual in that organisation can be phished.

3.2.2 Other Social Engineering Definitions

After the previously elaborated definitions, for the following definitions a more comprehensive overview in Table 3.3 will be offered.

Harris (Har13)

Harris [Har13] explains in the CISSP[9] exam guide SE as "nontechnical attacks" executed by using "persuasion, coercion (rubber-hose cryptanalysis), or bribery (purchase-key attack)" [Har13, p. 869]. He defines it in key terms and techniques throughout the book combined here in one definition as follows.

> **Definition 3.7 — Social Engineering (Har13).** "Gaining unauthorized access by tricking someone into divulging sensitive information."
>
> [Har13, p. 25, Key Terms]
>
> "An attacker falsely convinces an individual that she has the necessary authorization to access specific resources." [Har13, p. 193, Techniques]
>
> "Manipulating individuals so that they will divulge confidential information, rather than by breaking in or using technical cracking techniques." [Har13, p. 870, Key Terms]

Although Harris describes SE (Definition 3.7) with tricking, falsely convincing, or manipulating someone (SE Indicators 3.2 and 3.5), the explanation in Harris [Har13, p. 869] suggests that even bribery and coercion can be part of SE. That is, the targeted person can be an enabler (SE Indicator 3.1) of an attack *knowingly*. This view violates SE Indicator 3.3 and, thus, cannot express the SE used in this thesis. Getting unauthorised access or sensitive information are definitely malicious goals, therefore, SE Indicator 3.4 matches.

Kiltz et al. (KLD07)

Kiltz et al. [KLD07] implicitly mention an intentional communication to the targeted person in Definition 3.8, but do not name an attacker explicitly (SE Indicator 3.2). The targeted

[9] Certified Information Systems Security Professional (CISSP)

person complies by executing anything normally not doing and enabling successful SE (SE Indicator 3.1). Their examples like "fear of reprisals" do not clearly state whether the targeted person is aware of SE, the SE Indicator 3.3 would be more restrictive (less general). Thus, deceptive techniques exist ("trick"), but their scope covers other non-deceptive techniques as well and becomes too general (SE Indicator 3.5). To *exploit* someone is too vague, too, and malicious intent is not mentioned explicitly (SE Indicator 3.4).

> **Definition 3.8 — Social Engineering (KLD07).** "This refers to finding means to trick or pressure a person into doing things that they would not do under normal circumstances. Thereby certain traits of a person are exploited, such as the willingness to help or the fear of reprisals." [KLD07]

Fagnot (Fag07)

Definition 3.9 describes the manipulation of end-users who enable successful SE (SE Indicator 3.1). But Fagnot [Fag07] restricts the group of targeted persons to "IT end users" — a subset of potentially targeted persons as seen in this thesis (SE Indicator 3.2). All targeted persons are unaware of the attack (SE Indicator 3.3). Fagnot [Fag07] mentions malicious intent explicitly. Yet the definition focuses on retrieving sensitive information only which is just one goal beside others (SE Indicator 3.4). The mentioned "manipulation" suits SE Indicator 3.5 (deception).

> **Definition 3.9 — Social Engineering (Fag07).** "The act of manipulating IT ends users to obtain sensitive information. It is usually conducted over the phone and involves several end users so that none of them suspect the malicious intent." [Fag07]

Jeske and Schaik (JS17)

The human enabler requirement from SE Indicator 3.1 is fulfilled in Definition 3.10. Intentional communication between "criminals" and "victims" exist implicitly because it was not mentioned explicitly (SE Indicator 3.2). First obtaining information seems to be a malicious goal, but in the second sentence "breaking normal security procedures" generalises this a bit more; SE Indicator 3.4 ("Criminals") is identified. Here again "manipulation" and trickery matches SE Indicator 3.5 which also means the targeted person is unaware of the act (SE Indicator 3.3).

> **Definition 3.10 — Social Engineering (JS17).** "The act of manipulating individuals to divulge confidential information. Criminals usually try to trick their victims into breaking normal security procedures and releasing valuable information such as passwords and bank details." [JS17]

Grassi et al. (GGF17)

In NIST's special publication SP800-63-3 Grassi et al. [GGF17] define SE as follows.

> **Definition 3.11 — Social Engineering (GGF17).** "The act of deceiving an individual into revealing sensitive information, obtaining unauthorized access, or committing fraud by associating with the individual to gain confidence and trust." [GGF17]

SE Indicator 3.1 requirement is clearly met: the targeted individual is the enabler. The

communication is intentional and implicitly mentioned by using deception (SE Indicator 3.2). The same applies to a person unaware of SE (SE Indicator 3.3). Although an attacker is not directly named, an attacker is needed for maliciously deceiving the targeted person (SE Indicator 3.4). Possible attacker's goals are described and are not sensitive information solely like in other definitions. In Definition 3.11 deception and principles like building trusted relationships are described (SE Indicator 3.5).

Bundesamt für Verfassungsschutz (BfV16)

Germany's Federal Office for the Protection of the Constitution (Bundesamt für Verfassungsschutz) publishes annually a report called "Verfassungsschutzbericht". The report for 2016 defines SE as cited in Definition 3.12 [BfV16, p. 266].

> **Definition 3.12 — Social Engineering (BfV16).** "Ausspionieren des persönlichen Umfelds durch zwischenmenschliche Beeinflussung bzw. geschickte Fragestellung, meist unter Verschleierung der eigenen Identität (Verwenden einer Legende). Social Engineering hat das Ziel, unberechtigt an Daten, geheime Informationen, Dienstleistungen oder Gegenstände zu gelangen." [BfV16]

The human enabler (SE Indicator 3.1) is present ("zwischenmenschlich"). The communication is initiated intentionally (SE Indicator 3.2) as prerequisite for the interaction ("zwischenmenschlich", "Fragestellung"). Whether the targeted person is aware or not, is not expressed explicitly, but assumed (SE Indicator 3.3). The attacker shows malicious intent ("unberechtigt", SE Indicator 3.4), although the listed goals can be extended to vandalism or entertainment (still counting this as "yes"). It is nice to find "items" ("Gegenstände") as one potential goal. The term "Beeinflussung" is more general and covers deception only partially (SE Indicator 3.5); the example of pre-texting/impersonation clarifies only a bit.

Hadnagy (Had10)

Hadnagy's definition [Had10] from his book *Social Engineering: The Art of Human Hacking* was used in the paper "The Social Engineering Personality Framework" [UQ14]. With respect to the above SE indicators developed after that paper, the definition does not cover all aspects entirely.

> **Definition 3.13 — Social Engineering (Had10).** "the act of manipulating a person to take an action that may or may not be in the target's best interest. This may include obtaining information, gaining access, or getting the target to take certain action" [Had10]

The targeted person's action defines the outcome, hence, becoming a human enabler (SE Indicator 3.1). It is not clear how the targeted person is being manipulated. Also, an "attacker" who intentionally initiates a communication is not mentioned explicitly, but can be assumed (SE Indicator 3.2). Manipulation is a hint to an unaware targeted person, but not explicitly written down (SE Indicator 3.3). The malicious intent of an attacker is missing. Whether the manipulation results are harmful for the targeted person is optional; it can be harmful for the not mentioned organisation (SE Indicator 3.4). On the other side, he shows well-known examples of possible goals, but this makes this aspect too general.

The attacker may probably apply deceptive techniques because of the term "manipulating" (SE Indicator 3.5), however, not naming any deceptive techniques (too general). Hadnagy's use of "may" results in a vague definition.

On Hadnagy's website the SE definition[10] is abbreviated and expresses the meaning of the first sentence. The missing last sentence covers only the goal examples in Hadnagy [Had10]. This does not change the categorisation of the SE indicators.

Mitnick and Simon (MS02)

When it comes to SE, the classic references are Mitnick and Simon's books *The Art of Deception: Controlling the Human Element of Security* [MS02] and *The Art of Intrusion: The Real Stories Behind the Exploits of Hackers, Intruders & Deceivers* [MS05]. The Definition 3.14 from Mitnick and Simon [MS02] will be examined.

> **Definition 3.14 — Social Engineering (MS02).** "Social engineering uses influence and persuasion to deceive people by convincing them that the social engineer is someone he is not, or by manipulation. As a result, the social engineer is able to take advantage of people to obtain information with or without the use of technology." [MS02]

"To take advantage of people" by persuasion and manipulation fulfils the proposed SE Indicator 3.1 of the human enabler. It is assumed that implicitly the attacker initiates some kind of communication (SE Indicator 3.2). The term "deception" is a strong indicator for intentionally keeping a targeted person unaware (SE Indicator 3.3) as well as for applying appropriate techniques (SE Indicator 3.5). The goal can comprise anything of value to the attacker and that the process of achieving it is not limited to technology (SE Indicator 3.4). Regarding the latter, Mitnick and Simon are explicitly including the use without technology. However, to retrieve information is too restrictive as the sole goal. Because malicious intent is definitely found, but information retrieval the only goal, SE Indicator 3.4 are aligned with Definition 3.14.

Reverse Social Engineering

To clarify SE Indicator 3.2 about intentional communication initiated by the attacker regarding Mitnick and Simon's other Definition 3.15 on *Reverse* SE: setting up a situation where then the targeted person contacts the attacker counts as indirect communication. The same view expresses Gragg [Gra02] where the attacker causes a network problem, for instance, gets contacted by the targeted person and fixes the issue. After that the attacker may have gained trust for the real attack. However, causing such a network problem requires additional effort for the attacker [Gra02]. This first part can express SE and is also used in other papers.[11]

[10] "Any act that influences a person to take an action that may or may not be in their best interest.", https://www.social-engineer.org/about/ (last accessed 2019-03-26)

[11] "In a reverse social engineering attack, the attacker does not initiate contact with the victim. Rather, the victim is tricked into contacting the attacker herself." [Ira+11]

Definition 3.15 — Reverse Social Engineering (MS02). "A social engineering attack in which the attacker sets up a situation where the victim encounters a problem and contacts the attacker for help. Another form of reverse social engineering turns the tables on the attacker. The target recognizes the attack, and uses psychological principles of influence to draw out as much information as possible from the attacker so that the business can safeguard targeted assets." [MS02]

Mitnick and Simon continue with a second, completely different aspect mingled together in Definition 3.15, i.e., overloading the previous aspect. There, the targeted person detects a SE attack and reverses the attack, trying to attack the attacker with SE, cf. Persuasion Knowledge Model (Section 4.1). This means the targeted person becomes an attacker and uses SE. The author sees this not as a reversed 'novel' act, but a new SE attack starting when the communication reaches the former attacker, hence, a counterattack. However, the communication was intentionally initiated by the primary attacker and intentionally re-used by the counterattacker (SE Indicator 3.2). The counterattack is 'malicious' in the point of view of the former attacker, e.g., by facing prosecution (SE Indicator 3.4). Anecdote 3.2 tells the story of a fake police officer. The attacker tried to lure elderly people into handing over their money for safekeeping. An old man detected the SE attack and alerted the local police who arrested the couriers and the suspected caller [NDR20c].

Anecdote 3.2 — Counterattack on Fake Police Request. In January 2020, an 80 year old man in Hamburg was contacted via phone by an impersonated police officer to safekeep his money. The caller claimed that the police found his address on a burglar's list, suggesting that it contained future targets. The targeted person detected the attack and cooperated with the local police. He played along. The two persons collecting the money were arrested upon arrival at the old man's house. They claimed to be unaware of such fraud and were solely the couriers. The old man was able to recognise the voice of the phone caller in the police database. The suspected caller was arrested in Turkey. [NDR20c]

BSI (BSI15)

The German Federal Office for Information Security (BSI) provides "recommendations for standard security safeguards" like the *IT-Grundschutz Catalogues* [BSI15]. Here the latest English version from 2015 were chosen and not the most recent German one [BSI19b]. The introduction of threat G 0.42, Chapter G 0 "Elementare Gefährdungen", defines SE (Definition 3.16). Threat G 5.42 [BSI15] in Chapter G 5 ("Vorsätzliche Handlungen") also defines SE, but in a slightly different way: "social action" is substituted by "listening in". The SE description in M 5.150 [BSI15, p. 4319] is not as good covering the SE indicators as G 0.42. Therefore, G 0.42 will be used.

Definition 3.16 — Social Engineering (BSI15). "Social engineering is a method used to gain unauthorised access to information or IT systems by social action. Social engineering exploits human characteristics such as the willingness to help others, trust, fear, or respect for authority. Employees can be manipulated using social engineering so that they perform unauthorised tasks." (G 0.42) [BSI15]

SE Indicator 3.1	Targeted Person: Human Enabler
SE Indicator 3.2	Attacker: Intentional Communication
SE Indicator 3.3	Targeted Person: Unawareness
SE Indicator 3.4	Attacker: Malicious Intent with Goal
SE Indicator 3.5	Attacker: Deceptive Techniques

Table 3.2: Overview of SE indicators from Section 3.1

The exploitation of "human characteristics" makes clear that the enabler of successful SE is the targeted person (SE Indicator 3.1). The missing role of an attacker impedes finding the intentional communication SE Indicator 3.2. The "social action" shows that there exists an interaction and the explanation later in G 0.42 discusses this role. Hence, this will count as implicit. The unawareness SE Indicator 3.3 is not expressed explicitly, but "manipulation" and "exploits" hints to an implicit presence. Although, "fear" may indicate possible blackmailing or coercion (aware targeted person). Malicious intents (SE Indicator 3.4) are found as in to provide the attacker "unauthorised access" or the targeted person may even be persuaded to perform "unauthorised tasks". SE Indicator 3.5 fits into the terms "manipulation" and exploitation as well as into the example principles.

3.2.3 Definitions Overview

Table 3.2 lists the indicators of Section 3.1 for a smoother comparison. In Table 3.3 the findings of the previous definitions will be summarised.

3.2.4 Conclusion

None of the above definitions in Table 3.3 expressed entirely and explicitly all of the five SE indicators. With respect to Research Question 1.1 no suitable SE definition was found. As this is a founding part for this thesis to identify SE in, e.g., anecdotes, a suitable normative definition (Definition 3.17) needed to be developed as follows.

3.3 Uebelacker's Social Engineering Definition

Based on SE Indicator 3.1 to SE Indicator 3.5 the SE definition will be formulated. Referenced SE indicators are shown with words underlined. All SE indicators are present and mentioned explicitly. Key words in uppercase are used according to BCP 14[12] [RFC2119; RFC8174]. The terms "attacker" is defined in Definition 1.2 and "targeted person" (a natural person) in Definition 1.3. The unidirect, bidirect or indirect communication originates from the ontology by Mouton et al. [Mou+14a], used for the SE Indicator 3.2.

[12] "Key words for use in RFCs to Indicate Requirement Levels" [RFC2119] (updated by Request for Comments (RFC) 8174 [RFC8174]) defines the interpretation of key words in uppercase like "MUST", "MUST NOT", "MAY", "SHOULD", "OPTIONAL" etc. in Best Current Practice (BCP) 14

	SE Indicators					
Definitions	I3.1	I3.2	I3.3	I3.4	I3.5	Remarks
3.5	(y)	(y)	(p)	(i)	(y)	see Section 3.2.1
3.7	(y)	(i)	(t)	(y)	(y)	tricking implies intentional communication (I3.2); too general in I3.3 because of accepting bribery and coercion
3.8	(y)	(i)	(t)	(t)	(t)	targeted person may be aware in "fear of reprisals" attempts, i.e., knowingly complying (I3.3), and also too general regarding deception (I3.5)
3.9	(y)	(p)	(y)	(p)	(y)	too restrictive for I3.2 and I3.4
3.10	(y)	(i)	(y)	(y)	(y)	intentional communication (I3.2) was not mentioned explicitly
3.11	(y)	(i)	(y)	(y)	(y)	only part of deceptive techniques mentioned (I3.5)
3.12	(y)	(y)	(i)	(y)	(t)	unawareness not clearly mentioned (I3.3); deceptive techniques not elaborated enough, likely too general (I3.5)
3.13	(y)	(i)	(i)	(t)	(t)	no explicit "manipulator" (I3.2); benevolent attacker possible (I3.4)
3.14	(y)	(i)	(y)	(y)	(y)	attacker not explicitly communicating (I3.2)
3.16	(y)	(i)	(i)	(y)	(y)	attacker's role implicit (I3.2)

Table 3.3: Overview of SE definitions and identified SE indicators of the categories: **m**issing, **p**artial, **i**mplicit, **y**es, **t**oo general

Definition 3.17 — Uebelacker's Social Engineering Definition. The act of Social Engineering MUST be initiated by at least one attacker with malicious intent. It MUST intend to harm the targeted person, some other person or an organisation. The attacker's goal MUST comprise anything of value to the attacker, for instance, sensitive information, financial gain, valuable items, vandalism, hate crime, destroying reputation, publicity, entertainment or terrorism. The attacker MUST communicate intentionally with at least one targeted person whether the communication is unidirect, bidirect or indirect. The attacker MUST apply techniques of deception to make the targeted person comply with the attacker's request. A deceptive technique MUST be based on one or more persuasion principles applied to deceive the targeted person. The attacker MUST assume that the targeted person is unaware of the attacker's true intentions for a successful SE. The reaction of the targeted person to the attacker's action MUST be the key enabler for the Social Engineering to succeed.

4. Understanding Social Engineering

The previous chapter outlined the author's view on identifying and defining Social Engineering (SE) (Definition 3.17) based on SE indicators (Section 3.1). The diagram of the ontological model (Figure 3.2, Section 3.2.1) by Mouton et al. [Mou+14a] resembles this definition. To understand how SE works, this chapter focuses on the 'Techniques' (Deceptive Techniques, Section 4.4) and 'Compliance Principles' (Persuasion Principles, Section 4.5). That is, which techniques based on which principles an attacker may use against a targeted person. From an attacker's perspective, susceptibility to SE with applicable techniques and principles is key to succeed as the targeted person is the enabler (SE Indicator 3.1). The next sections comprise a knowledge-based model (Persuasion Knowledge Model (PKM)) and the notion of insiderness (especially knowledge insider). Besides the knowledge of attackers and their SE knowledge ('Persuasion Knowledge'), other factors of targeted persons influence their susceptibility (Section 4.3). These factors complement the general factors summarised in Chapter 2 which were categorised in human nature, culture, and personality (Hofstede's uniqueness in human mental programming [Hof01]). Here, the factors cover SE only. Finally, Section 4.6 depicts the Social Engineering Personality Framework (SEPF), suggesting based on a literature review that the level of susceptibility correlates with the targeted person's personality traits.

4.1 Persuasion Knowledge Model

Friestad and Wright [FW94] developed the Persuasion Knowledge Model (PKM) for consumer research[1] in 1994 as depicted in Figure 4.1. They modelled the persuasion relationship between consumers (target) and marketers (agent) where both sides have a certain degree of knowledge about a topic and about persuasion. Although this model was

[1] more on consumer research in Section 4.5

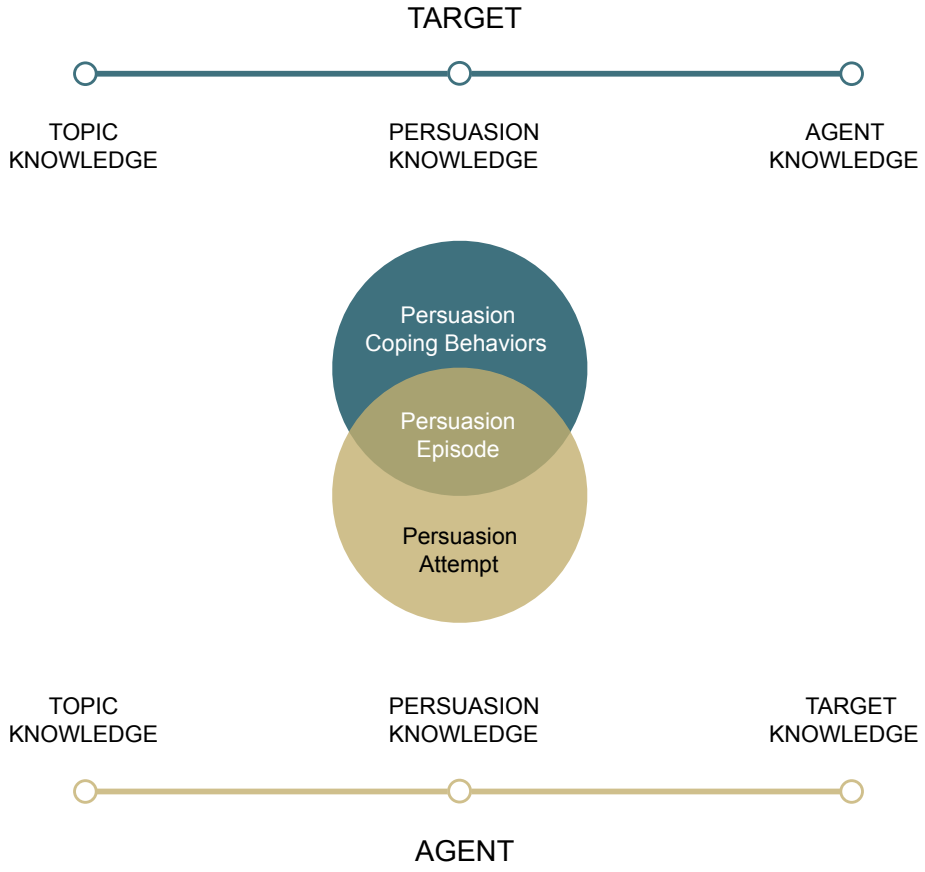


Figure 4.1: The Persuasion Knowledge Model (PKM) by Friestad and Wright [FW94] (original version in Figure C.3.3)

designed for consumer research, insights of this and similar fields (advertisement, marketing) fit very well into SE research (see Cialdini's persuasion principles in Section 4.5.1). The agent has additional knowledge about the target and vice-versa. The agent uses the knowledge for a persuasion attempt while the target tries to deal with the attempt with adequate behaviour. A consumer learns over time to interpret various persuasion attempts. Equally, marketers can alter their techniques with more precise or new knowledge about the target. Hence, the model is based on persuasion episodes and the outcome on knowledge of both sides. However, the primary focus and intended limitation on knowledge omits other factors contributing to a successful persuasion episode, such as cognitive biases or human nature (Chapter 2). The PKM roles 'agent' and 'target' map to the naming convention of 'attacker' (Definition 1.2) and 'targeted person' (Definition 1.3) in this thesis (Section 1.6).

All SE indicators except malicious intent (SE Indicator 3.4) can be found easily in the PKM: only the targeted person can let an episode succeed (SE Indicator 3.1); a communication channel is intended by the attacker (SE Indicator 3.2); if the targeted person cannot handle the persuasion attempt because of insufficient knowledge, the targeted person is unaware (SE Indicator 3.3); the attacker will use deceptive techniques to persuade the targeted person (SE Indicator 3.5). SE Indicator 3.4 covers the malicious intent of an attacker, how malicious a persuasion becomes is left open for interpretation (subtle advertisement for alcohol, tobacco, or gambling).

4.2 Insiderness

IBM stated in its Cyber Security Intelligence Index [Ser15] that outsiders were responsible for 45%, malicious insiders for 31.5%, and inadvertent actors for 23.5% of incidents affecting organisations. That is, 55% of the incidents can be associated to attackers with insider access [Ser15]. At first glance, this grouping seems plausible and is easily understood. An 'outside' attacker is not associated with the targeted organisation, a disgruntled employee can pose as a malicious insider, and even employees with best intentions can enable an attack unintentionally [Sys15]. Warkentin et al. [WCM11] see strong evidence that the major human threat to security in organisations are insiders (employees). Definition 1.1 of an attacker in Grassi et al. [GGF17] explicitly includes insiders (Section 1.6). The Threat Assessment and Remediation Analysis (TARA) definition of Tactics, Techniques, and Procedures (TTP) differentiates targeted persons between *External*, *Insider*, and *Trusted Insider* [Wyn+11, Appendix C.1]. A trusted insider describes a person within an organisation with administrative privileges. But what really defines an insider leaves open questions in a socio-technical context: is an insider a person who is member of an organisation (social/organisational domain), a person that has physical access to the premises (physical/technical domain) or digital access to the internal network (digital domain)? For instance, Pfleeger et al. [Pfl+10] defined an insider as "a person with legitimate access to an organization's computers and networks". A precise understanding of insiderness is necessary to assess insider threats.

4.2.1 Insider Knowledge

Focusing on knowledge as the Persuasion Knowledge Model (PKM) did, can clarify this situation. Using the term *insider knowledge*, i.e., knowledge about an organisation's internal matters which cannot be obtained from outside without an attack, can express what an insider looks like. Acquiring and using sensitive internal information can enable breaches of established security mechanisms — even if the attacker is not and never was part of the organisation. Current employees, former employees, contractors, and providers of outsourced services (e-mail, groupware, cloud storage, computing etc.) can be seen as knowledgeable entities in this case. Personal contacts of employees or even software manufacturers[2] who could gain insider knowledge widen this scope and can be subsumed as quasi-insiders. An attacker can target all these insiders, eliciting information and preparing for a future attack or just retrieve the desired information from insiders directly. For an organisation, holders of sensitive internal information become social assets.

If knowledge serves as one identifier of an insider, the *reachability* [PH13] aspect incorporates accessible internal objects in the physical and digital domain. That is, reachability consists of "reachable locations" and "accessible assets". Regarding access control these notions of insiderness can be weighted according to the difficulty of reaching the asset and can then be computed [PH13]. Whereas the degree of insiderness can be seen as a function of access and knowledge[3] [PH13]. While knowledge describes something an

[2] Software manufacturers or those whose software repository was breached can deploy backdoored software updates.

[3] "and, to some extent, trust, although it can be argued that trust is subsumed by the other two factors.", Bishop et al. [Bis+10] as cited in Probst and Hansen [PH13]

insider knows, accessing a restricted physical object (e.g., a room) consists of possessing an access enabler like a key, key card or security token. In comparison to biometric access control mechanisms, biometrics use an integral part of a person, inseparable from the body.

Knowledge of internal processes or even names can foster a successful SE attack: "Once a social engineer knows how things work inside the targeted company, it becomes easy to use that knowledge to develop rapport with legitimate employees" [MS02, p. 110]. For instance, Mitnick and Simon [MS02] mentioned *name-dropping* that correct (pre-gathered) names of colleagues could open a door (trust) for the attacker. An attacker may call employees and ask them to do something because of a person connected to higher positions (Chief Executive Officer (CEO) secretary) allegedly ordered the attacker to do so [MS02]. These kind of attacks are associated with the terms *president scam*, *CEO fraud* or even spear phishing (Section 4.4.3). One measure of insiderness applies psycho-social indicators, such as a "psychological profile of typical insider attackers (or non-attacking employees)" [PH13]. These indicators are used to examine insiders that may become likely an inside attacker. Although currently imprecise and lacking statistical validity [PH13], these indicators can support analyses of the "psycho-social security posture". As some inside attackers may act unintentionally and maliciously because of a preceding SE attack, they are also targeted persons. In that case, these psycho-social indicators may complement the Social Engineering Personality Framework (SEPF) in Section 4.6 which examines the susceptibility of targeted persons [UQ14]. However, the knowledge about more susceptible targeted persons can be interpreted as insider knowledge, too.

4.3 Factors influencing Susceptibility to SE

In Chapter 2 general human factors were presented, categorised in form of the three-layered human mental programming model by Hofstede [Hof01] (Figure 2.2). Categories from inherited universal human nature, learned culture specific to certain groups to personality specific to each individual were elaborated regarding cybersecurity and SE behaviour. This section will discuss a few combinations of as well as factors beyond these categories (*gullibility* in Section 4.3.1, *gender* in Section 4.3.2). "Most phishing studies found that *age* correlates with the likelihood to fall for phishing deception." [DZA12] For instance, a study by Darwish et al. [DZA12] showed that subjects between 18–24 years old are more susceptible than subjects over 25 years. In comparison, Lin et al. [Lin+19] wrote that older women were the most susceptible group to spear phishing e-mails (Section 4.4.3). The susceptibility of younger subjects declined over time, the older subjects' degree of susceptibility stayed the same. Furthermore, an agreeable *personality* is more vulnerable than a conscientious one (see SEPF, Section 4.6) [DZA12].

Darwish et al. [DZA12] also discovered that the *education* matters: subjects studying computer science are less likely to become a phishing victim than students of humanities. Internet *experience* and *digital literacy* have an impact on the *risk perception* which shapes susceptibility and precautionary behaviour [Sch+17]. In a study by Schaik et al. [Sch+17] subjects ranked the highest perceived risks as identity theft, keylogger, cyber-bullying and SE (out of 16 presented cyber-security hazards[4]). Jeske and Schaik [JS17] conducted a

[4] The term 'hazard' was attributed to "situations with the potential to do harm" [Sch+17]

Subject: Donation
From: "Mrs. Sigrid Rausing"<sigrid.rausing@bredband.net>
Date: 29.07.18, 22:38

Hello

I am Mrs. Sigrid Rausing, a Swedish business tycoon, investor, and philanthropist. Who distributed approximately 208.3 million to human rights organizations charity globally, and I also pledged to give away the rest of 25% this year 2018. However, your E-mail which was randomly selected by Google Inc. as an active web user to receive a donation of One Million Dollars as part of Sigrid Rausing charity project. Kindly Confirm ownership of your E-mail by contacting Sigrid Rausing via email for claim.
You can also read more about Sigrid Rausing via the link https://en.wikipedia.org/wiki/Sigrid_Rausing

Contact me via my Private E-mail: sigrid.rausing@yandex.com

Attachments:

S.R FOUNDATION.pdf	80,1 KB
S.R FOUNDATION.pdf	80,1 KB
S.R FOUNDATION.pdf	80,1 KB

Figure 4.2: This e-mail may not only start a conversation for a donation scam, but also contained the same text in PDF attachments. It originated from a probably hacked e-mail account of a customer of telkom.co.id.

survey (UK: sample size of N=154; USA: N=169) with questions about familiarity with online threats, use of security measures, internet attitude, internet experience, internet use, and demographics. Interestingly, the control bias (Section 2.6) may create an illusion of control with familiar threats [Tho99]. However, it is noteworthy that here the *perceived* risks resp. susceptibility (feeling) are examined, not the observed ones (reality). "Security is both a feeling and a reality. And they're not the same." [Sch08b]

4.3.1 Gullibility

The Cambridge Dictionary [Cam20] defines gullibility as "the quality of being easily deceived or tricked, and too willing to believe everything that other people say". Attackers may look for information to find potentially gullible targeted persons for SE. For instance, to be able to "craft a convincing socially engineered email and document, create malware/exploit that will bypass current antivirus detection" [Bar14, Section 9.4]. By doing so, an attacker can achieve a higher return on investment. Gullibility combined with the optimistic bias (Section 2.6) in targeted persons results in a belief system that they are unlikely to be targeted by SE and are more likely to defend themselves against SE compared to others [Bul+15]. The *DsiN-Sicherheitsindex 2020* analyses security behaviour, expertise and experience of German online users each year since 2014 [DsiN20]. They group online users into five types; one of which comprises the gullible user type. They show the biggest discrepancy of all types between (security) knowledge and associated behaviour. 34.9% of all surveyed online users were gullible users. Despite below average security behaviour and feeling quite secure, this group reported fewer security incidents which may correlate with the capability to detect incidents due to the average knowledge and the self-reported survey. The highest ranked incident were phishing attempts (20.4%) followed by receiving infected e-mails (19.7%).

When communication is involved (SE Indicator 3.2) the Pollyanna principle (Section 2.2.2) may lead some targeted persons to comply with malicious requests. Such requests may come via different media: as a facsimile in this inheritance scam where even a telephone call to the alleged inheritance executor (lawyer) in Spain was offered (Figure C.5.4). The scam e-mail in Figure 4.2 advertised that a donation is to be expected and the attacker should be contacted. It can be assumed that this is part of an advance-fee scam. Although the attached PDF files were identified as malicious by a few antivirus scanners (GData: 'PDF.Trojan.Agent.WNMPZC'; virustotal), they actually were not. One may argue that the AV vendors try to protect against such scams as well. However, scam e-mails with a malicious payload exist, see Anecdote 4.5 with a malicious Excel file attachment. Other well-known e-mail scams exist such as job offers for gullible future money mules (Figure 7.2). A similar offer may have caused a police officer to become a money mule resulting in court cases (Anecdote 7.2) [Bay. VGH, 16a D 12.2519; VG M, M 13 DK 12.3091]. The court ruled that the money mule acted in neglect or carelessness and naïvety. In Anecdote 4.1 a Polish woman in Germany called elderly targeted persons in Poland; Anecdote 4.10 narrates the story of telephone calls impersonating police officers, calling from Turkey to Germany. This resembles a typical Grandparent scam ('Enkeltrick'), initiated via a telephone call.

Anecdote 4.1 — Grandparent Scam via Telephone (Leh13). transcribed and translated phone call of German Anecdote C.1 from corresponding SpiegelTV report of Spiegel Online article [Leh13]: suspect Sylwia K. calls from Germany an elderly woman in Poland trying to persuade her to 'borrow' money.

Targeted Person:	Hi?
Attacker:	Hi Aunt!
Targeted Person:	is this Jadwiga calling?
Attacker:	yes, it is!
Targeted Person:	I don't recognise your voice.
Attacker:	I have a sore throat, you know.
Targeted Person:	poor you!
Attacker:	I have a huge problem since this morning.
(. . .)	
Attacker:	can you borrow me money till tomorrow?
Targeted Person:	how much?
Attacker:	20,000!
Targeted Person:	I borrow you nothing — I have nothing.
Attacker:	how much could you borrow me?
Targeted Person:	I don't know. . . I don't recognise your voice.
Attacker:	It's because I am at a bank.

Greenspan [Gre09] sees the difference netween gullibility and *credulity* that gullibility contains an action element (handing over money); whereas credulity describes a state of belief, e.g., believing something ridiculous or something lacking adequate evidence [Gre09]. The Cambridge Dictionary [Cam21a] defines credulity as "willingness to believe that something is real or true, especially when this is unlikely". Gullibility is defined in the Cambridge Dictionary [Cam20] (see above) broader including "being easily deceived or tricked". The targeted person is the enabler of a SE attack (SE Indicator 3.1). Hence, we have to address gullibility. However, credulity may foster a belief system that enables actions of gullible nature in the future. At least two of the four factors are at work to result in an action caused by gullibility [Gre09, Figure 1.1]:

Social situation: "presumably the con man was very persuasive, or there may have been others who vouched for his honesty";

Cognitive processes: "perhaps the victim was bad at reading people or naïve about the type of investment covered by the scam";

Personality: "perhaps the victim was a highly trusting or weak person who has difficulty saying 'no' ";

State: "perhaps the victim was exhausted or inebriated or highly infatuated with the con man"

Keiserens nye Klæder

Gullibility is a classic topic in tales, thus, nothing really new being part of human history for a long time. The "archetype for all folktales about gullibility" [Gre09] (here: credulity) was narrated in "The Emperor's New Clothes" ("Keiserens nye Klæder") [And37] by Andersen published in 1837 (Anecdote 4.2): this Danish tale of invisible clothes existed previously in slightly different versions in other cultures. In 1335 the Spanish prince Juan Manuel narrates in "El Conde Lucanor" a Persian tale about clothes invisible to men whose presumed father is not the real father [Wik18]. Even older stories told in India by Jinaratna ("Līlāvatīsāra", 1283) and Jineśvara ("Nirvāṇalīlāvatī", 1052) cover this deception: here supernatural clothes cannot be seen by people of illegitimate birth [Wik18]. These tales differ mainly in the reason why not to talk about the non-existing garments. It seems that the respective centuries and cultures influenced whether birth rights or intellect and competence were used. The common narrative could also be explained by the feeling of peer pressure or the desire of conformity to group behaviour (group attribution error; see persuasion principle in Section 4.5.1).

> **Anecdote 4.2 — The Emperor's New Clothes (And37).** Two alleged weavers ("swindlers") convinced the Emporer that their exquisite clothes are the most beautiful he can imagine. They are invisible to incompetent persons ("unusually stupid") or those who are unfit for their position. The Emporer found this property useful to distinguish fools and wise men in his empire. All people including the Emporer were deceived and did not admit that they were unable to see these clothes. Finally, a child speaks out during an Emperor's parade. Until then no one revealed their assumed incompetence or unfitness. [And37]

4.3.2 Different Outcomes Depending on Gender

The gender of an attacker or targeted person plays also a role in the effectiveness of SE. A targeted person may be influenced by the gender of the attacker. Marketing techniques can be found throughout SE stories, although primarily seen as non-malicious. A study of Rind and Bordia [RB96] showed that the amount of tipping when drawing a smiley on the bill in an upscale restaurant in Philadelphia was around 29% higher for waitresses than waiters (193 customers; mean of 2.17 customers per party). Vice-versa, the gender of a targeted person may influence the susceptibility for specific SE attempts. Costa Jr et al.[5] discovered that male and female subjects differ in personality traits across cultures. This is relevant for the Social Engineering Personality Framework (SEPF) in Section 4.6 which maps persuasion principles (Section 4.5.1) to affected personality traits (Section 2.8). Social roles, such as gender-based roles, can impact the reaction to a persuasion attempt in combination with the communication medium [GC02]. When discussing a topic, female subjects agreed less with the message in e-mails than face-to-face. There was no difference found in male subjects [GC02]. Gender-based behaviour expectations may be one explanation according to Guadagno and Cialdini [GC05]: female targeted persons tried to bond with the influence agent, whereas male subjects wanted to maintain their independence [GC05]. The gender role assumption is that male behaviour "often manifests in attempts to demonstrate one's independence from others in successful performances" [GC02]. Female roles are oriented towards communal goals, often fostering interpersonal cooperation [GC02]. The majority of studies concluded that female subjects are more susceptible to phishing than male subjects [DZA12]. Lin et al. [Lin+19] combined age and gender in their spear phishing experiment. It revealed that older women are more susceptible than younger men. However, older men and younger women showed a similar degree of susceptibility. [Lin+19, Fig. 1]

4.4 Deceptive Techniques

The ontological model of SE attacks developed by Mouton et al. [Mou+14a] in Figure 3.2 distinguished between the (SE) "Techniques" and "Compliance Principles". The latter expresses the persuasion principles by Cialdini which will be presented in Section 4.5.1. Mouton et al. [Mou+14a] mentioned the following SE techniques: phishing, pretexting, baiting, and quid pro quo. Without any direct association with SE, Definition 4.1 describes deception from the broader socio-psychological perspective. The intentional verbal or

[5] Costa Jr et al. (2001) as cited in Parrish Jr et al. [PBC09]

non-verbal communication is a key element (SE Indicator 3.2). SE Indicator 3.5 requires *deceptive techniques*, synonymously used for manipulation, tricking, and persuasion. Each deceptive technique consists of at least one persuasion principle. Hence, these parts of the ontological SE *attack* model [Mou+14a] can be applied to acts of SE, too. The list of deceptive techniques is long and would hardly be complete. They exploit human traits such as those mentioned in Chapter 2 or described later as persuasion principles. To name a few techniques here briefly: *name-dropping* [MS02] (see also Section 4.2.1); *distraction* to move the focus of the targeted person, see Anecdote 9.4 and Stajano and Wilson [SW09] in Section 4.5. This section will elaborate on the road apple attack (Section 4.4.1), impersonation (Section 4.4.2), phishing and spear phishing (Section 4.4.3) in greater detail.

> **Definition 4.1 — Deception (BV07).** "Deception is most commonly defined as intentional attempts to mislead others through words or behaviors. Deception can involve misrepresenting one's actual beliefs, knowledge, feelings, characteristics, or experiences. The term lying is commonly used to describe explicit verbal deception [...]. However, deception is more than intentionally providing others with false verbal statements; it also includes verbal omissions or the withholding of information [...]." [BV07]

Techniques and persuasion principles from marketing and sales are often consulted in SE research because of their dual-use character for malicious acts. *Product placement* in movies ('Schleichwerbung'), *viral marketing* techniques or using *social media influencers* are well-known techniques that may apply deceptive techniques. Edward Barnays jumped onto the bandwagon of the women's liberation movement to place cigarette advertisement (Anecdote 4.3). He used the movement to 'liberate' women to be able to smoke in public with confidence, rebelling against the predominant "traditional social and cultural prohibitions" [Bra96]. Nowadays the ubiquity of digital technology leads to improved and automatable deceptive techniques, e.g., *dark patterns*. Dark patterns describe deceptive internet patterns or designs in online services, such as social networks. Their main purpose is to lure users into performing actions or change behaviour that may have a negative impact [TAB30]. They belong to the group of neuromarketing techniques. Dark patterns are used to distract from a higher bill or hidden extras as well as to address emotions to buy products [TAB30]. One technique may sneak items into the online shop basket of the targeted person. If malicious intent is involved, dark patterns can express SE (SE Indicator 3.4). Bogenstahl [TAB30] categorises them as unethical, sometimes fraudulent.[6] Related are the intentionally crafted game elements to make children addictive to the game Fortnite by Epic Games [Ols19].

Like market analysis or examining the target group for future marketing campaigns, similar actions can precede acts of SE. Information gathering without all SE indicators involved (especially the human enabler, SE Indicator 3.1), such as dumpster diving (Definition 3.2) does not express SE. Pre-attack information gathering can be used to fuel more convincing techniques, e.g., crafting spear phishing e-mails (Section 4.4.3). In particular insider knowledge (Section 4.2.1), can make SE more persuasive. Of course, technology exists

[6] "Der Einsatz von Dark Patterns ist unethisch, mitunter unlauter und ggf. betrügerisch. Insbesondere sind auf die Ausnutzung menschlicher Wahrnehmungsschwächen ausgerichtete Dark Patterns für unerfahrene Nutzende schädlich, [...]" [TAB30]

supporting social engineers by collecting and organising the information: Recon-ng[7] is a reconnaissance framework. FOCA[8] (Fingerprinting Organizations with Collected Archives) scans for metadata in documents found on websites via search engines.

> **Anecdote 4.3 — Edward Bernays: Torches of Freedom (Bra96).** Edward Bernays[9] helped George Washington Hill, president of American Tobacco, to advertise explicitly targeting women. In the 1920s, women were smoking rarely in public due to "traditional social and cultural prohibitions" [Bra96]. Many of these prohibitions were questioned by the Women's Liberation Front. Hill wanted to change the smoking behaviour and win the underrepresented female market for Lucky Strikes ("a gold mine"). The consulted psychiatrist A. A. Brill stated "Today the emancipation of women has suppressed many of their feminine desires. More women now do the same work as men do. Many women bear no children; those who do bear have fewer children. Feminine traits are masked. Cigarettes, which are equated with men, become torches of freedom." [Bra96] For the 1929 New York City Easter parade Bernays recruited debutantes smoking cigarettes. They titled the stunt 'Torches of Freedom' which in New York had even more significance. Bernays successful advertisement changed the way public relations worked. Decades later, he became an anti-smoking advocate. [Bra96]

4.4.1 Technique: Road Apple Attack

In research, the *Road Apple Attack* with USB flash drives is an often applied technique [KSR17; Las+13; PDP13]. It originates from the observational lost-letter study by Farrington and Knight (1979)[10]. Pieters et al. [PDP13] adapted the road apple attack with USB flash drives as described in Anecdote 4.4. The communication is indirect and intentional by the attacker and researchers (SE Indicator 3.2). The targeted persons enable the attack to successfully enter the organisation's premises (SE Indicator 3.1). The attacker tries to deceive the targeted persons by flash drives with the organisation's logo (SE Indicator 3.5) who may be unaware of this trickery (SE Indicator 3.3). The attacker of Anecdote 4.4 has a malicious intent of copying sales data (SE Indicator 3.4). The reasons a targeted person takes the flash drive and plugs it into an office computer may explain the behaviour and probable countermeasures. A targeted person might feel responsible to help and wants to identify the owner of the flash drive. Or the targeted person hopes to find sensitive information about the organisation.

[7] https://github.com/lanmaster53/recon-ng
[8] https://github.com/ElevenPaths/FOCA
[9] Edward Bernays was a nephew of Sigmund Freud. The influence of psychoanalysis was huge in that time. Cigarettes were seen as a symbol of freedom as well as a phallic symbol offered by a man to a woman. Smoking was a "sublimation of oral eroticism" [Bra96]
[10] Farrington and Knight (1979) as cited in Lastdrager et al. [Las+13]

> **Anecdote 4.4 — Road Apple Attack (PDP13).** A USB flash drive prepared with malware and displaying the organisation's logo, is left by the attacker in a publicly accessible location, such as a canteen or parking lot. A targeted person (employee) finds the flash drive and enters the organisation's building. In the office, the targeted person connects the flash drive to the Microsoft Windows computer. The malware (rootkit) installs itself via autorun. The malware searches for and collects sales data on the computer or network shares. The malware encrypts the collected data before transmission to bypass the firewall which cannot inspect encrypted data. The attacker outside of the organisation's premises receives the encrypted sales data and decrypts it. [PDP13]

The origin of Anecdote 4.4 dates back to 2006. Novel attack vectors for the same deceptive technique were developed with the ongoing miniaturisation of hardware. Schilling and Steinmetz [SS16] used the USB Armory stick[11] from Inverse Path (F-Secure) that resembles a USB flash drive. However, it contains an ARM mini computer that masks itself as a mass storage device and simultaneously registers a new (USB) network device. The latter feature can phone home almost undetected using the host machine's network access. With much publicity Kaspersky researcher Schouwenberg discovered and exposed the *Stuxnet worm* [Kus13]. It is assumed that a Russian contractor of the Iranian nuclear power programme received a flash drive not knowing of its malicious payload (SE Indicator 3.3). Without the action of the contractor, the attack would not have been successful (human enabler, SE Indicator 3.1). The flash drive infected an Iranian uranium enrichment plant successfully. One reason was the Stuxnet's ability to infect computers without internet connection. An infected computer places its malicious payload onto an uninfected USB flash drive. This flash drive then is able to infect other systems when plugged in — like in times when the internet was not wide-spread at home and malware travelled via floppy disks such as the boot sector virus *Stoned* (1987). Stuxnet was powerful by carrying zero-day exploits for Microsoft Windows, unknown to the white hat community [Kus13].

4.4.2 Technique: Impersonation and Imposture

Imitating another person (impersonation) is an omnipresent part of human life. People narrate stories and imitate their characters when using direct speech, even when telling a story to oneself, people slip into the different roles (narrative psychology, Section 1.8). Actors often play other characters unless starring as themselves. Some stories make it into movies and are based on real events such as the book 'Captain of Köpenick' [Zuc67] (Anecdote 3.1). Others are fictional and are based on books: the book 'Alias Madame Doubtfire' by Anne Fine was made into the well-received movie 'Mrs. Doubtfire' starring Robin Williams; the novels 'A Series of Unfortunate Events' by Daniel Handler where imposter Count Olaf (Neil Patrick Harris) disguises as a different character in each episode to snatch the inheritance of the Baudelaire children; or famous in Germany, the 'Die Feuerzangenbowle' (The Fire-Tongs Bowl) movies based on the book by Spoerl [Spo62]. Heinz Rühmann played the role of a pupil imposter who never went to school and wanted to experience it, even with a PhD already in his pocket. In these aforementioned stories,

[11] https://inversepath.com/usbarmory

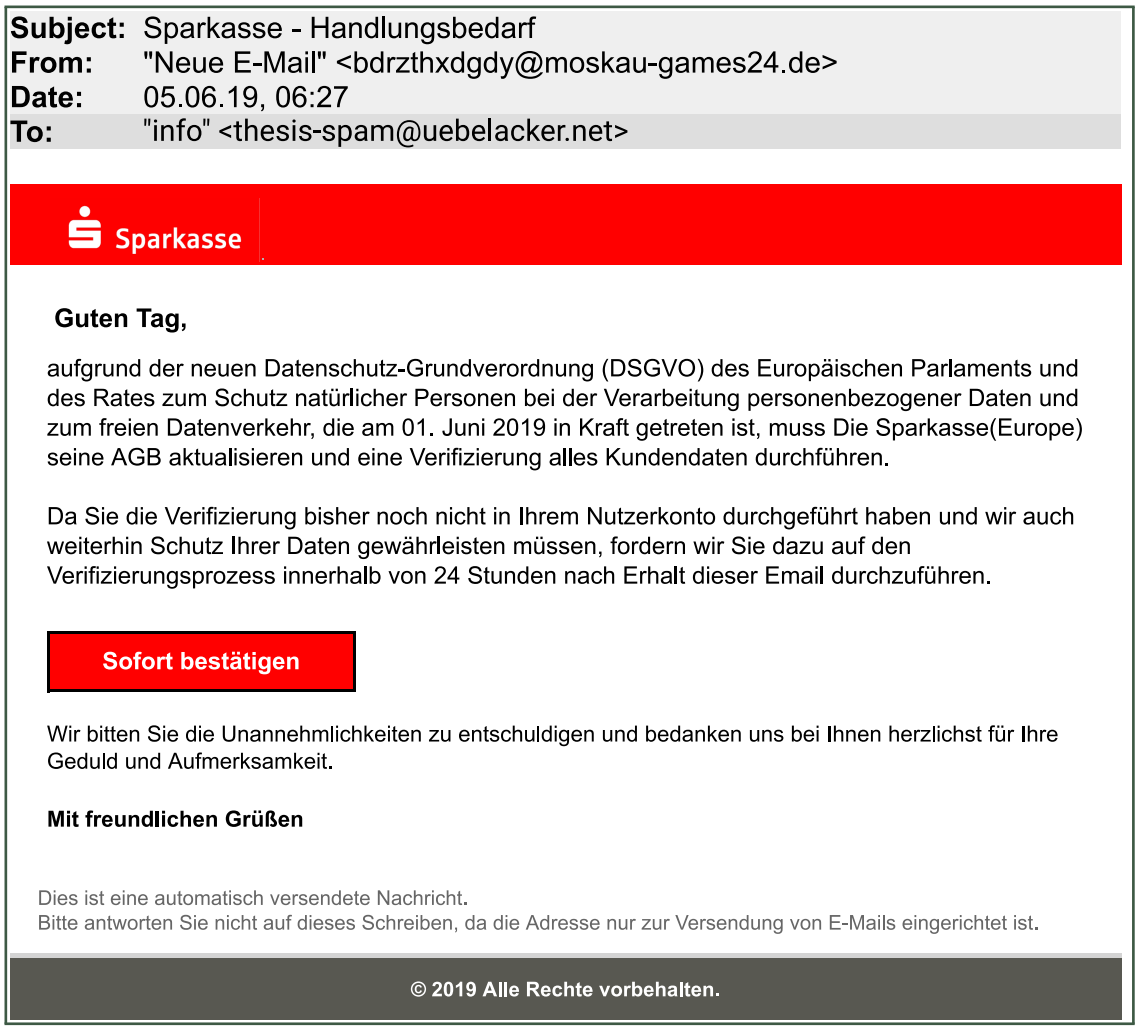


Figure 4.3: Urging recipients to follow a phishing link (using the URL shortener `7i.fi`) and enter bank credentials for verification purposes within 24 hours because of alleged GDPR requirements (recipient address changed)

movie actors slip into the roles of the books' characters who, as well, imitate to be someone else. That is, not only as an entertaining factor in books and movies, also people's minds narrate interactions by switching the roles of the actors.

No wonder, impersonation is also used maliciously. One definition in the Cambridge Dictionary defines impersonation as "the act of attempting to deceive someone by pretending that you are another person" [Cam21b]. The term 'imposture' ("the act of pretending to be someone else in order to deceive others" [Cam21c]) describes a person who pretends to be another persona in a malicious context (imposter). An attacker can impersonate an existing individual — probably known to the targeted person — like a relative or a CEO or slip into a persona such as a police officer. Depending on this type (impersonation, imposture), the technique of how to impersonate the individual or play the role of a persona may differ. Malicious impersonation and imposture are widely used in SE, starting with easily spoofed sender addresses in e-mails establishing trust or *catfishing* (also known as *romance scam*, Definition 4.2) in online networks. Catfishing and grandparent scam ('Enkeltrick',

Anecdote 4.1) are founded on impersonation. Attackers can disguise as head hunters to acquire sensitive information of the targeted person [MS02, p. 22–26] or use recent news coverage (availability bias, Section 2.3.2) to craft corresponding attacks. When the GDPR was widely covered, phishing e-mails appeared, e.g., pretending to originate from a bank (Figure 4.3). Di Martino et al. [DiM+19] used the General Data Protection Regulation's (GDPR) 'Right of Access' in their experiment to retrieve personal data by impersonating the genuine requester. They were granted full access to personal data in 15 of the 55 targeted organisations. The personal data consisted of sensitive information such as geolocation or financial transactions [DiM+19]. The COVID-19 pandemic with corona relief funds bear another opportunity for attackers. The Hamburger Corona Soforthilfe (HCS) stopped its relief programme for self-employed shortly after phishing websites appeared mimicking the original site [NDR20a]. Or older targeted persons were distracted by a health care imposter allegedly collecting COVID-19 contact details, while an accomplice stole valuables [NDR20b].

> **Definition 4.2 — Catfishing (Sch+17).** "The act of building a fake relationship online by pretending to be someone else, creating an online romance through a false persona or fake social media profile." [Sch+17]

4.4.3 Technique: Phishing

For the digital age, e-mail is quite an old technology and medium. The Simple Mail Transfer Protocol (SMTP) Request for Comments (RFC) was published in 1982 [RFC821]. Similar to facsimile or traditional mail (letter post, postcards), the sender address in an e-mail ('From:' header field) can be spoofed easily. Spoofing sender addresses is one form of impersonation. In *phishing* (Definition 4.3) the content of the e-mail body is used to fool recipients into believing that the phishing e-mail originates from a legitimate source. Sometimes the sender address is also manipulated. Like unsolicited commercial or bulk e-mail (spam), *technique propagation* [Sch00] makes it easy to send out a lot of phishing e-mails with minimal investment.

> **Definition 4.3 — Phishing (JS17).** "The act of sending an e-mail to a user falsely claiming to be an established legitimate enterprise. The aim is to scam the user into surrendering private information that will be used to steal the user's identity." [JS17]

Successful phishing requires a human enabler. Without the interaction of the targeted person, a phishing link would not be clicked and sensitive information on a phishing website would not be entered (SE Indicator 3.1). The communication is mostly unidirectional and intended by the attacker (SE Indicator 3.2). The attacker tries to retrieve sensitive information maliciously (SE Indicator 3.4) to exploit the targeted person who must be unaware of it (SE Indicator 3.3). Phishing uses deceptive techniques (SE Indicator 3.5). Anecdote 4.5 shows the case of a phishing e-mail whose malicious payload (infected Excel document) started the infection of the network of an electricity supplier, leading to a blackout [BfV16, p. 264–265].

Anecdote 4.5 — Phishing E-Mail Resulting in Ukrainian Electricity Blackout. In December 23rd 2015 a west-Ukrainian region was experiencing a blackout for a few 100.000 households. The local electricity supplier was targeted before by a phishing e-mail with an infected Excel document attached. Step by step the attacker got access to the network and systems which were partly SCADA systems (ICS). To impede the mitigation and disinfection process, the attacker launched a DDoS attack on the service hotline of the supplier. [BfV16]

An example phishing e-mail is depicted in Figure 4.3 where a sender address spoofing was omitted. The e-mail is crafted without any personal information like a personalised salutation. It targets customers of the savings banks called Sparkasse without any specific branch mentioned. The introduction of the GDPR was used as a recent event to lure potential Sparkasse customers to a phishing website. For Sparkasse customers this might seem a genuine e-mail, others would just ignore it. The attacker does not usually know of which bank the targeted person is a customer. A recurring characteristic of typical phishing e-mails are required responses by the targeted person in a timely manner (persuasion principle 'Scarcity of Time' in Section 4.5.2). Figure 4.4 shows the same time pressure requesting urgent action to avoid the suspension of an e-mail account. There, the salutation is just the string preceding the '@' symbol, but could be in some cases the last name which would appear more convincing. Both examples share a link to a phishing website. However, in a typical display as HTML messages, these links only appear when the mouse hovers over the 'button'. Figure 4.5 delineates the same e-mail, but representing the message body as text only. The link directs to a website most likely hacked. To further mask a malicious URL, attackers can register look-alike domains, see IDN homograph attack below. In comparison to Figure 4.3, the phishing link contains the recipient's e-mail address. Hence, the attacker puts a little bit more effort into creating these e-mails as each of them needs modifications of salutation, e-mail address, and phishing link for each recipient. This unique link enables the attacker to identify who clicked the link[12] even if nothing was entered on that website.

More sophisticated phishing that further personalises each e-mail is called *spear phishing* (Definition 4.4). That is, including personal information can establish trust in the genuineness of an e-mail. Personal information might be collected manually when high profile targeted persons are attacked, e.g., by the impersonation of their CEO (CEO fraud, Section 4.2.1). But automation with the aforementioned tools is possible. Anecdote 7.3 describes spear phishing attack for carbon emission certificates in Germany, documented in court documents. The malware *EMOTET* used the existing contacts of the infected victims to reply to previous conversations. It abused existing trust relationships by sending spam or phishing e-mails. Infected systems were also used to install other malware, such as the banking trojan horse *Trickbot* [BSI19a]. The successful spear phishing attack that infiltrated the internal network of the German parliament Bundestag was covered with a lot of publicity (Anecdote 4.6) [BfV16, p. 262–263].

[12] and not just the e-mail address from the URL, but also IP address, browser, operating system etc.

Subject: Account Alert: su261@cl.cam.ac.uk
From: Webmaster <accounts@post.com>
Date: 17.07.2017 00:54
To: su261@cl.cam.ac.uk

Dear su261,

Dated: July 15, 2017.

This is to inform your email account (su261@cl.cam.ac.uk) service is currently out-dated and some of your incoming & outgoing messages status already showing pending in our server.
Your service to send/receive E-messages could be suspended next 24 hours time if not verified.

You are requested to quickly click below verify button to update/activate your service in full and start enjoying your E-messages service again.

Click here to verify

Note: Your email account send/receive services will permanently be disabled if you fail to verify correctly.

© 2017 Email Administrator. All Rights Reversed.
********** ********** ********** ******

Figure 4.4: Phishing e-mail received at University of Cambridge (**HTML version**: phishing link not shown without interaction)

Subject: Account Alert: su261@cl.cam.ac.uk
From: Webmaster <accounts@post.com>
Date: 17.07.2017 00:54
To: su261@cl.cam.ac.uk

```
Dear su261,

Dated: July 15, 2017.

This is to inform your email account (su261@cl.cam.ac.uk) service is currently
out-dated and some of your incoming & outgoing messages status already showing
pending in our server.
Your service to send/receive E-messages could be suspended next 24 hours time if
not verified.

You are requested to quickly click below verify button to update/activate your
service in full and start enjoying your E-messages service again.

Click here to verify <http://augustlodge.com/en.php?name=su261@cl.cam.ac.uk>

*Note:*Your email account send/receive services will permanently be disabled if
you fail to verify correctly.

©2017 Email Administrator. All Rights Reversed.
********** ********** **********
********** ******
```

Figure 4.5: Phishing e-mail received at University of Cambridge (**text version** of Figure 4.4: phishing link visible)

Anecdote 4.6 — Phishing E-Mails Infecting German Bundestag. Spear phishing attacks against government entities are a growing concern and became more publicly recognised when in May 2015 the German Bundestag experienced almost a complete shut down of their internal network. Such targeted e-mails containing a malicious link were sent on August 15th and 24th 2016 again to infiltrate German political parties. These spoofed e-mails originated from an IP address well known for APT 28 attacks, suspecting Russian intelligence agencies. [BfV16]

Sensitive personal information such as credentials can be used to get the targeted person's attention, e.g., in sextortion scams. These credential may originate from data leaks [Kre18]. After the hacked bitcoin exchange Mt. Gox in 2011 which included the customers' e-mail addresses, the author as a former client received a lot more scam e-mails (transfer bitcoins and receive twice as much bitcoins in return), some of which were phishing e-mails. The leaked data was also processed by haveibeenpwned.[13] The leaked data set for Mt. Gox customers looked like:

- UserID: 5948
- Username: uebelacker
- Email: sven@uebelacker.net
- Password: <password hash>

Definition 4.4 — Spear Phishing (Cap+14). "Spear phishing is a form of cyberattack attempting to infiltrate a system or organization for cybercrime or espionage purposes. Cyberattackers find inside information specifically relevant to users and craft fake email messages, usually impersonating well-known companies, trusted relationships, or contexts." [Cap+14]

IDN Homograph Attack

Internationalised Domain Names (IDNs) [RFC5890] extend the limited ASCII characters for domain names with Unicode characters (UTF-8). With a set of allowed Unicode characters, domain names such as `übelacker.de` (U umlaut) can be displayed in applications. Because of the underlying ASCII-based representation of domain names, IDN must be transformed by a bijective function (Punycode). Then `übelacker.de` is represented by `xn--belacker-55a.de` and vice-versa. This transformation is also available for the Top Level Domain (TLD) part.

The challenge with Unicode characters is that some look quite alike, e.g., the ASCII character 'o' (U+006F) resembles the Unicode character for the Greek omicron 'ο' (U+03BF). Others differ just a little bit; therefore, organisations try to find those and register such domains for themselves or domain registrars agree to not provide such look-alikes. For instance, the professional business network XING with the domain `xing.com` also acquired `xıng.com` (`xn--xng-jua.com`) with a Turkish 'ı' without a dot (U+C4B1). For targeted persons these minor differences are challenging — maybe a defective pixel or some dust on the screen causes this appearance. Besides registrars and domain owners

[13] https://haveibeenpwned.com/

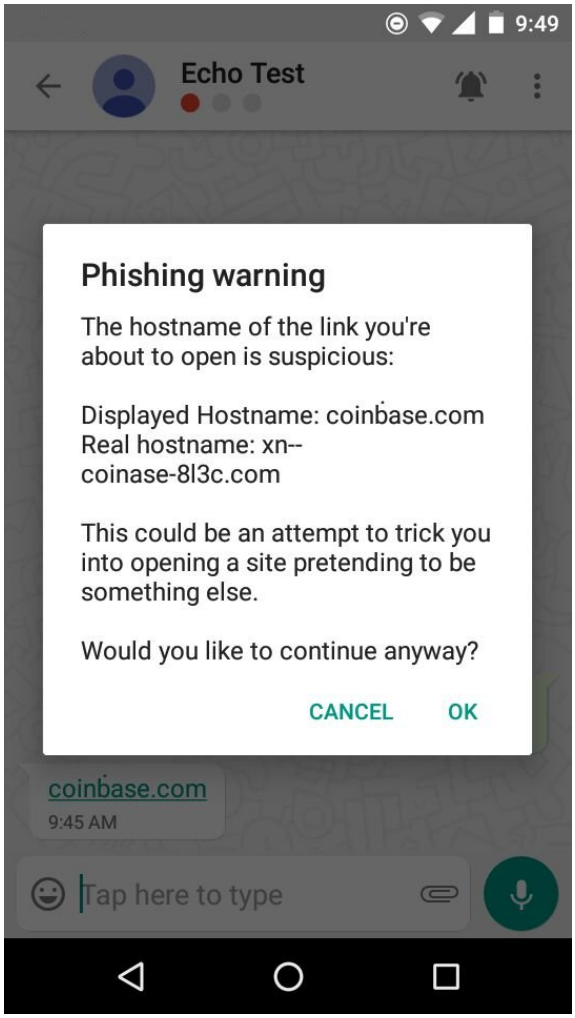


Figure 4.6: Threema IDN homograph attack prevention: showing the IDN encoded domain name prior to accessing the URL[15]

implementing security measures, UX designers can support end-users by highlighting or intervening when look-alike domain names are handled. The instant messenger Threema[14] implemented a phishing warning when parsing such domains. Figure 4.6 shows a dialog box of a sent domain that contains a Unicode character (U+1E03) of the letter 'b' with an additional dot on top. The domain `coinḃase.com` (`xn--coinase-8l3c.com`) looks very similar to the correct domain `coinbase.com`, a cryptocurrency exchange. UX designers may also help end-users by making clickable URLs in HTML phishing e-mails more visible and intervene with an extra step before calling these URLs.

Opportunistic SE

Some of the aforementioned scams, such as the classic 'Spanish prisoner' or the 'Lettre de Jérusalem', jumped on the bandwagon of recent events. This chapter showed that the introduction of the GDPR resulted in phishing e-mails as shown in Figure 4.3. The COVID-

[14] https://threema.ch/en
[15] source: https://github.com/threema-ch/threema-web/issues/791, accessed 2019-04-29

19 pandemic was used for scams as well. For the targeted persons, recent events discussed in the mass media are more vivid in memory, see availability bias about 'vividness of an information' (Section 2.3.2). By adding a risk factor (if not already existing like with COVID-19) attackers can reuse these events. If such an event opportunity arises, it can be embedded into the scam story, e.g., in a phishing e-mail. The technique though stays almost the same whether SE adapts recent news about GDPR or data breaches and urges targeted persons to perform actions. Definition 4.5 reflects this type of SE. Edward Bernays' historic marketing stunt of the freedom torches (Anecdote 4.3) resembles 'Opportunistic SE' (Definition 4.5) in marketing context and in real-life.

Definition 4.5 — Opportunistic Social Engineering. Incorporates recent events into Social Engineering to craft more vivid stories for the targeted persons.

4.5 Persuasion Principles

In advertising and sales, persuasion principles are widely applied to influence targeted persons [BV07]. The act of persuasion is intended to change how a targeted person feels, reacts etc., see Definition 4.6. As Akerlof and Shiller [AS15] put it: advertisers are hired to enhance the sales, systematically use trial and error techniques to understand how to trigger customers to buy their products — even against the well-being of the customers. Section 4.3.2 showed research about tipping behaviour when a waitress draws a smiley on the bill [RB96]. Cialdini [Cia07] developed the psychology of persuasion originally for understanding the effectiveness of marketing and sales campaigns with its underlying human factors. For instance, he found that when waiters hand over the bill, then return handing out more mints while saying they were nice customers, the tipping was higher compared to the value of the mints (see reciprocity principle in Section 4.5.1).

Definition 4.6 — Persuasion (BV07). "Persuasion is a method of changing a person's cognitions, feelings, behaviors, or general evaluations (attitudes) toward some object, issue, or person." [BV07]

In Section 4.4 deceptive techniques (SE Indicator 3.5) concerned the "Techniques" in the ontological model of SE attacks by Mouton et al. [Mou+14a] (Figure 3.2). The techniques discussed can be used to deceive targeted persons intentionally and maliciously (SE Indicator 3.4) who then enable an attack to become successful (SE Indicator 3.1). These psychological methods, applied to get targeted persons comply with malicious requests, are based on Cialdini's book on *Influence: The Psychology of Persuasion* [Cia07], more in Section 4.5.1. Because the term 'persuasion principles' is widely used in the literature related to SE, it will be used here as well. It is crucial to clarify, that persuasion principles can be found not only in SE; if these principles are used in deceptive techniques targeting possibly unaware persons, they may be part of SE.

Modic and Lea [ML13] researched compliance of targeted persons towards scams. An initial study (*N*=779) from Modic et al. [MAP18] resulted in a modular psychometric tool that relied on underlying salient factors from pre-existing empirical and theoretical research. This *self-reported* (online) questionnaire measures the targeted persons' susceptibility to persuasion. Their Susceptibility-to-Persuasion scale (StP-II) consists of 138 items in initially nine subscales [MAP18]. Another analysis of scam victim experiments in the UK revealed seven principles elaborated by Stajano and Wilson [SW09] based mostly on face-to-face attacks (*observed* in real life situations; bidirectional communication; SE Indicator 3.2). Their principles are extracted from various entertaining experiments for TV[16]. The majority of experiments used deceptive techniques (Section 4.4; SE Indicator 3.5) such as impersonation. Although not intended to be malicious acts but experiments (SE Indicator 3.4), the majority expressed SE. Targeted persons were unaware (SE Indicator 3.3). Some of which overlap with Cialdini's principles or other previously mentioned behaviours. While Cialdini's principles focused initially on marketing and sales, the compliance principles here explicitly covered scams and may explain how targeted persons fell for the scams (SE Indicator 3.1).

Distraction

"While you are distracted by what retains your interest, hustlers can do anything to you and you won't notice." [SW09] The Dual Process Theory (Section 2.4) defines two types of information processing: system 1 (peripheral, heuristic) and system 2 (systematic, rational) [Kah11]. Our working memory has limitations (Section 2.4.1) on what we can process to make decisions [Mil56]. This influences the upper boundary of our cognitive capacities for security-related tasks (Section 2.4.1) [Ben+15]. Attackers may overload a targeted person's information processing to let it switch to system 1 for heuristic decision making. System 1 would be prone to some cognitive biases by which system 2 would not be equally affected because of more rational decisions. Gragg [Gra02] added that when a targeted person is confronted with an unexpected perspective, the targeted person needs time to process which may be scarce. An attacker can use this as a distraction to cloud a targeted person's judgement.

In the Real Hustle TV series they conducted experiments such as the 'Monte' scams where a sleight-of-hand trick is the key distraction technique. In the later Anecdote 9.4 presented at the SE Poetry Slam, the slammer will tell the physical penetration testing story of using a puppy for distraction.

Social Compliance

"Society trains people not to question authority. Hustlers exploit this 'suspension of suspiciousness' to make you do what they want." [SW09] Cialdini calls this principle 'Authority'. In an organisational context, the CEO and Chief Information Security Officer (CISO) are authority roles in the fictional Anecdote 4.7, see also CEO fraud in Section 4.2.1.

[16] 'The Real Hustle', BBC, 2006–2012, https://www.bbc.co.uk/programmes/b006m8mf

> **Anecdote 4.7 — Authoritative E-mail Request.** On Friday noon, the targeted person receives an e-mail allegedly sent by the organisation's CISO. An attacker faked the sender address of the CISO. The attacker describes in the e-mail the migration plan towards an improved organisation-wide multi-factor authentication. For a smooth procedure, administrators will work over the weekend to prevent parallel authentication services and login issues next week. The attacker excuses the short notice and explains the urgency of this enhanced security mechanism which was authorised by the CEO. The CISO's e-mail provides a link to the new service website where the targeted person must enter the password and provide a mobile phone number. After entering the password and mobile phone number, the website presents a success message. However, the website was set up by the attacker to collect this sensitive information.

Herd

"Even suspicious marks will let their guard down when everyone next to them appears to share the same risks. Safety in numbers? Not if they're all conspiring against you." [SW09] The social proof principle from Cialdini incorporates this principle.

Dishonesty

"Anything illegal you do will be used against you by the fraudster, making it harder for you to seek help once you realize you've been had." [SW09] This principle does not match exactly with one of Cialdini's. Commitment and consistency may be related if the attackers lures the targeted person into performing a minor illegal activity or a lie. Then the attacker might request additional illegal actions or lies. In the 'The Real Hustle' one experiment was to sell counterfeit notes to targeted persons knowing about it, but probably not being aware of the full scale of the scam. However, the targeted person would likely be aware of an illegal action and SE Indicator 3.3 (unawareness) would not fit easily.

Deception

"Things and people are not what they seem. Hustlers know how to manipulate you to make you believe that they are." [SW09] Deception is a key component of SE (SE Indicator 3.5), see also Section 4.4.

Need and Greed

"Your needs and desires make you vulnerable. Once hustlers know what you really want, they can easily manipulate you." [SW09]

Time

"When you are under time pressure to make an important choice, you use a different decision strategy. Hustlers steer you towards a strategy involving less reasoning." [SW09] The time principle resembles Cialdini's refined principle 'Scarcity of Time' (Section 4.5.2).

4.5.1 Cialdini's Principles of Persuasion

The Dual Process Theory of Section 2.4 can be found in the Elaboration Likelihood Model (ELM) of persuasion: Petty and Cacioppo [PC86] developed this general theory "for organizing, categorizing, and understanding the basic processes underlying the effectiveness of persuasive communication" [PC86]. The 'persuasive communication' reflects the presented SE indicators 'intentional communication' (SE Indicator 3.2) and 'deceptive techniques' with persuasion principles (SE Indicator 3.5). Their two routes of persuasion consist of (i) the central route where messages are processed in "thoughtful consideration". Thus, the resulting attitudes of targeted persons are based on the message content and would endure longer (less likely to change if contradicting arguments are presented). The (ii) peripheral route works on heuristics [Kah11] and the attitude changes rely more on the cues of the information context, e.g., attractiveness of the attacker or information source [PC86].

The Dual Process Theory and ELM also found their way into Cialdini's research [GC05]: the central route of information processing is more likely used if cognitive resources are available, the topic is important or the targeted person has some knowledge about it, or the information is presented in written form. Whereas the peripheral route is chosen otherwise. Depending on the route of information processing, information may be more persuasive based on the quantity of arguments and perceived credibility of the attacker (peripheral) or the quality and veracity of an attacker's argument (central) [GC05]. Cialdini's framework of persuasion covers six basic principles [Cia07] that may increase the likelihood of a successful attack. That is, the sole application of one Cialdini principle does not automatically create SE attacks, but they can be part of SE as defined in SE Indicator 3.5. The principles of *Authority* and *Scarcity* can be refined further and will be discussed afterwards (Section 4.5.2). Akerlof and Shiller [AS15] state that Cialdini's principles express most of the psychological biases upon which behavioural economics rely. Scheeres [Sch08a] showed that Cialdini's principles can be applied in the context of SE.

The persuasion principles most commonly found in a literature review of SE anecdotes were: Authority (63.3%), Liking (13.3%), Reciprocity (11.1%), and Commitment (10.6%) [Bul+18]. The following description of principles were adapted from Bullée et al. [Bul+15]. Each principle can also be understood in stories (Cialdini: mental frames) which resemble the *mode of thinking* of targeted persons (Narrative Psychology, Section 1.8) [AS15]. Sagarin et al. [Sag+02] conducted studies on the resistance to deceptive persuasion. Attempts to minimise the susceptibility of targeted persons are more likely to succeed if they address two aspects: (i) the "perceived undue manipulative intent" (of the attacker) and (ii) the "perceived personal vulnerability to such manipulation" [Sag+02]. The former may refer to the extent a targeted person is aware of a SE attempt (SE Indicator 3.3). The latter ("illusion of invulnerability" [Sag+02]) resembles the effects of overconfidence of Section 2.5. Muscanell et al. [MGM14] analysed Cialdini's principles with respect to internet scams. They developed best practices of how to minimise the likelihood of a successful attack for each principle. The best practice defences are added below. Furthermore, good questions which targeted person can ask themselves are presented based on the work of Muscanell et al. [MGM14] and Bullée et al. [Bul+18].

Reciprocity

"refers to the giving of something in return. The target feels indebted to the requester for making a gesture. Even the smallest gift puts the requester in an advantageous position" [Bul+15]. Anecdote 4.8 tells the case of the Hare Krishna Society using the reciprocity principle to collect donations [Cia07]. "Reciprocity helps establishing trust with others and refers to our need for equity" [UQ14].

Mode of Thinking: targeted persons want to be part of stories where they can reciprocate favours and gifts [AS15].

Best Practice Defence: the targeted person should reject the initial malicious request. Targeted persons have to accomplish to discover the true motives of the attacker [MGM14].

Self-Questioning: "is it likely this person really did me this kindness without expecting anything in return?" [MGM14]

> **Anecdote 4.8 — Hare Krishna Flower Gift.** In former approaches the devotees of the Hare Krishna Society were simply asking for donations on the street which did not work very well. After they introduced the "benefactor-before-beggar" [Cia07] system, targeted persons donated more money: members of Hare Krishna handed 'gifts' to passersby (targeted persons), e.g., books or flowers. Flowers were pressed into the hands of targeted persons telling them it were a gift and returning the flowers was not acceptable. Then the devotees asked the targeted persons to contribute to the Hare Krishna Society. By changing their fund raising scheme, the Hare Krishna Society raised more money than before. [Cia07]

Conformity (also known as Social Proof)

"is imitating the behavior of other people. Members of the in-group have a stronger feeling of group-safety compared with members of the out-group." [Bul+15] The group's behaviour is more likely to be consulted and copied in situations in which the targeted persons feel insecure. This resembles the herd principle of Stajano and Wilson [SW09]. Although not knowing the previous customers of a Hotel room, new customers will more likely (+33%) reuse towels when a note says that previous customers in that room reused towels 75% of the time. The fictional Anecdote 4.9 comes from the saying 'don't look a gift horse in the mouth' and applies a similar approach to information security. Luring targeted persons into inserting USB sticks with malicious software is also present in the road apple attack (Anecdote 4.4). Here, the targeted persons receive the USB device via a postal service.

Mode of Thinking: to follow others, targeted persons want to tell their story of assuming others show better judgements (information explanation) or wanting

to avoid incurring "disapproval by failing to conform (in the social conformity explanation)" [AS15].

Best Practice Defence: targeted persons need "to understand that other peoples' actions can sometimes be wrong and under which circumstances" [MGM14], especially when in groups, e.g., mob behaviour.

Self-Questioning: "would I do the same if I was alone in this situation?" [Bul+18]

> **Anecdote 4.9 — Gift Horse Packet.** The targeted person works in public relations and receives a packet at the office. It consists of a gift originating from a company from which the targeted person has never bought anything. The gift card explains that based on customer analysis of corporate clients, the targeted person might be interested in this kind of product as well. Other happy customers of the same business in the same area really like this gadget and use it all the time. 75% of their customers order it as give-aways for their clients. The targeted person can try it out for free. The gift reveals itself as a USB powered mug heater displaying the temperature with different coloured LEDs. If an additional software of an attached CD is installed, the program even depicts the temperature history in a neat graph on the computer. The software installed not just the visualisation component, but also a malicious payload.

Liking (also known as Similarity)
"someone puts that person in a favorable position. People tend to like others who are similar in terms of interests, attitudes and beliefs." [Bul+15] Ryan and Mauch [RM10] created the fake identity Robin Sage that listed the same schools as targeted persons on social networking platforms (Section 4.5.1). They were able to extract sensitive information from targeted persons because of trust gained through similarity. Perceived similarities — even superficial ones like shared names or birthdays — augment compliance because it may come from the same social group [UQ14], thus, liking persons with similar qualities. "If you make it plain you like people, it's hard for them to resist liking you back" [Buj02].

Mode of Thinking: targeted persons want to be liked and, thus, they must take part in stories where they are liked or not [AS15].

Best Practice Defence: targeted persons should imagine removing any likeable persons (upon realisation of their likeability) from a situation and reevaluate the situation in isolation [MGM14].

Self-Questioning: "would I say yes to this request if someone else were asking me?" [MGM14]

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." [Bul+15] This principle addresses directly the previously mentioned loss aversion (Section 2.3.1). If something is perceived as becoming scarce soon, subjects sense losing freedoms and try to counteract [UQ14]. Something that seems to become unavailable in the future motivates averting the loss [Wes08]. Adam Smith also elaborated on loss aversion in the 18th century [ACL05]. The scarcity refinements will be discussed below.

Mode of Thinking: targeted persons become part of stories where they think that they might 'lose' presumably rare goods [AS15].

Best Practice Defence: things that are scarce are not always valuable or good. Targeted persons should learn to recognise and anticipate their own experience when something becomes scarce. The anticipation of this "arrousal" can calm down the decision-making [MGM14].

Self-Questioning: "is this still an attractive offer if it wasn't scarce?" [Bul+18]

Commitment & Consistency

"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 will honor it, which increases the likelihood of compliance." [Bul+15] The foot-in-the-door technique is based on this principle. Once targeted persons agree to a small commitment, they would more likely to stick to that commitment when asked for a bigger one. The effect is higher if targeted persons' commitment is in written form or visible for the public.

Mode of Thinking: targeted persons want to stay consistent with their decisions in their stories [AS15].

Best Practice Defence: (i) targeted persons should learn to identify when being influenced by this principle, e.g., listening to their feeling not wanting to comply. (ii) Before any response, targeted persons should imagine how they would feel when looking back to a complied request [MGM14].

Self-Questioning: "if I could go back would I do the same thing?" [MGM14]

Authority

"is the principle that 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 behavioral trait of responsibility transition." [Bul+15] For instance,

upon receiving a call from new customers, real estate agents referred to colleagues that they advertised as experts in a specific field. The real estate agency experienced a higher rate of appointments and signed contracts. Anecdote 4.10 translates the warning notice of Figure C.6.5 about fake police calls to retrieve valuables from targeted person. Police officers as trusted authority figures are impersonated to get targeted persons comply. The authority refinements will be discussed below.

Mode of Thinking: to become a part in authority stories, targeted persons must pay deference to authority figures [AS15].

Best Practice Defence: targeted persons should examine the validity and credibility of the authority they faced by the following two questions [MGM14]:

Self-Questioning: "(1) is this person truly an expert? (2) How truthful do I expect this person to be based on his or her position?" [MGM14]

> **Anecdote 4.10 — Grandparent Scam: Call by Fake Police Officer.** In Figure C.6.5 the State Office of Criminal Investigation (LKA) of Lower Saxony, Germany, warns about criminals impersonating police officers. The fake police officer (attacker) calls the targeted persons telling the story of an arrested burglar who carried a note of names and addresses. One entry contains the details of the targeted person. The attacker asks for help to capture accomplices and about valuables that might get stolen. As a precaution the attacker offers to collect valuables that the (fake) police will keep safe. Sometimes the attackers discredit banks as they might be involved in this criminal activity to convince the targeted persons to avoid banks for safekeeping.

Robin Sage & Anna Brett

Ryan and Mauch [RM10] conducted a 28-day experiment in 2010 where they created the fake identity of a young, attractive female called 'Robin Sage' in various social networking sites and on mailing lists. The fake identity pretended attending educational facilities which helped alumni and alumnae to gain trust (*Liking/Similarity*) without even having met her (SE Indicator 3.5). She posed as a 'Cyber Threat Analyst'. The researchers were able to obtain sensitive information violating Operations Security (OPSEC) and Personnel Security (PERSEC) procedures in this short time period. Besides *Liking* the cognitive bias *attribute substitution* (Section 2.6) may be involved because of her appearance [Thi16]. The communication style was bidirectional [Mou+14a] (SE Indicator 3.2) since a return channel for communication was needed to establish trust. The targeted persons were unaware (SE Indicator 3.3) and the key enabler to provide sensitive information (SE Indicator 3.1). The fictional malicious attack (SE Indicator 3.4) in Anecdote 4.11 is based on the Robin Sage experiment. There, the attacker uses a fake job title to persuade targeted

persons to install malware unknowingly. The job title as well as position in an official authority means that the attacker may use the refined principle *Authority by Knowledge* (Section 4.5.2). Also, because the targeted person may perceive the worm as an urgent matter, *Scarcity of Time* principle (Section 4.5.2) can be identified.

> **Anecdote 4.11 — Fake Social Networking Contact.** A targeted person gets contacted via a business networking site. An attacker writes that attacker visited the same college as the targeted person but during different years. The attacker tells an old story about one teacher and asks the targeted person if the targeted person remembers this teacher. The attacker's profile shows involvement with a law enforcement agency and an attractive person with similar interests. The targeted person cannot remember seeing the attacker on campus, but that's because the attacker attended the college later.
>
> After a few brief messages, the attacker reveals that at the moment they are examining a wide-spread worm encrypting hard drives to require ransom payments. One security company was able to detect it already and hopefully others will follow soon. The attacker should not talk about it but since the conversation to the targeted person is so kind, the attacker reveals the link to the beta version of a security tool for detection. The targeted person installs the tool and runs it resulting in no infection. The tool itself contains a malicious payload unknown to the targeted person and infects the targeted person's computer.

In a related study Huber et al. [Hub+09] developed a proof-of-concept bot for social networks that gathered sensitive information by chatting to targeted persons. In this so-called "automated social engineering" the bot started bidirectional communication intentionally with prior identified targeted persons of an organisation. One created fake identity was *Anna Brett*, a 22 year old student searching for the 'Royal Institute of Awareness'. Her profile stated that she was single and attractive pictures were chosen intentionally. The bot found ten male singles of that institute on Facebook (pre-attack information gathering). With these ten targeted persons an automated chat was initiated about Anna Brett's alleged interest for the institute. For those who answered, the 'bonding' was successful and the 'execute' phase started: the targeted persons were asked to answer an online questionnaire for a friend allegedly pursuing a PhD at the 'Cambridge Computer Laboratory'. If the targeted persons followed the link to the 'survey on password security' within three weeks, the bot was successful. In comparison to the Robin Sage experiment, if this fully automated chat bot would have succeeded as it initially was designed, the proof-of-concept worked. Similar to Robin Sage, the fake identity would apply the persuasion principles of *Liking/Similarity* as well as *Authority by Knowledge*. That is, with this automation of chat bots using SE (deceptive techniques, persuasion principles), these types of attacks may scale and become cheaper for the attacker. However, due to the corresponding universities lacking a Research Ethics Board (REB), the experiment design was changed. The second part of a Turing test with students revealed that the majority of subjects were able to distinguish between chatterbot ('Anna') and human ('Julian') conversation.

4.5.2 Refinement of Principles

The principles of *Authority* and *Scarcity* can be refined into subprinciples as follows.

Scarcity of Time

When time is scarce, time pressure occurs when having to make decisions. This is a common and questionable practice in sales to let some products appear to be out of stock soon. Similarly, displaying the number of (competing) customers viewing the same product can motivate to buy instead of reevaluating the decision properly. The phishing e-mails in Figure 4.3 or Figure 4.4 demanded a reaction time of 24 hours of the targeted person to avoid the deactivation of a bank account or e-mail account.

Buying a home is for most people a once-in-a-lifetime event spending a lot of money with not a lot of experience. The main focus of buyers is on the new home and hopefully finding an agreement with a significant other. The time for financing this endeavour is tight: finding a bank and getting a credit. The customers are experiencing time pressure and a lack of focus due to an overwhelming amount of open tasks which can be exploited, for instance, by unnecessary transaction fees [AS15, ch. 4]. Framing can be used in this case as well to change the perspective whether the seller pays the fees or the buyers. In sum, the price should be the same, but our perception of the fees can change in relation to the down payment [AS15]. Anecdote 2.1 describes the situation in the German real estate market before the 'tax credit for first-time home buyers' (Eigenheimzulage) were about to discontinue. Home buyers experienced time pressure to buy a home to still receive the tax credit.

Scarcity of Information

Information can also become less available. The attitude to value scarce resources higher comes in handy, when an attacker creates information not available to everyone, but many know that it exists. A targeted person may desire to get hold of that information depending on its content.

Authority by Knowledge

The perceived knowledge of a person may convince a targeted person to comply. An attacker can pose as knowledgeable with fake academic titles or even refer to a knowledgeable person. The Milgram shock experiment [Mil65] demonstrated that with a lab coat and a scientific setting, 66% of the subjects were obedient to administer electro shocks to human test subjects (actors) that would have been lethal.

Authority by Hierarchy

The physical appearance of an attacker can transport an aura of authority, e.g., by wearing a police uniform, work clothes or specific status symbols [Bul+15]. By impersonating someone in an organisation's hierarchy that has privileges, an attacker can exploit that position. For instance, Anecdote 9.5 will discuss the role of janitors' work clothes. As mentioned above, an attacker may refer to a person in a higher management position or state to have an appointment in respect thereof (see tailgating Anecdote 1.2 at the TUHH). Anecdote 4.7 shows a phishing attempt where a CISO was impersonated to lure employees into entering their login credentials in a phishing website. The case of Willi Herold documents the magnitude and its consequences of the Authority by Hierarchy principle (Anecdote 4.12).

Anecdote 4.12 — Standgericht Herold (Wös15). 1945, end of the World War II, Willi Herold, a chimney sweeper from Chemnitz, found a highly decorated uniform of the Nazi military police (German: Feldjäger). He was able to command soldiers by impersonating the military police. He took power of the Emsland camp "Aschendorfermoor". People obeyed his orders and mass executed prisoners and members of the resistance. The so-called "Standgericht Herold" (drumhead court-martial) was later captured. Herold and five accomplices were sentenced to death in 1946 by the British allies. [Wös15]

4.6 Social Engineering Personality Framework

Besides susceptibility to SE originating from human nature or cultural background (Hofstede's human mental programming [Hof01], Figure 2.2), the personality of targeted persons has an impact whether they become a victim. In Section 2.8 the Five-Factor Model (FFM) personality traits and their probable influence on susceptibility were described. A literature review about SE susceptibility related to personality traits resulted in the development of the Social Engineering Personality Framework (SEPF) [Qui13; UQ14]. For instance, Parrish Jr et al. [PBC09] were consulted who developed a model to look into susceptibility towards phishing and personality traits. The SEPF analysed which personality traits make targeted persons more vulnerable to specific persuasion principles. Hirsh et al. [HKB12] called their approach *personalised persuasion* when framing messages are constructed according to each personality profile. That is, Cialdini's persuasion principles (Section 4.5.1) were mapped to the FFM personality traits, see Figure 4.7. Furthermore, Uebelacker and Quiel [UQ14] suggested *coping strategies* (Cialdini just called them 'Defense' [Cia07]) for each personality trait to minimise the susceptible surface:

Conscientiousness:

targeted persons high on the conscientiousness trait prefer to follow rules, policies, and social norms. Conscientious targeted person are motivated by achievement, order, and efficiency [HKB12]. Techniques that exploit such rules and motivations can make use of the persuasion principles authority, reciprocity, and commitment & consistency [UQ14]. However, if sensible security policies exist that also address SE, e.g., with a suitable behavioural codex, conscientious targeted person may not be more susceptible.

Coping Strategies: the authors suggested sensible and comprehensible policies and awareness trainings adapted for each employee, especially for reactions involving time pressure (scarcity of time). An organisational culture that supports and fosters security-related behaviour can be beneficial, e.g., when a targeted person reporting an incident does not experience a blame and shame response [UQ14].

Extraversion:

sociability, a subtrait of extraversion, is related to the principles liking and social proof (conformity) [UQ14]. Another subtrait is 'excitement seeking'. Modic and Lea [ML12] showed that extraversion and sensation seeking correlates with scam

compliance. Obtaining something scarce can be assumed as exciting, adding higher susceptibility towards the principle of scarcity. Contrarily, applying the commitment & consistency principle would most likely be unsuccessful because extraverted individuals show a lower preference for consistency [UQ14]. In summary, targeted persons with a higher level of extraversion may tend to be more vulnerable to liking, social proof, and scarcity as well as less vulnerable towards commitment & consistency principle-based SE.

Coping Strategies: rewards and social attention are motivating extraverted individuals [HKB12]. Hence, addressing the motivational system can include rewards for achieved trainings including a ranking system in comparison to not mentioned colleagues with optional posting of results to the internal social network; organisation-wide system for improvement suggestions that display the name of the suggester on request [UQ14].

Agreeableness:

'trust'[17] is one subtrait of agreeableness. Subjects who are more trusting are less concerned about privacy issues, such as revealing one's location via location based services [JS06] or online information sharing [TJ15]. Thus, subjects would react to SE attempts with a more gullible behaviour (Section 4.3.1), e.g., disclosing sensitive information at ease or trusting a detrimental relationship to an attacker. Uebelacker and Quiel [UQ14] suggested that the trusting nature opens doors for SE attacks based on the persuasion principles authority, reciprocity, liking, and social proof [UQ14]. The motivational system of agreeable subjects consists of valuing communal goals and interpersonal harmony [HKB12]. An attacker may use an importunate attitude to exploit the harmony-seeking motivation [UQ14].

Coping Strategies: awareness training should tell stories about experienced SE attacks to improve the 'who to trust' perspective [UQ14].

Openness to Experience:

Modic and Lea [ML12] showed that openness correlates with scam compliance. Subjects with a higher degree on openness are more susceptible towards SE using scarcity [UQ14]. If targeted persons feel that perceived options and freedoms become scarce, they will try to counteract (see loss aversion, Section 2.3.1). No definite lower or higher vulnerability was assumed for other principles: subjects would be more susceptible if strong fantasy were involved, but also less susceptible with respect to computer proficiency [UQ14].

[17] The term trust can be found in various disciplines and even in one discipline it may be understood differently. Here, the personality subtrait in psychology is meant obviously. "In social sciences, it [trust] is a property of the system that forms as a result of interaction between agents of that system." [FRS05] In computer science trust was introduced and used in different contexts. After a literature review, Grandison and Sloman [GS00] defined trust for internet applications as "the firm belief in the competence of an entity to act dependably, securely and reliably within a specified context" (they assume dependability covers reliability and timeliness). However, the abundant use, e.g., in trusted computing, trust management, or trusted code has been disputed because of its "manifold and contradictory meanings" in "Why Trust is Bad for Security" [Gol06].

Coping Strategies: Uebelacker and Quiel [UQ14] see edutainment and gamification approaches as supportive tools to reach subjects open to experience. The subjects should play a creative part in these trainings sessions to address the motivational system of creativity combined with intellectual stimulation [HKB12].

Neuroticism:

For phishing, a SE technique, Parrish Jr et al. [PBC09] see neurotic targeted persons less vulnerable due to computer anxiety. They behave more carefully because of their underlying pessimism [UQ14]. On the other hand, targeted persons with a higher degree in neuroticism are motivated by threats and uncertainty [HKB12; Leu17; ML12]. An attacker may use these motivations for attacks, such as applying the authority principle [UQ14]. However, the SEPF authors expected no direct influence per principle, but rather a general assumption of a smaller susceptibility, although targeted persons may be more vulnerable to authority (threats) and scarcity (uncertainty of scarce things).

Coping Strategies: Uebelacker and Quiel [UQ14] had no suggestions here. Trainings and policies for neurotic targeted persons may be able to address desired reactions and behaviour in threatening or uncertain situations.

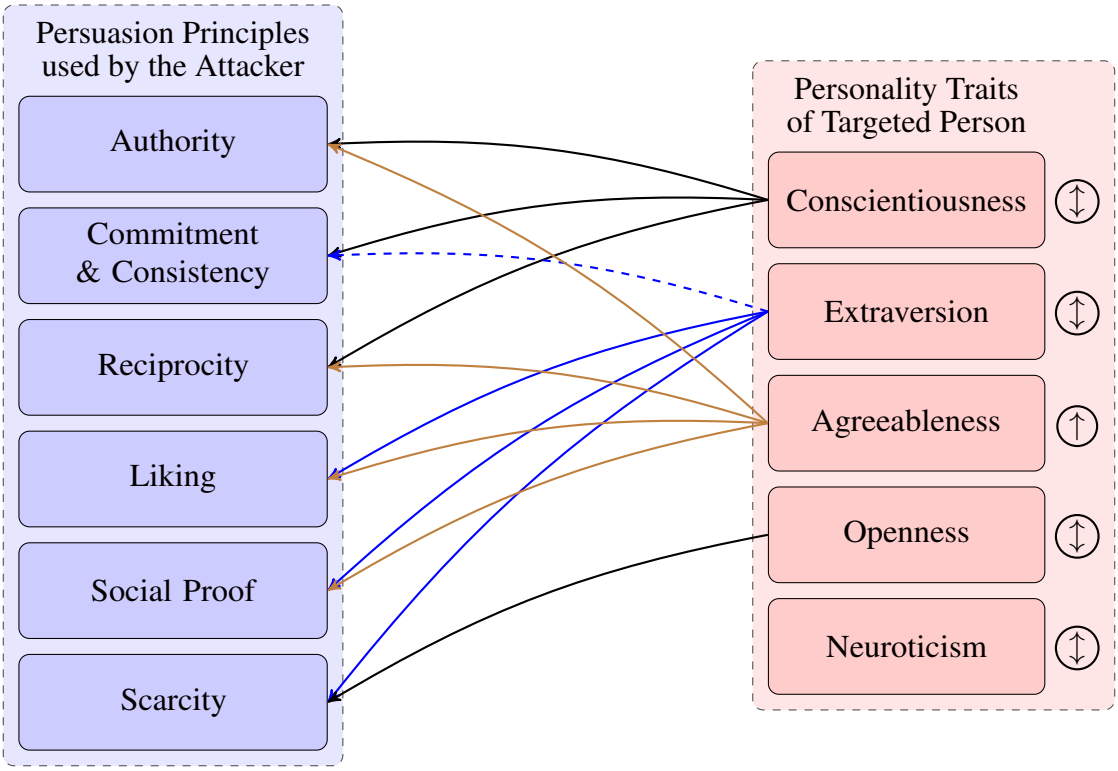


Figure 4.7: **SEPF**: Specific FFM personality traits of a targeted person may increase (solid line) or decrease (dashed line) the susceptibility to Cialdini's persuasion principles. Some of which can be used by an attacker (for better readability some arrows are coloured). General personality assumptions about susceptibility (higher, lower, or both) for each trait are depicted by corresponding arrows (↑, ↓, ↕). Figure adjusted from Uebelacker and Quiel [UQ14] (general susceptibility for neurotic individuals changed; naming convention applied)

II

Social Engineering Evidence

5. Evidence-Based Research

Section 1.7 elaborated why sources are called “anecdotes”. This term was chosen after questioning the quality of evidence with a quite logical empiricism view [SMP17]. There, anecdotal evidence expresses one of the weakest forms of scientific evidence: it could comprise story-telling or hearsay, non-representative or non-reproducible cases with volatile falsifiability, and somehow cherry-picked and biased. On the other hand, the court documents to be presented in Chapter 7 can lead to insights from a truth finding process; controlled experiments or questionnaires can be verified if data are available and also repeated. In Social Engineering (SE) research where manipulation and deception of subjects play a major part, ethical questions arise (Section 5.2.1). To subsume all stories found under one level of evidence, the category with the lowest scientific claim was deemed appropriate: anecdotes.

In many scientific disciplines finding evidence and drawing the correct conclusions from it, is a standard requirement, but in different nuances. Besides possible misinterpretation and erroneous analysis of data or misconfigured experimental design, incorrect conclusions may lead to serious consequences. The human factors and SE are intertwined inextricably (SE indicator 3.1). The factors leading to successful SE have been analysed in disciplines like psychology. The SE research in this thesis examines empirical data, mainly in the form of anecdotes.

Traditional medical research is based on the interpretation of empirical data and can be seen as a related field for some aspects of SE. Evidence-based medicine relies on practices how to use evidence for a more detailed view epistemologically. It demands a stricter classification to draw conclusions. Some countries and medical organisations introduced definitions for the level of evidence regarding the strength of acquired data sources and practices. These standardised levels range from Randomised Controlled Trials (RCTs) over cohort studies to expert opinion. However, these hierarchies of evidence strength are

also criticised: due to the narrow scope of RCT as well as its high internal validity, but hard to draw generalisable conclusions, its ranking as a "gold standard" [Car07] has been disputed. One should not forget the etiology part of medicine: discovering root causes (cf. general knowledge discussion in Section 5.2).

This chapter will start with an introduction to the philosophy of science (Section 5.1), followed by an overview of current research on a "Science of Security" in Section 5.2. Section 5.3 will discuss the knowledge base needed for cybersecurity research. Finally, Section 5.4 will close this chapter with evidence-focused SE research before the next chapter provides plausible sources of SE events.

5.1 Philosophy of Science

Philosophy of science creates a meta-level based on other disciplines of science: "a second-order reflection upon the first-order operation of the sciences"[1] [SMP17]. In the last century, the Vienna Circle challenged the then predominant metaphysics among other things by focusing on the so-called *logical empiricism*. They grounded science on logic combined with empirically gathered information.

One of their two main tenets covers *empiricism and verification*: how to verify scientific insights to show us anything about the world. Moreover, how can these insights be converted into mathematical and deductive logical statements. An achievement would comprise developing a "framework of general knowledge" based on first-order logical statements [SMP17]. This criticises inductive reasoning as proposed by David Hume in the 18th century. Hume's problem of induction deals with the challenge to infer logical conclusions from sheer empirical observations. For instance, there is no proof that the sun will rise tomorrow even if humans have witnessed it for millennia. Cases exist where the "laws" of physics need redacting, caused by new empirical findings. These laws may validate an abstraction (model) of the world at one point in time, but not the world itself.[2] So, what can these laws tell us about the world at all? Karl Popper questioned the verification of such logical statements and proposed that falsifiability of scientific claims becomes possible.[3] He stated that inductive science does not exist. His proposal imposed limitations upon many scientific disciplines. Kuhn's view on science differed and he objected in 1962 to the assumption that science works on the foundation of logical statements.[4] Instead, scientists use cognitive models for working on paradigms [SMP17]. To understand the empirical practice in science, methods describing phenomena can be seen as "mechanistic explanations".[5]

The second tenet of logical empiricism focuses on the *reduction of science* which emphasises a unified understanding of science. One facet explores the transferability of reduced logical statements, such as logical abstractions, from one discipline to another. The other

[1] "Vienna Circle" in the Stanford Encyclopedia of Philosophy, 2016, as cited by Spring et al. [SMP17]
[2] Cartwright, N., "How the Laws of Physics Lie", 1983, as cited by Spring et al. [SMP17]
[3] Popper, K.R., "The Logic of Scientific Discovery", 1959, as cited by Spring et al. [SMP17]
[4] Kuhn, T.S., "The Structure of Scientific Revolutions", 2012, as cited by Spring et al. [SMP17]
[5] Glennan, S. and Illari, P., "The Routledge Handbook of Mechanisms and Mechanical Philosophy", 2015, as cited by Spring et al. [SMP17]

covers the unity of methodologies between disciplines through reduction, e.g., how to evaluate evidence. Spring et al. [SMP17] warn against applying reduction. It can deprive scientists of developing accurate tools supporting their own discipline better. By reducing a complex system its representation will be reduced as well. Mitchell argued that our representation should be equally complex, resulting in a pluralism of models — a science of complexity.[6] Nowadays, philosophy of science names *integrative pluralism* (Mitchell) and *mosaic unity* (Craver) as its most recent approaches. Disciplines can agree on adding constraints to their models to coordinate with other disciplines, offering common *interfield explanations* (Darden/Maull) [SMP17].

5.2 Science of Security

When classical scientific approaches like logical empiricism are forced upon cybersecurity, researchers may struggle and their scientific insights are questioned. The classical view on science as natural sciences identifies physics as a gold standard. For a Science of Security with respect to SE, the Science Council's definition can be of assistance (Definition 5.1), i.e., the design of a "systematic methodology based on evidence" [SC09]. Spring et al. [SMP17] collected and summarised the five most discussed obstacles for a practice of a science of security. Spring et al. see these obstacles as misguided and showed how to overcome them (see below).

> **Definition 5.1 — Science (SC09).** "Science is the pursuit and application of knowledge and understanding of the natural and social world following a systematic methodology based on evidence."

Untenable experiments:
security experiments are criticised to be of untenable nature, e.g., being too risky or unethical in practice. Three main reasons are mentioned: missing proper control (laboratory vs. real-life experiments), prohibitive privacy requirements and ethical constraints, and fast technological change through new developments. Section 5.2.1 for more.

Reproducibility is impossible:
in cybersecurity, controlled environments to repeat experiments or other research to validate results are not feasible. The evaluation effort in computer science comprises the categories repetition, replication, variation, reproduction, and corroboration according to Feitelson [Fei15].[7] More in Section 5.2.2.

No laws of nature:
the criticism relates to the demand that security may be called a science iff 'laws of nature' exist and are applied similar to physics (in the classic meaning of the term 'science'). Law-based systems can be broken down into universal laws, conditional

[6] Mitchell, S.D., "Biological Complexity and Integrative Pluralism", 2003, as cited by Spring et al. [SMP17]
[7] Feitelson [Fei15] as cited by Spring et al. [SMP17]

laws as in causal statements (classical first-order logic), and into laws with confirming or falsifying specific abilities. An abstraction of reality (Mitchell[6], 2003) leads to explanation gaps hindering researchers to draw generalisable conclusions: 'all models are wrong, some are useful' [SMP17]. Because of the intrinsic simplification of 'laws', law-based systems can hardly support the causality demand for real world assumptions (Cartwright[2], 1983). Spring et al. [SMP17] repudiate this very logical empiricists' claim to present 'laws of nature' in security research.

No single ontology:
this obstacle suggests that a common ontology must be achieved for each science, so it must for security. That is rarely the case in other disciplines. But what is needed are "clarity of expressions", not a single language. Spring et al. [SMP17] suggest an "integrative pluralism" where translations between disciplines exists. For instance, Section 4.1 explains the "Persuasion Knowledge Model" [FW94] whose actors are an "agent" and a "target". For reasons discussed in Section 1.6, the attacking party 'attacker' (Definition 1.2) and the targeted one 'targeted person' (Definition 1.3), which translates to "agent" and "target" for Friestad and Wright [FW94], were chosen. Terminology differs per discipline as well as the interpretation of terms changes over time within one discipline (interpretation drift). If the terms used represent in clear, unambiguous words their meaning in the corresponding context, no harm should be done.

'Just' engineering:
Security engineering has already its place in security research. Its goal is to create dependable systems when facing "malice, error, or mischance" with a focus on "tools, processes, and methods" for designing, implementation, and testing [And08, p. 3]. With the criticism that security research is 'just engineering', security research becomes negatively connotated. Simultaneously, the criticism assumes science is not existent in security research based on the critics' definition of science [SMP17].

5.2.1 Claim: Untenable Experiments

Egelman et al. [Ege+13] conducted one experiment in a laboratory environment to examine the impact of password meters on choosing a strong password. Password meters motivated users in better password choices in the laboratory. Then they compared the outcomes to a second experiment in a real-life scenario: it revealed that password creation behaviours were "heavily dependent on context" [Ege+13], meaning that subjects created stronger passwords for accounts perceived as important, but not for all accounts. Without this additional control group, drawing conclusions based on the 'sterile' laboratory experiment would have lacked crucial information. Hence, experimental design is of uttermost importance. The comparison of different designs enables researchers to identify feasible and epistemologically helpful approaches. For instance, Dimkov et al. [Dim+10] conducted experiments for physical penetration tests involving SE (more below in "Ethical Considerations") and concluded that their custodian-focused methodology fosters more realistic results than their environment-focused one.

In a logical empiricist's world, a Randomised Controlled Trial (RCT) setup represents

experiments to falsify theories deductively. RCTs play an important part in evidence-based medicine, but in security research RCTs complement other approaches — like case studies or model-based reasoning — where applicable, cf. discussion about falsifiability in the Philosophy of Science (Section 5.1). For experiments in security research where *control groups* like A/B tests are feasible, they facilitate profound statistical analyses. Especially, when security connects with other disciplines, e.g., social sciences or psychology, RCTs offer useful and well-studied methods and techniques [SMP17]. One challenge for experimental design comprises whether to choose a laboratory environment with fewer influencing factors or a real-life environment including various distractive and biased factors. Security research is not alone in this battle, but must clearly beware of drawing premature conclusions. To overcome the missing control complaint, Spring et al. [SMP17] recommended to identify a "usable intellectual structure" in a broader set of structured observations such as qualitative research methods besides natural experiments.

How to discover general knowledge if scientists are facing *rapid technological change* which may hinder repeating an experiment or observing a phenomenon more than once? Unrepeatable phenomena are nothing new in other disciplines, for instance the extinction of dinosaurs. While in medicine the approval of new drugs can take years using RCT, SE experiments in cybersecurity can work at a quite different pace. The speed of conducting experiments that include human beings may not scale as results may suffer from side effects, e.g., if unwanted time pressure is applied to questionnaires. But other techniques can improve the speed like when blood tests or wearable body sensors become ubiquitous and affordable at the same time. Hence, medical experiments are also gaining speed as well as providing more precise data due to technological advancements. The rapid technological change makes them not less valuable for science. Computer science offers not only for itself automatable features enhancing the speed of analysis. The human subjects, however, cannot be hurried easily by technological advancements, contrary to technical analysis methods — especially if the goal is to examine long-term (side) effects. This may explain the missing longitudinal studies in cybersecurity research [Leu17]. Spring et al. [SMP17] stated that a quickly changing technology landscape may challenge the generalisation of results. "Generality turns out to be hard to find, and highly valued when it is" [SI18] and is "something much less than universal" [SI18]. But generalisation tactics crossing research fields can build general knowledge and result in the understanding of similarities and differences, e.g., of paradigms [SI18]. For SE research where the human factor is an inextricable part (see SE Indicator 3.1), rapid change in the technology landscape challenges the repeatability of SE experiments, in particular when technology is important for SE.

Another challenging aspect in security experiments are *ethical and privacy constraints*:

Ethical Considerations

In this research where malicious deception plays a major part, it is a concern of how to conduct experiments ethically and results to be reproducible for validation or invalidation. Furthermore, scientific insights should be shareable ethically. Reproducibility will be covered in Section 5.2.2.

Normative ethics express two major viewpoints of what is seen as ethically 'right' or 'wrong': *utilitarianism* as the the predominant type of the so-called consequentialism

and *deontology* [MMV13]. In *utilitarianism* the consequences of one's action justify an approach, similar to "the end justifies the means"[8]. That means for utilitarians that applied to research, the results or the outcome outweigh the means if the society benefits — even if the subjects (minority) are left with disadvantages. Then an approach would be deemed ethically acceptable. *Deontologists* have a different view and consult the act itself (here, e.g., the experimental apparatus) to determine their ethical view, e.g., in an experimental setting to always ask and get informed consent by the subjects — even if the general outcome would be hindered. This makes deception experiments harder to design where the SE susceptibility is of interest. Based on these two approaches, Mouton et al. [MMV13] examined in "Social engineering from a normative ethics perspective" three main environments SE can be executed in:

(i) Public communication (for entertainment purposes),
(ii) Penetration testing, and
(iii) SE research.

The two latter environments are applicable here — although they could be combined as both want to enlighten the field of SE. All environments have in common that the operator's intention is not to harm the deceived subjects. According to the SE Indicator 3.4 (Attacker: Malicious Intent with Goal), these environments must, however, mimick a malicious intent. Hence, the operator impersonates an attacker creating a fictitious malicious intent as they are not exactly malicious for ethical (and probably other) reasons. This may influence the experimental outcome if the targeted person (test subject) perceives the ficticious attack differently from a real attack. For instance, when using mock-phishing to find out how many employees are susceptible, the fictitious attack must be almost indistinguishable from a real one (hosting phishing websites on dubious external services or hacked Wordpress instances etc.) to make the insights count and equally minimise any counterproductive effects [MS17b].

How different experimental designs can combine ethics and proper deceptive research is discussed in Dimkov et al. [Dim+10]. They cover two methodologies of how to design and conduct physical penetration tests as SE experiments. The typical penetration tests are conducted in the digital realm, e.g., using network vulnerability scanners. Physical ones switch to the physical domain, for instance, penetration testers try to enter restricted areas in a building without proper access privileges. The results could show how employees react to seeing unknown persons without a compulsory company badge or discovering that someone is trying to pick a door lock. Dimkov et al. [Dim+10] defined actors such as employee (targeted person), custodian (asset owner), security officer (responsible for organisational security), penetration tester, contact person for the targeted person (only in custodian-focused setup), and coordinator (experiment owner; only in custodian-focused setup).

The *environment-focused* methodology contains *one* deceptive act of a penetration tester against an employee. The employee as a targeted person is tested to hand over an asset (here: laptop) which belongs to a so-called custodian to an unknown person (attacker).

[8] first known appearance in Ovid's Epistulae Heroidum as "exitus acta probat" [Ovid] which translates to "the outcome justifies the deed"

Whereas the *custodian-focused* setup adds the custodian (laptop owner) as another targeted person of the tester's deception. Neither the employee who is asked to hand over the laptop, nor the custodian know about the true intentions of the experiment. That is, the penetration tester tries to deceive the custodian *and* the employee. The custodian's reaction may influence the experimental results; hence, the deceptive interaction is widened. Besides an additional experiment coordinator, a contact person is introduced who knows about the experiment and is contacted in case of an emergency. This contact person knows about the deception and deceives the custodian. This refinement of the environment-focused method can reveal the custodian's security awareness as well. Dimkov et al. [Dim+10] evaluated criteria, such as reliability, repeatability (see Section 5.2.2), reportability, two forms of *respectfulness*, and realism. Concerning ethics, respectfulness regarding actors and trust relations are applicable here. The "respect for people" (one ethical principle in the Belmont Report [PR78]) arises from the concern that a direct[9], physical, deceptive interaction with subjects is even more intense than using a digital medium. Respect for the deceived actors contains, for instance, a standard procedure for debriefing excluding no one. The existing trust relationship between employees should not worsen due to the experiment.

Finn and Jakobsson [FJ07] justified deception for specific cases:

(i) The experiment's risk[10] is minimal for the targeted persons regarding the physical and psychological harm. Also, the welfare and rights of the subjects are not violated.
(ii) The deception is an essential component for the outcome. That is, subjects cannot agree to the deception beforehand. Thus, the subject needs to be debriefed afterwards to alleviate this issue. However, in some cases debriefing can cause harm.
(iii) The expected knowledge is deemed to be of scientific importance.

Dimkov et al. [Dim+10] and Finn and Jakobsson [FJ07] cited the quite general Belmont Report [PR78] for experiments in biomedical and behavioural research. It describes the three main principles of *respect for persons*, *beneficence*, and *justice*. Respect for persons can be accounted for the normative deontology approach and beneficence for the utilitarianism approach [MMV13]. While the Belmont Report was published in 1978, more recent publications, e.g., Thomas et al. [Tho+17], focus on the 2012 Menlo Report [KD12] covering ethical research and privacy concerns in Information and Communication Technology (ICT). The Menlo Report is based heavily on the Belmont Report. It adjusts its principles to ICT research as "legal restrictions and requirements have expanded considerably since the 1980s" [KD12]. The Menlo Report introduces a fourth principle "Respect for Law and Public Interest" [KD12, Section C.5]. This principle for ICT was included implicitly in the Belmont Report's 'beneficence' principle. Kenneally and Dittrich [KD12] created it explicitly for addressing issues of compliance, transparency, and accountability. They see it as an important factor to avoid problems of credibility, trust or confidence in presented findings. An overview of all principles can be found in Table 5.1.

[9] Dimkov et al. [Dim+10] used the term 'direct' as in direct communication with a different meaning than used in SE definitions, cf. Mouton et al. [Mou+14a] (Indicator 3.2, Section 3.2.1, Figure 3.2) or the SE Definition 3.17. 'Direct' means here the bidirectional communication and interaction in the *physical* domain solely.

[10] Finn and Jakobsson [FJ07] distinguish later between actual and perceived risk.

Principle	Application
Respect for Persons	"Participation as a research subject is voluntary, and follows from informed consent; Treat individuals as autonomous agents and respect their right to determine their own best interests; Respect individuals who are not targets of research yet are impacted; Individuals with diminished autonomy, who are incapable of deciding for themselves, are entitled to protection." [KD12]
Beneficence	"Do not harm; Maximize probable benefits and minimize probable harms; Systematically assess both risk of harm and benefit." [KD12]
Justice	"Each person deserves equal consideration in how to be treated, and the benefits of research should be fairly distributed according to individual need, effort, societal contribution, and merit; Selection of subjects should be fair, and burdens should be allocated equitably across impacted subjects." [KD12]
Respect for Law and Public Interest	"Engage in legal due diligence; Be transparent in methods and results; Be accountable for actions." [KD12]

Table 5.1: Proposed guidelines for ethical assessment by the Menlo Report [KD12]

Considerations on Privacy

One may claim that prohibitive privacy requirements result in untenable experiments. The Menlo Report [KD12] covers privacy in their ethical guidelines as well as respecting subjects. By explicitly offering an ethical approach they guide researchers how to conduct ethical experiments. Their basic guidelines may foster the publication and acceptance in the scientific community as well as sharing results lawfully and ethically. These guidelines enhance the standard scientific method and should be a prerequisite when designing SE research, especially deceptive experiments.

5.2.2 Claim: Impossible Reproducibility

Dimkov et al. [Dim+10] questioned whether SE penetration tests are *repeatable* because human behaviour is "unpredictable". The question whether insights can be reproduced, drives researchers to determine how to reflect on their research: one might wish to get confirmation of the results, others may like their experimental design to be used and enhanced, and some may want to offer fellow researchers the ability to 'disprove' one's theories. Feitelson [Fei15] suggested a terminology to distinguish between types of redoing experiments: *repetition*, *replication*, *variation*, *reproduction*, and *corroboration*. Under *repetition*, *replication*, and *variation* Feitelson counts experiments that are rerun exactly as the original one. *Repeated* experiments use the original artefacts, *replicated* apply recreated artefacts to the previous experimental apparatus, and *variated* ones modify a parameter in a measurable manner intentionally. To attempt regaining the insights of another experiment, *reproduction* builds on the same experimental idea with a similar setting and same procedure, but allows "newly created appropriate experimental apparatus" [Fei15]. Whereas,

corroboration approaches find evidence differently on various levels (refute hypothesis, reveal limitations of results and techniques). Corroboration delves even deeper regarding scope and generalisation, beside not being bound to the predetermined procedures. That is, it can vary the techniques as well. For increasing confidence in scientific results, Feitelson [Fei15] sees corroboration as the best approach. It can additionally evaluate the confidence and limitations of different procedures. Feitelson showed which use cases with their requirements can support which expected outcome or research question, including the level of generalisability and scope [Fei15, Table 2].

A different aspect, not covered in this criticism on reproducibility, are selectively reported results. Insights of repeated experiments are underrepresented in scientific venues especially if they do not reveal any novelty [SMP17]. But they are extremely important for follow-up research based on it. However, selectively published research (after acceptance by peer-review) often outweighs repeated research.

According to Feitelson [Fei15], repetitions are hardly feasible with human subjects, similar to Dimkov et al. [Dim+10]. Furthermore, repetition does not provide any novel insights to generalise hypotheses. Replication can show that the replicated setup expresses the documented, original design. In repeated or replicated experiments executed with humans, the same results will most likely not be achievable due to different conditions (individual stressors etc.). Such resembling results can be analysed statistically and offer insights in that way. Through variations researchers may examine the scope and generalisability with respect to the results. Reproduction can support the confidence in not only the results and their generalisability, but also in the procedures. However, these procedures must be designed properly to avoid problems in design, e.g., sample bias, or provide sufficient data to draw conclusions at all [SMP17]. A researcher can minimise the sample bias effect by knowing the bias and then applying controls or design changes. In SE research, biases contain social and cultural aspects. For instance, a lot of research has access to subjects in WEIRD societies only (Western, Educated, Industrialised, Rich and Democratic; Section 6.3.1) [HHN10]. The types of corroboration and reproduction seem promising for SE experiments.

5.3 From Structured Observations to a Knowledge Base

The SE research in this thesis covers qualitative research. In qualitative research two main theories are recognised. The data-driven *inductive* approach where theories are grounded on prior observations. Whereas theories are tested against a posteriori gathered information in the *deductive* approach.[11] Hume, Popper, and others have militated against inductive reasoning. But induction and deduction do not need to be disjunct approaches. Mathematical modelling can generate a cycle interlocking induction and deduction by referencing each other, as depicted in Figure 5.1 [SMP17]. This cycle can be understood as a truth finding process for structured observations (observation, construction of models, deduction, and validation).

A literature review was used to formulate hypotheses for the “Social Engineering Personality Framework” [UQ14] (Section 4.6), suggesting possible correlations between

[11] talk by Alice Hutchings on “Qualitative Research”, 2017-02-23, University of Cambridge, UK

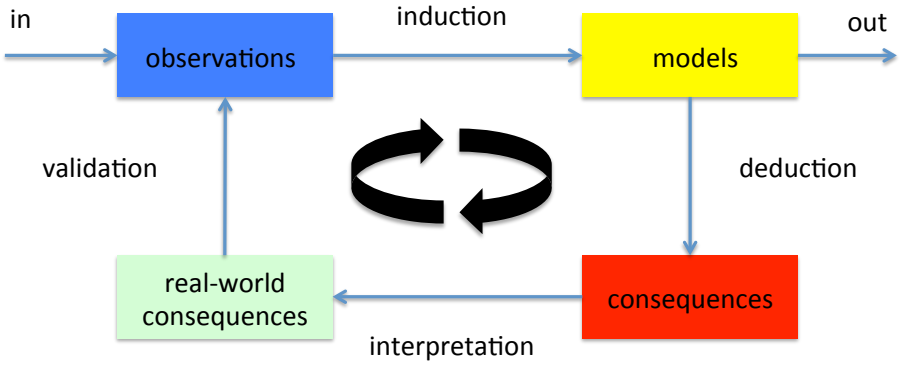


Figure 5.1: Mathematical modelling cycle for structured observations in Spring et al. [SMP17, Figure 1]

susceptibility to SE and personality traits, thus, providing inductive hypotheses for a future deductive examination. SE indicators in Section 3.1 were found through a modelling cycle similar to Figure 5.1: the first observations were based on anecdotes and alleged acts of SE. Different sources described their anecdotes as SE, some provided their SE definition as well. The created SE indicators were intended to sharpen the view on SE in this thesis. Then this 'indicator model' was checked whether anecdotes fit well or whether the SE indicators lack anything which others consider SE. After some refinement cycles of the SE indicators, it was expected to find an appropriate, existing SE definition. By developing a qualitative metric[12] ('specificity') the definitions were assessed concluding that none expressed the SE indicators explicitly. Hence, an appropriate Definition 3.17 based on the SE indicators needed to be created upon which this research can be grounded.

5.4 Evidence-Focused Social Engineering Research

With the previous sections in mind, sources of SE evidence can come in many facets and quality as well as for various scientific use cases. In this thesis the research focuses on a variety of sources which differ in quality. If SE research focused on historical empirical data solely, novel attack vectors would be ignored. One member of the TRE$_S$PASS' Technical Advisory Board mentioned that 'in security the past is a poor predictor of the future' [D2.5.1]. However, prediction may not be achievable without prior evidence-focusing insights. Thus, the first step is to search for sources of SE — SE acts that already happened (known threats) or are within the realms of possibility (expert opinion, brain storming). The latter may not be seen as high-ranking evidence or evidence at all. But just focusing on known, existing SE acts leads to a more partial view on this topic. Here, plausible, imaginable acts will be consulted to complement historical sources in the hope that this approach reduces overlooked threats. That is, the observed, but also imagined acts could fit with different starting points very well into the above cycle for building a knowledge base (Section 5.3). Of course, the quality differs, but these hypothetical acts foster future experimental design and the attentiveness for finding existing, similar acts of SE. A too strong focus on known, historical threats may become counterproductive

[12] Section 3.2 indicator metric of SE definitions: missing < partial < implicit < yes < too general

for revealing new types of threats.[13] In conclusion, it will be called evidence-*focused* SE research here as the first step towards evidence-*based* research.

In the following chapters possible sources will be outlined to find plausible evidence on how acts of SE have been performed.

[13] Kang, M.-C. (2013). "Responsive Security: Be Ready to be Secure" (DOI: 10.4324/9781315145907) as cited in *TRESPASS Information Testing and Degradation Tools* [D2.5.1]

6. Social Engineering Evidence

After this chapter's discussion on Social Engineering (SE) evidence from interviews (Section 6.1), literature (Section 6.2), and experiments (Section 6.3), some of the sources mentioned are extracted in the next chapters and three methods are elaborated in more detail: analysis of court documents (Chapter 7), brainstorming with Lego (Chapter 8), and the Social Engineering Poetry Slam (Chapter 9).

6.1 Soliciting Social Engineering Evidence

Interviewing someone who has experienced an act of SE seems straightforward. Experts on SE may reveal interesting insights as well. None of the anecdotes collected might even present any SE act due to the fact that the term SE is used in different meanings. It is up to the evaluation afterwards to identify SE, e.g., applying the SE indicators of Section 3.1 (Research Question 1.2). Hence, a qualitative 'expert opinion' should be accompanied by an 'expert evaluation'.

At first, one must decide *whom* to ask and *how*. First hand information can come from targeted persons who had experienced SE (maybe fell victim) as well as social engineers (active or former criminals, even serving their sentence). Interviewees can also be experts and researchers of various disciplines, for instance, criminologists. This latter group would probably provide more second hand knowledge. The method on *how* to interview the subjects must correspond to their environments and requirements, such as anonymity. Methods may be in-person interviews (Section 6.1.1), questionnaires with free text components, modelling sessions or competitions (Section 6.1.2).

6.1.1 Interviewing Active and Former Attackers

Asking Social Engineers that are active criminals is a non-trivial pursuit. If an active Social Engineer is found willing to answer questions, one may still speculate about whether the responses are honest and describe real events or not. Ethically as well as legally it is hard not to become a confidant of a serious crime *and* protect the interviewee against prosecution. The main anchor for truth finding relies on the honesty of the interviewee (as always) which can not be validated easily by the interviewer, and even harder by others. It should be examined if the information obtained maps the SE reality correctly. For other researchers it is almost impossible to repeat this kind of research or do follow-up interviews with the same interviewees. Thus, the insights of such interviews contain a strong narrative character and are of qualitative nature. They are mostly unrepresentative, but can offer an epistemological way to get an idea of attack opportunities. Interviewing offenders offers researchers an opportunity to dive into the criminal domain and supports the understanding of an offence [HH18]. Researchers can start understanding novel SE attacks on which they can base similar follow-up experiments or interviews.

Hutchings and Holt [HH18] interviewed researchers who had interviewed cybercrime offenders. They present common pitfalls and ethical safeguards. While the offender interviews of Hutchings and Holt [HH18] cover cybercrime in general and do not mention any SE, one may assume that cybercriminals apply SE techniques, too. Hutchings and Holt highlight the reasons for such qualitative interviews: "correcting misunderstandings; providing a deep understanding; identifying new lines of inquiry; accessing a hidden population; providing a voice; and perceiving quantitative approaches as being insufficient" [HH18, page 79]. The researchers conducted interviews in-person or via online video conferencing with semi-structured techniques. Face-to-face interviews took place in public locations, e.g., cafes. Recorded interviews were later transcribed. The interviewees were recruited in online forums, Internet Relay Chat (IRC), conferences etc. and sometimes using snowball sampling.

Some funding organisations require to publish the research data as Open Access. This aspect in addition to further ethical issues like guaranteeing pseudonymity, had a major influence on their publication. Some researchers may want to protect their criminal interview partners from law enforcement (and avoid retaliatory attacks against themselves). Hence, data was de-identified enabling the researcher to neither confirm or deny interviewing a specific subject [HH18]. The researchers required informed consent and communicated that they only release data to law enforcement if requested by law (subpoena etc.). As a safety measure, Hutchings and Holt [HH18] advised the offenders not to talk about future activities.

One ethical question arose about the incentives that would motivate cybercriminals to give interviews. Although some candidates asked for financial compensation, researchers feared negative results as well as ethical concerns raised if cybercriminals were paid. Some subjects were motivated just "to safely have their voice heard" [HH18]. Good insights on how attacks work and are executed are deemed necessary for research, but seem counterproductive for the attacker herself. It can be bad for the future success rate and 'return-on-investment' per attack or even angry 'colleagues' — similar to magicians telling their tricks or security practitioners sharing effective countermeasures publicly. Therefore,

effective safeguards are required against de-identification.

Kevin Mitnick [MS02; MS05], a convicted Social Engineer and now a consultant, is well-known for his books on SE. During his prison time, the "Free Kevin" movement formed around the hacker group 2600 who released the documentary "Freedom Downtime"[1] including interviews with Kevin. Mitnick executed attacks — some only successful due to SE — that were hardly violating any criminal laws of that time. Interviews with former offenders that applied SE seem easier to execute compared to active criminals.

Persuasion principles applicable for SE are widely used in sales and marketing. Diana Li [AS15, p. 62] tried to talk directly to car salesmen about their selling tricks (cf. Section 4.3.2). There was a rumour that 50% of the profits were made by 10% of the customers. In her first attempt all but one salesman "clammed up" and refused. The second time she surrounded the important questions with placebo ones and got more profound insights. One can imagine that she deceived the interviewees by deviating from the important questions — a possible ethical challenge if briefings were omitted afterwards.

6.1.2 Other Interview Approaches

Besides interviewing experts in-person, via telephone, or in questionnaires, the SE **poetry slam** (Chapter 9) offers only limited interaction with the participants at the time of the presentation. Whereas in the **Lego modelling** (Chapter 8) the researchers can leave the observer role and interact with the participants.

The main idea behind asking for SE evidence is to 'let people tell their stories'. In a modelling session for an attack tree (see Kordy et al. [Kor+13] about the attack-defence tree tool), members of the TRE_SPASS project sat together and brainstormed which values per attribute domain to put on each node. The modelling scenario comprised the "IPTV case study" which described a future payment device combined with an IPTV set top box especially for persons with reduced mobility. The experts came from various disciplines and discussed the kind of attacks they could imagine. At the end a huge attack tree was created which partially contained SE attack vectors.

6.2 Literature

Literature of all types can be a valuable source of documented SE cases, such as **court documents** (Section 7). 'Classic' literature would cover *The Art of Deception: Controlling the Human Element of Security* by Mitnick and Simon [MS02]. Mouton et al. consulted sources such as "news articles, technical reports, research reports, films or blogs" [MLV16]. Mouton et al. experienced "that there are limited practical examples of social engineering in literature" [Mou+14a].

The literature on documented *real* cases of SE are one source besides fictional anecdotes. Scientific literature can offer more: thorough documentation of SE experiments. Experiments are designed to provide insights of SE in a controlled setting with properly documented procedures and outcomes. Experiments gain a dedicated category in Section 6.3.

[1] "Freedom Downtime" (2001), 2600 Films, directed by Emmanuel Goldstein

In the broader sense, the cybersecurity literature tries to explain or educate by example. Standards such as the IT-Grundschutz Catalogues [BSI13; BSI15] describe attacks by short stories that can become anecdotes, cf. Anecdote 1.3. That is, professional sources on SE may provide fictional, feasible stories or modified real cases to explain an attack and reasonable safeguards fitted to a specific topic. Security standards, policies, requirements or guidelines may become a source even when lacking exemplary stories. Attack vectors can be generated based on invalidating policies [Iva+15]. Policy invalidation can jump start creation of abuse and misuse cases involving SE. One specific case may expand into many SE anecdotes. However, attacks on such policies must be checked against the proposed SE indicators (Section 3.1; Research Question 1.2).

Furthermore, Computer Security Incident Response Teams (CSIRTs) and law enforcement agencies report on or publish warnings of criminal events. In 2017 the State Office of Criminal Investigation (LKA) of Lower Saxony warned about criminals impersonating police officers, see Anecdote 4.10 (warning poster in Figure C.6.5). Even online information on criminal activities may count as a source for SE, e.g., malicious job advertisements or job application e-mails with malicious payload [LKA18]. Investigative journalism and historic documentation of criminal activities can offer sources, too: the documentation of the 'Standgericht Herold' (Anecdote 4.12) was published by the German newspaper 'Welt' [Wös15]. Some events become belles-lettres if entertaining enough, such as the 'Captain of Köpenick' [Zuc67] of Anecdote 3.1.

6.3 Social Engineering Experiments

Controlled experiments applying SE or Randomised Controlled Trials (RCTs) as discussed in Chapter 5 play another role by creating *new* evidence that may lead to novel insights. In comparison to the analyses of literature like court documents, researchers create these new sources intentionally, similar to interviews. Because deception is one key enabler of SE (SE Indicator 3.5), researchers must include it in an experimental apparatus targeting persons. Researchers need to adapt SE indicators from Section 3.1 (Research Question 1.2) to their experimental design where the 'Targeted Person' becomes the test subject and 'Attacker' describes the researchers.

Targeted Person: Human Enabler (SE Indicator 3.1):
the experimental design must ascertain that SE can only succeed through the activity of the test subject.

Attacker: Intentional Communication (SE Indicator 3.2):
researchers communicate with the test subject directly or indirectly as well as initiate the communication.

Targeted Person: Unawareness (SE Indicator 3.3):
the experiment must be designed to keep the subject as unaware of the involved SE as possible.

Attacker: Malicious Intent with Goal (SE Indicator 3.4):
as there should ethically not be any malicious intent of the researcher to harm the test subject ("containment" [HS14]), the experimental design must present *fictitious* malicious intent.

Attacker: Deceptive Techniques (SE Indicator 3.5):
the experiment must express at least one deceptive technique towards the test subject, see also Finn and Jakobsson [FJ07] about the justification of deception for specific cases in Section 5.2.1.

The last two indicators are the most challenging regarding the outcome: the *malicious intent* becomes fictitious for SE Indicator 3.4. According to the SE definition (Definition 3.17), such experiments imitate SE, but are not exactly SE. Furthermore, deceptive techniques may function differently in the confines of a laboratory. SE experiments in laboratory environments can be challenging if the environment impedes finding insights applicable to real world situations (untenable experiments in Section 5.2.1), e.g., in the password meter experiment [Ege+13]. Hence, *external validity* [HS14] must show that the discovered nonbiased artefacts (such as the individual level of susceptibility to SE) are transferable to real life situations as well. The challenge and claim of generalisable results was discussed in 'Science of Security', Section 5.2.

Penetration Tests

Besides *controlled* experiments, corporate penetrations tests involving SE can examine personnel susceptible to SE. The results are rarely shared in the scientific community. Sometimes professional penetration testers tell their anecdotes pseudonomised, e.g., Anecdote 9.4 of the SE Poetry Slam or Long [Lon08, Chapter 5] on breaking into a building.

6.3.1 Challenge: The WEIRD Researcher Bias

For behavioural sciences Henrich et al. [HHN10] criticised research on the basis that its "standard subjects" not only represent a very small and specific subpopulation, but also that generalised conclusions are drawn. The subpopulation often used is described as coming from so-called 'WEIRD' societies. They represent more likely outliers compared to the rest of humanity. The abbreviation WEIRD stands for Western, Educated, Industrialised, Rich and Democratic. The majority of researchers has easy access to and therefore studies WEIRD subjects, e.g., experiments in academic environments with undergraduates solely. The earlier mentioned tipping behaviour of diners parties in an upscale restaurant in Philadelphia happened in a restaurant located on the campus of Temple University where University members like to eat [RB96]. The conclusions drawn must reflect such limitations. Henrich et al. [HHN10] present WEIRD samples from various domains like "visual perception, [...] cooperation, [...] reasoning styles, self-concepts and related motivations" [HHN10].

Some domains relate to SE research. Researchers must be aware of this bias when conducting experiments in particular. The motivation to conform is most likely weaker in

many WEIRD populations [HHN10].[2] Conformity is one of the persuasion principles developed by Cialdini [Cia07] (Section 4.5.1). Citizens of the USA are by far the most individualistic society in Hofstede's study [Hof14a], even compared to other Westerners [HHN10, Section 5.1]. Similarly, the freedom of choice, especially at work, is perceived differently in the world [Iye10]. It is important to consider these aspects that WEIRD samples are "highly unrepresentative" [HHN10] (see also claim in 'Science of Security', Section 5.2), e.g., when designing questionnaires and experiments or planning interviews. Regarding the omnipresent deception in SE research, Henrich et al. [HHN10] explicitly examined experimental practices which apply deceptive techniques. Hence, a researcher must be aware of the challenge to generalise results (external validity in experimental design [HS14], Section 6.3) and feel motivated to redo research with subjects other than in WEIRD societies.

[2] "A meta-analysis of studies performed in 17 societies (Bond & Smith, 1996), including subjects from Oceania, the Middle East, South America, Africa, South America, East Asia, Europe, and the U.S., found that motivations for conformity are weaker in Western societies than elsewhere." [HHN10, Section 3.2.3]

7. Court Documents as Evidence

In court the understanding of the term ‘evidence’ differs from its meaning in science, previously discussed in Chapter 5. In Germany with its continental European legal tradition, a court can decide which four sources of evidence to consider and how to rate the significance of each evidence. The four sources of evidence in the Code of Criminal Procedure (Strafprozessordnung) are statements of witnesses (‘Zeugenbeweis’), evidence of experts (‘Sachverständigenbeweis’), inspections (‘in Augenscheinnahme’), and reading out of documents (‘Urkundsbeweis’ [§249 StPO]). Contrarily, evidence in Anglo-American court processes is legally permissible if it presents the following four characteristics (CISSP exam guide [Har13, p. 1056]).

- *relevant*: “a reasonable and sensible relationship to the findings”
- *complete*: expressing the whole truth
- *sufficient / believable* to convince accepting the validity of the evidence
- *reliable / accurate*: factual and not circumstantial evidence consistent with the facts

This chapter still is about scientific evidence. The legal characteristics resemble scientific criteria. German court documents are created based on the pragmatic aspects of practices (useful in court) and not entirely based on scientific criteria [Döl84, p. 269]. The general file analysis (‘Aktenanalyse’) serves three main functions to examine court documents: *communication* (1) to inform other instances about decisions and *legitimation* (2) of decisions and their justification. The third function is to offer a *control* (3) mechanism to enable revised decisions in future processes [Döl84]. The last one resembles the scientific criteria to refute or validate previous scientific results. Furthermore, the truth finding process and the high expectations on the quality of evidence make court documents a valuable resource for Social Engineering (SE) research.

The judicial process asks the 'how' and 'why' questions about an event.[1] In comparison, forensics do not ask these questions, moreover focus on "reconstructing a historical explanation" [SMP17]. In court, the cases and court decisions are thoroughly documented. Furthermore, they can often be easily accessed by the public (second function of court documents: legitimation [Döl84] / 'in the name of the people') — unless there are reasons which create barriers, more in Section 7.1.2. Hence, court documents can be re-evaluated and the conclusions drawn can be validated or refuted by fellow researchers. An appeal on points of law can amend a previous court decision insofar until the final highest judicial authority is reached or a revision is prohibited by a court decision. That is, the truth finding process has a final amendment where it must end. In most jurisdictions the amendment is annotated and therefore visible in the preceding court document. If a case was closed, the scientific truth finding does not need to stop at that point.

Court documents offer historical data for research: they exclusively cover reported crimes that made it to court. During a hearing a historical event is examined and witnesses as well as suspects can be interviewed. The latter aspect of an 'interview' presents an element many other SE sources lack. Here, only reported crimes in documented *and* available court cases limit a generalisable, *quantitative* understanding of SE events. That is, conclusions are of a more qualitative than quantitative nature in both 'interview' sources. Accessible court documents of highest judicial decisions identify valuable jurisprudence for following court cases as guidance and for the public. Interviewing suspected criminals, similar to active and former social engineers in Section 6.1.1, can be bound to the applied method to achieve the desired saturation level. Saturation means reaching a predefined point where sufficient interviewees participated, see, e.g., snowball sampling in interviewee groups [HH18]. Rudimentary quantitative analysis in the entirety of available cases is advisable. For instance, Sillaber and Uebelacker [SU19] examined court documents qualitatively based on a Mayring 5 oriented content analysis [May14] and quantitatively according to Dölling [Döl84]. For SE research concerning court documents, it is more important to find stories expressing SE with novel attack strategies or ideas qualitatively. The quantitative aspect of how many times a specific anecdote was found, helps prioritising which anecdotes to choose in trainings etc. But novel SE strategies may likely outweigh this in favour of a general understanding of the SE landscape.

However, for this thesis the mere stories count of how SE was executed, who played which role (Section 7.2), and why it was successful by gaining what (Section 7.3). The modus operandi found in court documents can shed light on the "distinct methods of operation" [Har13] of attackers, thus revealing deceptive techniques (Section 4.4) and persuasion principles (Section 4.5) of SE cases. In some rare cases, access to court documents can be challenging (Section 7.1). Phishing is one type of SE and used here to narrow down the results exemplarily.

7.1 Obtaining Court Documents

The accessibility and completeness of court documents differ per country [SU19]. Some countries like Austria offer a centralised public database, others have decentralised sources

[1] CISSP [Har13] discusses the similar Motive, Opportunity, and Means (MOM).

like Germany where commercial portals offer central searches. One widely used German portal is the commercial juris database.[2] The juris database was consulted regarding (spear) phishing cases in the following. Phishing will be understood as in Definition 4.3 and its subgroup spear phishing as in Definition 4.4.

7.1.1 Phishing in the German juris Database

The juris database was searched for §263a StGB[3] for computer fraud and the term 'phishing' to identify phishing cases. §263a StGB is applicable when a phisher exploits the online credentials gathered for monetary gain, e.g., online banking fraud [Hoe15]. Not to miss phishing cases where 'phishing' was not mentioned, §263a StGB complemented the search. The following results originate from a pilot study by Pham [Pha15].

In July 2015, the juris database contained about 1.3 million court documents. 178 court documents were found based on these search terms without any restrictions on time or location. Out of those 119 were tagged with §263a StGB, 52 with the search term 'phishing', and seven contained both. 35 of the 178 documents were court decisions included to illustrate the application of laws, but not revealing any detailed information on the modus operandi of attacks. SE context[4] was found in 29 documents in total where seven were categorised under §263a StGB, 19 under the search term 'phishing', and three were tagged with both (ten times §263a StGB, 22 documents with 'phishing'). 1996 was the earliest year found for computer fraud offences involving SE [OLG D, 2 Ss 437/97], the last year in the data set was 2014 [AG KS, 2850 Js 26209/14]. The first court ruling on an offence of phishing was from 2005 [LG BN, 3 O 236/06], the last one was documented in 2013 [LG K, 3 O 390/13].

The amount of reported cases can be found in Germany in the Police Crime Statistics (PCS) and the federal situation report on cybercrime ("Bundeslagebild Cybercrime") of the Federal Criminal Police Office (BKA). Reported crimes express the bright field, hence, a subset of all incidents that had occurred. Only a few reported cases end in a court session and a subgroup of court sessions results in available court documents. To understand the huge discrepancy: in the years 2005–2015 more than 230,000 computer fraud incidents (key 5175(00) in PCS) were filed; for the same time frame a total of 119 documents were found in the juris database. From 2007 to 2014 the federal situation on cybercrime provided 35,138 reportings on phishing, but only 52 phishing cases were found in the juris database. One court document (juris database) can include multiple phishing reports (PCS). These numbers are not easily comparable without knowing how many reported cases all combined court documents contain.

During this [Pha15] and the second analysis [SU19] researchers faced the following challenges:

[2] https://www.juris.de/ run by the German company Juris GmbH, Saarbrücken.
[3] §263a StGB: general law covering computer fraud in the German penal code (Strafgesetzbuch)
[4] here: SE identified by principles of Cialdini [Cia07] and Stajano and Wilson [SW09]; the analysis was conducted before developing the SE indicators (Section 3.1).

7.1.2 Challenges with German Court Documents

Obtaining court documents and extracting useful information for SE research can be challenging and cumbersome. Below the challenges are categorised.

Completeness of search results:
Not all court decisions are documented and stored in the juris database. If a suspect does not appeal, a German court can write a shortened ruling resulting in lower granularity. Hence, the document will be rarely considered relevant enough for incorporation in a court document collection. One may say that only higher court decisions are seminal for subsequent cases.

Completeness of each document:
In court the main actors in SE cases are defendants, plaintiffs, and witnesses. Their role in SE and in particular in phishing differs. A plaintiff can be a victim of phishing, a bank, a money mule[5] or even the state; a defendant can be a bank, a money mule etc. That is, at least one of the necessary parties involved in SE can be interviewed in court. Predominant roles in phishing cases are discussed later in Section 7.2.

How detailed a document becomes, relies on the evidence. It may leave research questions unanswered, e.g., if defendants do not want to incriminate themselves. Furthermore, not all evidence is provided in court documents. Court documents do not grant unlimited access to all records, e.g., the bill of indictment with more evidence is missing. Regarding data quality in the empirical approach, data are complete when "of sufficient depth, breadth, and scope for the task at hand" [BS06]. For in-depth SE research these data may reveal insufficient detail.

Findability
using the juris search engine means how easy it is to receive the desired information. The juris database contains amended court documents, for instance, by added relevant articles as search parameters. This enabled the search for §263a StGB (computer fraud) and to identify phishing cases that were not indexed with the search term 'phishing'. Misspelling in original court documents occurs and needs to be taken into account. The final search terms were extended with additional typo phrases, e.g., 'pishing' [OLG HAM, 31 U 31/15].

Privacy
in court documents is a valid aspect, especially if the court found a suspect not guilty. Pseudonymisation of names of suspects is a minor inconvenience for research. A bigger barrier is the restricted access to all case records. The former mainly

[5] In one case in the juris database, the money mule sued his bank because a transaction was revoked ("Stornobuchung"). The case was dismissed. The money mule received 7,000 EUR from a bank account that was phished. The money mule withdrew 6,000 EUR (less a 1,000 EUR commission) and sent it via Western Union to St. Petersburg. The phished targeted person (witness) contacted the bank about this issue and the money was transferred back into the targeted person's account. Then the money muled sued the bank [LG BN, 3 O 236/06].

hinders identification of the actors' genders, the latter the complete understanding of a specific case.

Accessibility at no charge
of court documents is vital when courts rule 'in the name of the people' in a free democratic basic order. The research in this chapter was conducted using the commercial juris database in a university library which paid the license fee. No alternative *central* database free of charge was known at that time. Some German states have their own (decentral) databases, e.g., North Rhine-Westfalia[6], Berlin-Brandenburg[7] (provided by juris GmbH) or Lower Saxony[8]. An overview of all federal and state databases can be found on justiz.de.[9] The Hamburg based non-profit association openJur e.V. also provides court documents pseudonomised and centralised in one database at no charge.[10] Related is the previous definition of accessibility where available data must be easily and quickly retrievable [BS06]; Whether court documents are *easily* accessible is to question. This was especially the case described in Anecdote 7.1.

> **Anecdote 7.1 — Obstacle to acquire court documents (LG Osnabrück).**
> A phishing case in the news [EH16] got my attention. It resulted in a sentence of about 6.5 years for a group of phishers. The phishers used malware on victims' computers to intercept online banking activities. It was not described how the machines were infected, i.e., whether SE was involved. The attackers acquired replacement SIM cards from the victims' telecommunication providers. These SIM cards were used to receive a copy of the mobile TANs for bank transactions. It was not clear how the phishers were able to fool the telecommunication providers.
> A search for the court documents nine month after the trial was unsuccessful. Neither the juris database nor the online portal for court documents of Lower Saxony[8] showed any results. An enquiry form on the court's own website returned an error[11] after submission. An S/MIME signed e-mail was sent to the post room address of the court (lgos-poststelle@justiz.niedersachsen.de, 2017-05-04). The e-mail server `izn1.Niedersachsen.de` replied with the error "`JUSTIZMAILGW01 #550 5.4.6 Mail loop detected`". As the last option I tried to contact the court via classic mail to their P.O. box. I did not receive any response.

Shareability:
A publicly available document can be shared unless licensed otherwise. The documents downloaded from the juris database cannot be shared. Juris GmbH denied a

[6] https://www.justiz.nrw.de/BS/nrwe2/
[7] http://www.gerichtsentscheidungen.berlin-brandenburg.de/ (no TLS encryption)
[8] http://www.rechtsprechung.niedersachsen.de/ (no TLS encryption)
[9] https://justiz.de/onlinedienste/rechtsprechung/index.php
[10] https://openjur.de/
[11] "Zurückgewiesene Anfrage: Ihre Anfrage konnte leider nicht bearbeitet werden. Wir bitten dies zu entschuldigen. Die ID ihrer Anfrage lautete WQsb@goRFUoAAUYB8[...]"

The file `courtdocumentsphishingworkflow.pdf` hasn't been created from `courtdocumentsphishi`
Run '`dot -Tpdf -o courtdocumentsphishingworkflow.pdf courtdocumentsphishingworkflo`
Or invoke LaTeX with the `-shell-escape` option to have this done automatically.

Figure 7.1: Predominant workflow of phishing cases in court documents (adjusted from Sillaber and Uebelacker [SU19])

request to share and publish their redacted and augmented versions of court documents. Fellow researchers who want to repeat the analyses must acquire a license to access the data *sources*. This shareability of the original documents should not be confused with shareable knowledge [SI18]; the research results are shareable and were shared though.

Readability:
Although human readable, the database search results are not easily machine readable and thereby processable for automatic analyses. The PDF or HTML formatted documents follow no schema for validation such as the JSON Schema.[12] Automatic analyses depend heavily on the formatting. Decentrally published documents lack a standardised (machine readable) publication format combined with a descriptive language for validation.

The first four points impact the level of detail to find novel SE attacks as well as to better understand the modus operandi qualitatively. As mentioned earlier, insights about novel attack strategies may outweigh quantitative analyses. Therefore, these first four challenges may impact a profound understanding of SE based on court documents.

7.2 Roles in Phishing Cases in Court

The parties in a court setting are different from those in a pishing attack. Defendants, plaintiffs, and witnesses may play a role in a SE situation. In court not all of the phishing roles are present, e.g., the most frequently found plaintiff-defendant relationship was between phishing victims and banks. When the majority of phishing cases are combined, a predominant phishing workflow arises, depicted in Figure 7.1. Anecdote 7.3 [VG B, 10 K 333.10] describes an outlier of the typical phishing workflow involving an emission certificate trade portal instead of banks.

The 'victim' is a *targeted person* (Definition 1.3) who got phished by a phisher ('*attacker*', Definition 1.2) successfully (Section 1.6 for naming convention) which is the prevalent reason for a court process. The communication and interaction between phisher and victim is the important part to understand why a possible act of SE succeeded. In a court case these two actors can also comprise multiple persons with the same role, e.g., a group of phishers or multiple victims. However, a money mule can also be a 'victim' of a phisher,

[12] https://json-schema.org/

more in Section 7.2.1. The identified SE roles (phisher, victim, money mule) provide a framework for the creation or complementing of SE anecdotes. For each attack targeting phishing victim or money mule the SE indicators can be used to classify SE (Section 3.1). Also, the Persuasion Knowledge Model (Figure 4.1) of Friestad and Wright [FW94] can be consulted (Section 4.1).

7.2.1 Becoming a Money Mule

A money mule is a financial agent who launders money illegally obtained by a criminal network (phisher). They make it difficult for law enforcement to trace money flows, e.g., by registering businesses or using bank accounts under their names hiding the phishers' identity [LJ16]. Money mules in court documents often claimed not to have known their criminal role as a money laundering agent (Definition 3.4). It is not possible to ascertain if money mules did not want to incriminate themselves or were indeed unaware of any wrongdoing. In the latter case the money mule was successfully lured into that financial operation by probably using SE. Thus, financial agents as victims of SE can exist as targeted persons in SE anecdotes.

In a court case [VG M, M 13 DK 12.3091] and the appeal [Bay. VGH, 16a D 12.2519] a Bavarian police officer acted as a money mule (Anecdote 7.2). The court of appeal's judgement dealt with the disciplinary procedure (money mule vs. state), but also provided insights about the case. Neither the money mule nor the customers were phished here, the document was found by searching for the German computer fraud paragraph. SE at the right time let the presumably gullible police officer comply.

> **Anecdote 7.2 — Police Officer as Money Mule.** A police officer was recruited as a financial agent for a fake online shopping company. His bank account was used to receive bank transactions from online shopping customers who did not receive their products. He subtracted his commission and sent the remaining money via MoneyGram to a person in the Philippines. He became a money mule for the fraudster and claimed not to be aware of being used. The court ruled that the police officer acted in neglect or carelessness as well as naïvety to not suspect any fraud. The money mule had financial difficulties at that moment. According to the documents, these were the enablers making the police officer become a scam victim.
>
> [Bay. VGH, 16a D 12.2519; VG M, M 13 DK 12.3091]

The recruiting of a money mule can start with a lucrative job offer via e-mail. Figure 7.2 does not originate from any court document, but just from the author's inbox[13] as an example. Such e-mails can be useful to imagine missing parts of any SE anecdote. The sending IP address (if correct in the header) came from a dynamic address pool of a German Internet Service Provider (ISP) (Versatel). The e-mail was sent from a most likely hacked e-mail account of an e-mail server located in Switzerland (cyon.net) handling the e-mails of their clients; the `From:` field matched the e-mail server domain. A `Reply-To:` field was added to receive responses to the free mailer address at gmx.com. The same e-mail address was used in the e-mail text. This is a widely used practice and it may be

[13] The author was able to analyse the e-mail header which would have been stripped in a court document.

Subject: Agentur für Arbeit Onlineangebote für den Februar 2016
From: "Lastname" <firstname@example.com>
Date: 13/04/18 12:28
To: "Sven Uebelacker" <thesis-spam@uebelacker.net>

Guten Tag Sven Uebelacker,

unsere Gesellschaft ist ein führendes Finanzunternehmen und wir suchen ab sofort zuverlässige Mitarbeiter zur Vervollständigung unseres Teams in der EU.
Das Gehalt beträgt ca 3900 Euro monatlich bei ca 4 Arbeitssunden pro Woche.
Der Arbeitnehmer hat keine eigenen Ausgaben und muss keine spezielle Kenntnisse haben.
Auch Berufstätige sind für diese Arbeitstätigkeit geeignet.
Ihr Tätigkeitsfeld ist die Geldflussoptimierung. Sie erhalten die Mittel dierekt auf Ihr Bankkonto überwiesen, und müssen es abzüglich Ihrer Provision von 20% weiterleiten.
Die Arbeit ist in ganz Europa angeordnet und derzeit noch zu besetzen.
Sie sollten Zielstrebigkeit zu Ihren Stärken zählen und grundlegende PC Kenntnisse besitzen. Problemloser Umgang mit PC sowie telefonische Erreichbarkeit sollten für Sie auch kein Problem sein.
Kontaktaufnahme mit uns:

Wenn Sie sich angesprochen fühlen, wollen wir Sie kennenlernen, hierzu mailen Sie uns Ihre Bewerbungsunterlagen an: firstname2.lastname2@gmx.com

Ihre privaten Unterlagen behandeln wir natürlich vertraulich.
Mit freundlichen Grüßen

Lastname

Figure 7.2: Job offer e-mail promises a high income if the author becomes a money mule (‘`From:`’ field and free mailer address anonymised; recipient address changed)

assumed that the free mailer address can be accessed by a criminal network: they created this free mailer account or they hacked an existing one that showed a good reputation to avoid suspicion by the free mail provider.

The e-mail claimed to come from a leading financial institution which wanted to complete their team in the EU. The job offered a position in money flow optimisation with just a few requirements such as rudimentary computer skills and a telephone. The applicants should receive money on their bank account. The applicants would keep 20% commission before transferring it somehow (not mentioned). The expected monthly income would be around 3,900 EUR for working four hours per week with no expenses. The e-mail subject resembled a monthly job offer mailing coming from the Federal Employment Agency (Agentur für Arbeit). Hence, the e-mail targets primarily job seekers who are probably unemployed. The salary is incredibly high and motivating. This makes it hard to determine if the applicant would apply because they are gullible or they do not care to become involved in criminal activities when in need of money easily earned.

7.3 Predominant Goal: Financial Gain

All analysed court documents of phishing cases had as predominant goal direct or indirect financially-motivated attacks. They are therefore categorised as Financially-Motivated Social Engineering (FMSE) according to Verizon RISK Team [Ver19]. The major “criminal business model” [SI18] was to obtain money via bank transactions and money transfers *directly*. One special case of spear phishing for carbon emission certificates was identified, see Anecdote 7.3 [VG B, 10 K 333.10]. Carbon emission certificates can be traded similar

to a currency; they have monetary value. They can be seen as *indirect* financial goals. If the attacker's motivation was to harm the company without using the value of the certificates, the certificates represented a financial loss for the company. The attacker disrupted the business process, e.g., because of the company's need to acquire new emission certificates. It is more likely to assume a financial interest as six additional German accounts of other companies were spear phished successfully. Interestingly, the sink of the certificate transactions could not be fully identified. After the certificates reached the London based company 'Total Global Steel' the trail to the money sink was lost, although the International Transaction Log (ITL) should have transaction information. It is not clear why this was not possible. The international transaction system seems to be implemented insecurely.

Anecdote 7.3 describes a phishing attack where SE was identified. The attacker acted with malicious intent (money, business disruption) and had a goal (SE Indicator 3.4). She communicated intentionally (SE Indicator 3.2) to urge targeted persons to install important security updates to protect the trading portal. The unaware targeted person (SE Indicator 3.3) was deceived to log in to a phishing website (SE Indicator 3.5) and enabled the attack successfully (SE Indicator 3.1).

> **Anecdote 7.3 — Spear Phishing for Carbon Emission Certificates.** An employee of an organisation in the cellulose industry received a personalised e-mail asking to install security updates. The updates should ensure the security of the emission certificate trade portal. Prior to installation the phishing victim should check their account details on a website to get the computer registered and linked with the service. A link to a phishing website was provided. The phishing victim entered the account credentials of his organisation. Unknown phishers transferred 76 minutes later two types of emission certificates to two Danish emission certificate accounts. The attacker transferred these certificates later to an organisation in London and then to unknown destinations. The German emission certificate registry (Deutsche Emissionshandelsstelle (DEHSt)) is connected directly to the United Nations Framework Convention on Climate Change's (UNFCCC) ITL. They noticed six additional successfully phished accounts in their national registry due to spear phishing e-mails. [VG B, 10 K 333.10]

7.4 Usefulness of Court Documents

In comparison to other SE sources, court documents reveal their truth finding process. Court documents are based thoroughly on the high quality standards for court evidence. They cover almost exclusively the bright field of SE, representing a truth finding abstraction of reality and not a fictional SE story. Due to the time between a SE event, the conviction and finally the documented and published court cases, the documents cannot cover SE incorporating more recent technological advances. The small number of cases found with phishing shows *relevant* cases for the court process. These cases provide sufficient information to become judged at court. Hence, sufficient evidence exists to start a trial. This comes in handy that court documents can provide reliable SE evidence sufficiently complete.

Evidence in court differs from SE evidence needed for research. Court documents must be publicly accessible to foster a scientific process based on the original sources, see Open

Data initiatives including the “FAIR Guiding Principles” for scientific data management and stewardship [Wil+16][14]. However, not all records are available without limitation like the bill of indictment containing more evidential material. The term ‘anecdote’ was used here as well to remain consistent for all SE sources, although in jurisdiction a different terminology depending on the evidence type is applied. Court documents offer a rich, but limited source of high quality SE evidence. Novel insights for SE research can be found, e.g., the spear phishing for acquiring carbon emission certificates (Anecdote 7.3) or the elaboration on money mules whether accomplice or not (Section 7.2.1).

[14] https://www.go-fair.org/fair-principles/

8. Lego Modelling

One approach in the EU FP7 project TRE$_S$PASS[1] comprised modelling Socio-Technical Systems with Lego bricks [D4.1.2; D4.2.2; D4.3.3]. When discovering anecdotes of Social Engineering (SE), Lego offers a unique bottom-up approach for modelling and brainstorming. Participants of different disciplines can come together, communicate virtual concepts, and highlight gaps in their understanding. With different roles in modelling sessions, the participants not only form an interdisciplinary, but also a multifunctional team. The ease to understand how to use Lego bricks and the haptic experience as well as the mostly positive memory of 'playing' with Lego from childhood fosters a fruitful discussion. Lego bricks may spice up the creativity and maybe make use of the playfulness of an inner child. Furthermore, different perspectives on a scenario modelled with Lego is a key feature (3D). A predefined colour-coding can help to convert a physical model into modelling languages like Unified Modelling Language (UML) or into modelling tools like ArchiMate [HCH14]. At that point specific knowledge on how to model with the ArchiMate tool is required.

This modelling approach was applied for creating SE attacks in a brain storming session prior to the creation of our SE indicators. This chapter describes the modelling session, its outcome, and the lessons learned.

8.1 Lego Modelling of a Socio-Technical System

The scenario originated from a Lego modelling session at the TRE$_S$PASS winter school in January 2016 at the Technical University of Denmark (DTU), Copenhagen, executed by Lizzie Coles-Kemp (Royal Holloway University London (RHUL)). It aims to model

[1] Technology-supported Risk Estimation by Predictive Assessment of Socio-technical Security

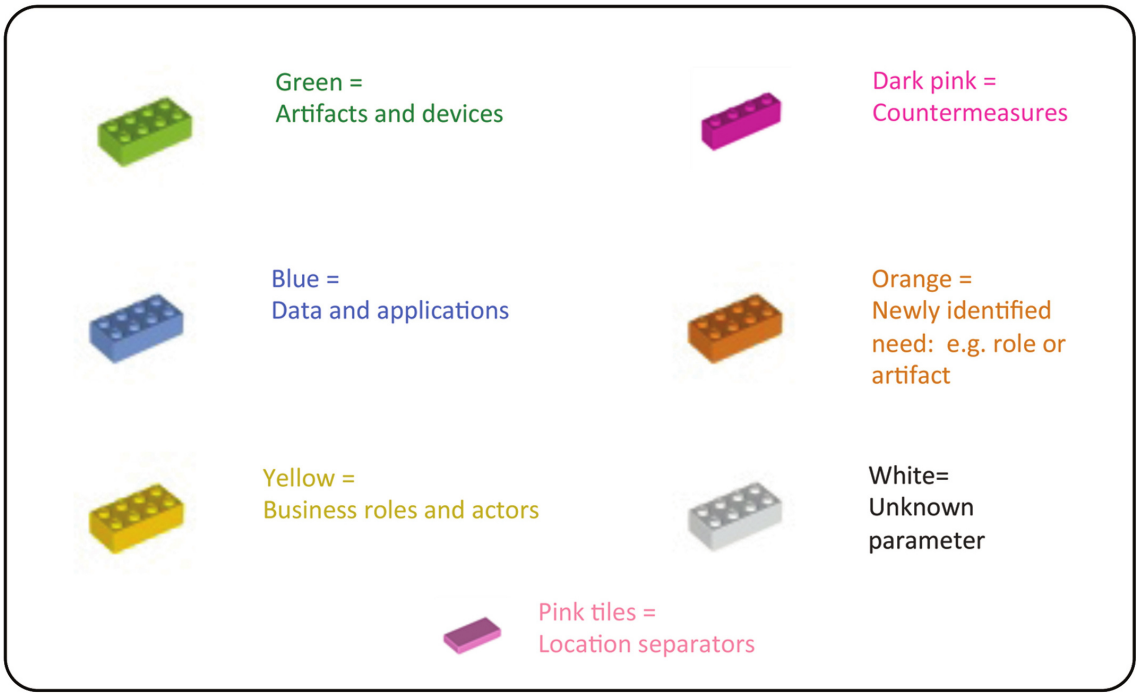


Figure 8.1: Exemplary colour key definition for scenario artefacts: "Modelling the model; brainstorming with Lego; physical, logical, and human variables" used in the TRE$_S$PASS winter school

attack vectors of a Socio-Technical Systems in cloud computing. A similar, less complex scenario was examined by TRE$_S$PASS partners in the paper "Tool-Based Risk Assessment of Cloud Infrastructures as Socio-Technical Systems" by Nidd et al. [Nid+15] for formal risk assessment (but not for Lego modelling).

8.1.1 Cloud Computing Scenario

The Socio-Technical Systems consist of various roles and personas, digital infrastructure as well as the physical floor plan. The physical space comprises two rooms, a hallway, an internal room and a data centre, including doors and access control to all three spaces as depicted in Figure 8.2. The technical infrastructure comprises physical servers 1–2, virtual machines 1–4, and virtual firewalls 1–2. The hypervisors are connected to a physical switch (SW1). The scenario refers to a private investment and trading enterprise migrating to a cloud based service. The goal of the attack was given and consisted of stealing a confidential document stored on one virtual machine inside the computer centre.

8.2 Modelling Session

The following modelling session was part of a seminar at Hamburg University of Technology (TUHH), a course where students have to do independent investigations on a given topic. The seminar "Human Factors in Information Security", organised by the author, addresses bachelor and master students of all disciplines. Participants have to evaluate recent scientific advances on human factors research. One aspect in the summer semester

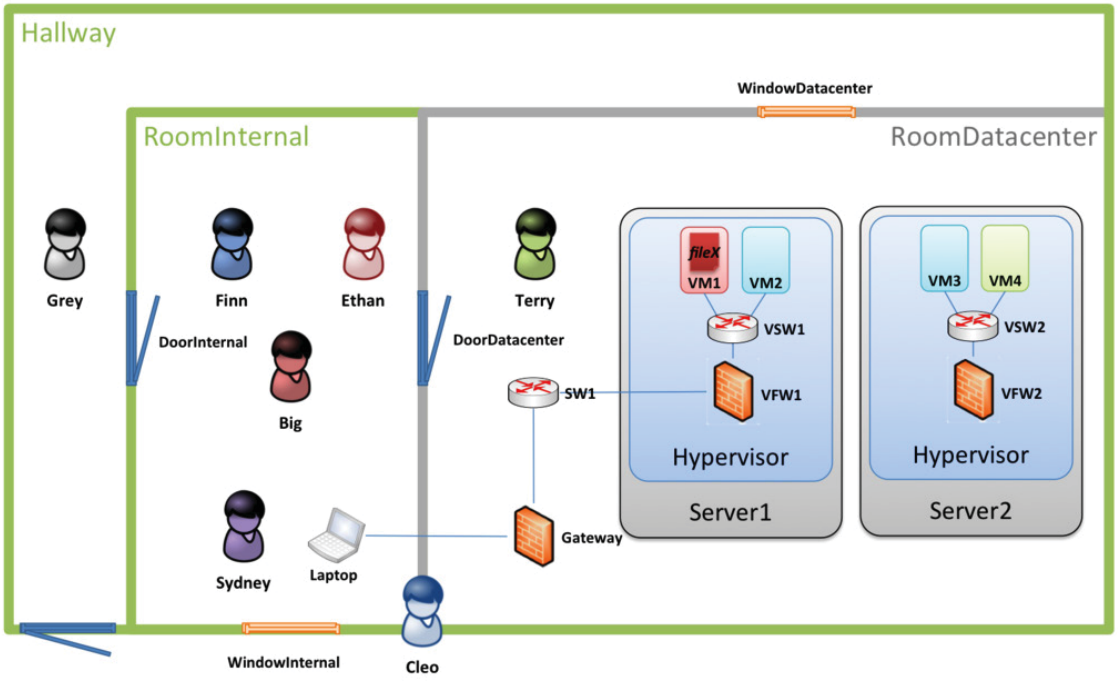


Figure 8.2: One map of TRE$_S$PASS' cloud computing scenario including socio-technical components (persons, physical barriers, and digital infrastructure) used in TRE$_S$PASS winter school

2016 was the task[2] to develop an attack story based on the TRE$_S$PASS cloud computing scenario. Eleven TUHH students participated in this session on May 27th 2016. Students formed two groups and discussed their attack ideas based on the input learned in previous sessions to create a rich picture of an attack. The attacks developed had to include social interaction using human factors as an enabler. These restrictions seem not too strict for a Socio-Technical Systems, however, including a human factor enabling the attack (SE Indicator 3.1) and requiring social interaction (SE Indicator 3.2) can hint to a SE attack. The malicious goal of an attacker was preset in the scenario by accessing a confidential document (SE Indicator 3.4).

The students were given role definitions of the actors and corresponding access privileges in the physical and digital realm (Figure 8.3). Furthermore, a recommendation on how to use different colour schemes was provided by exemplary colour key definitions of artefacts, actors, etc. (Figure 8.1). The modelling session took around 60 minutes including a presentation of the modelled attack ideas.

8.2.1 Student Group 1

The first group presented an attack described as Anecdote 8.1 and depicted in Figure 8.2.G1.1.

[2] I am grateful that I could use the Lego brick sets from robotik@TUHH (Sarah Latus). Their classes address pupils to spark interest in robotics and inform about future engineering careers (https://dual.tuhh.de/).

Actors	Departments & Roles	Owns	Physical access	Digital access
Ethan	Organised Crime	FileX	Room Internal	VM1
Finn	Finance Officer	-	Room Internal	VM2
Terry	IT support Technician	-	All	None
Sydney	System Administrator	-	All	All
Cleo	Cleaning personnel	-	All	None
Attacker	-	-	None	None

Figure 8.3: Role definition and access privileges of actors used in the TRE$_S$PASS winter school

> **Anecdote 8.1 — Attack Outline Lego Group 1.** The confidential file was copied by a robbery at gun point coercing an employee to provide the confidential document. All witnesses (big boss on the left; employee at computer) were shot at the end to prevent the identification of the attacker (red cap; gun in hand). The detectability of such an executed attack is extremely high. However, there can be challenges in the post mortem analysis to know what kind of data was on interest and what was copied (not "stolen").

The enabling human factors (SE Indicator 3.1) are predominantly fear experienced by the targeted persons and their hope for not being shot if complying with the attacker's request. Social interaction (SE Indicator 3.2) leads to a successful attack. The attacker has malicious intents (SE Indicator 3.4). As there were no manipulation by (psychological) triggers involved (~~SE Indicator 3.5~~) and all parties were well aware of the attack (~~SE Indicator 3.3~~), this story cannot be identified as SE. The created anecdote showed that the modelling process would have required more supervision and intervention based on the SE indicators.

8.2.2 Student Group 2

Group 2 came up with a more sophisticated attack to gain access to the file as follows in Anecdote 8.2.

> **Anecdote 8.2 — Attack Outline Lego Group 2.** The attacker identified the cleaning personnel (Cleo) as a weak link as Cleo (red shirt, long neck) has to take care of her handicapped son (green bricks) and needs financial support. She was contacted outside the company's premises by the attacker (white body, silver head) who gave her an USB stick containing malware (Figure 8.2.G2.1). Cleo's goal was to use the computer of the systems administrator Sydney once he left the company (Sydney at work in the background of Figure 8.2.G2.2). Cleo entered the office with the USB stick as shown as a red, transparent brick in Figure 8.2.G2.3. Once Sydney has left, Cleo inserted the USB stick to the computer (Figure 8.2.G2.4). The malware created a backdoor for the attacker to copy the confidential file (black line to the outside world, Figure 8.2.G2.5).

The targeted person is the cleaning personnel in who colleagues of the organisation established trust. To conclude on this attack story, a (direct) social interaction between cleaning personnel and attacker (SE Indicator 3.2) is clearly enabling the attack (SE Indicator 3.1). The attacker initiated his action with malicious intent (SE Indicator 3.4). Although this comes close to the SE definition (Definition 3.17), the targeted person is *knowingly* complying with the attacker's request by accepting the bribe (~~SE Indicator 3.3~~). Like blackmailing or coercion, bribery was excluded as it cannot be categorised under deceptive techniques (~~SE Indicator 3.5~~). The detectability is expected to be lower than in Anecdote 8.1. Group 2 created another persona: the handicapped son of Cleo. This is acceptable here because the main personas were not changed; this additional persona merely helped to understand Cleo's motivation. Overall, this anecdote is more sophisticated than the previous one, but does not count as SE according to the SE indicators.

Cleaning Personnel as Favoured Suspect

Cleaning personnel is often considered in discussions about persons susceptible to SE. A similar anecdote was presented in the SE Poetry Slam (Anecdote 9.5) where penetration testers impersonated cleaning personnel. It can be assumed this happens for various reasons:

- Cleaning personnel is often employed by an external company. The identification with the employees' organisation can be questioned.
- Janitors are rarely seen during working hours (social detectability) and are easy to blame indirectly (and they are rarely part of any team other employees cooperate with directly).
- Cleaning personnel receives wide access privileges and can access more rooms than the average employee in general. Some employees may argue that they themselves are therefore less likely a suspect.
- With respect to the wider access privileges for cleaning personnel, employees may envy them and may (unconsciously) suspect them
- Susceptibility to bribery or malicious intentions with financial gain is expected to correlate with lower wage.
- A cleaning company may employ temporary personnel via a temporary agency. This can lead to alternating cleaning personnel with whom employees may hardly familiarise.

Regarding the latter point, in 1994/1995 "The Hacker Quarterly" magazine of the hacker group called '2600' published an article about "Janitor Privileges" [Voy94]. The author described an attack vector on how to get into an organisation by applying as a temporary worker at a janitorial company for night shifts. Anecdote 8.3 outlines this article. In the whole process the attacker may impersonate someone or use deception. The article does not focus on this and the author does not mention SE at all.

But the question here remains — with respect to the above mentioned aspects of employees' *feeling* of insecurity — whether cleaning personnel is an overestimated threat in reality.

Anecdote 8.3 — Janitor Privileges (Voy94). To enter the premises of an organisation with malicious intent, an attacker gathers information about the outsourced cleaning company. First, to discover which janitorial company is present, the attacker may call the organisation itself and asks if they are happy with the current cleaning company and would recommend it. An observation of the building at cleaning time may also be feasible to identify logos of the janitorial company.

The attacker calls the cleaning company and asks for a good temporary agency who would hire temporary personnel for them. The temporary agency can have fewer security checks and the attacker could apply as student for night shifts. By stating which area the attacker can work, the attacker can narrow down possible other organisations. The attacker may reject some offers until the desired organisation comes up.

Once inside the organisation, the attacker needs to gather information about security guards, supervisor behaviour, security cameras etc. The article continues on how to carefully "steal" documents or other information, e.g., using a small, inconspicuous notepad.

8.2.3 Outcome & Lessons Learned

To conclude, both modelled attack stories do not convey SE according to the SE indicators. It is no surprise that SE Indicator 3.1, SE Indicator 3.2, and SE Indicator 3.4 were included correctly in compliance with this session's prerequisites. That the other two SE indicators were not present, is a lesson learned to provide these additional requirements in future sessions. At the time of this session these two requirements had not been postulated yet. The modelling session served also the purpose to support the finding process for the SE indicators. This modelling method with Lego remains still a useful approach when conducted with more guidance and intervention concerning all of our SE indicators.

Furthermore, the goal to teach students about Lego modelling carried away the advantage to gain insights from cross-functional groups of various professions with different technical expertise. The majority of subjects, though, were tech-savvy students and a Western, Educated, Industrialised, Rich and Democratic (WEIRD) bias can also be assumed.

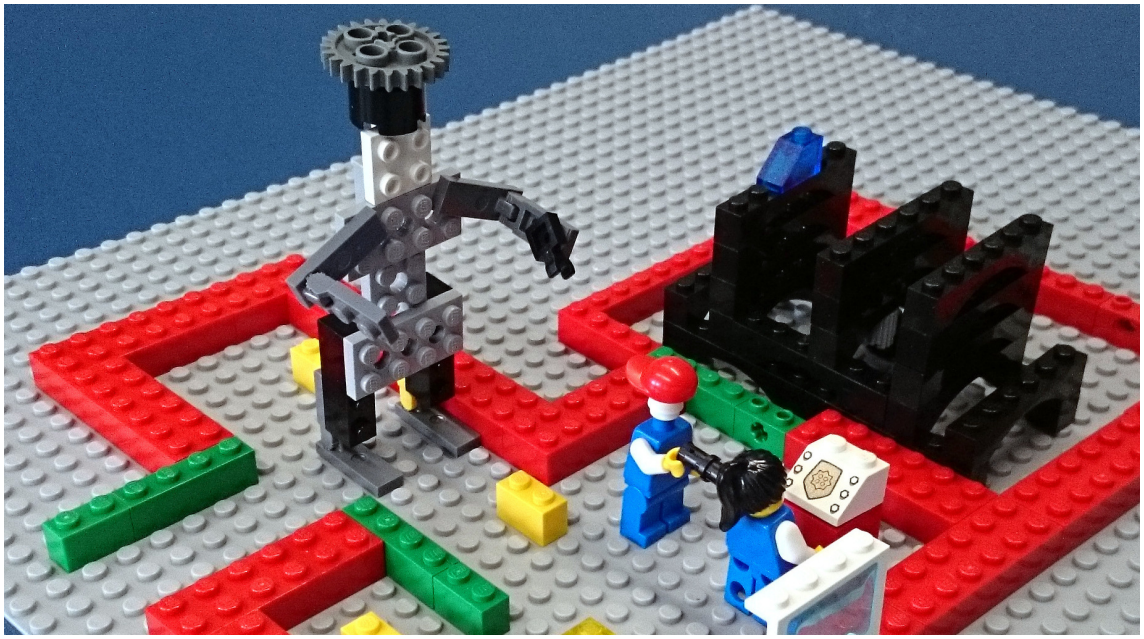


Figure 8.2.G1.1: Holdup murder to gain access to a confidential file; attacker (red cap) points gun at an employee accessing the file on a computer

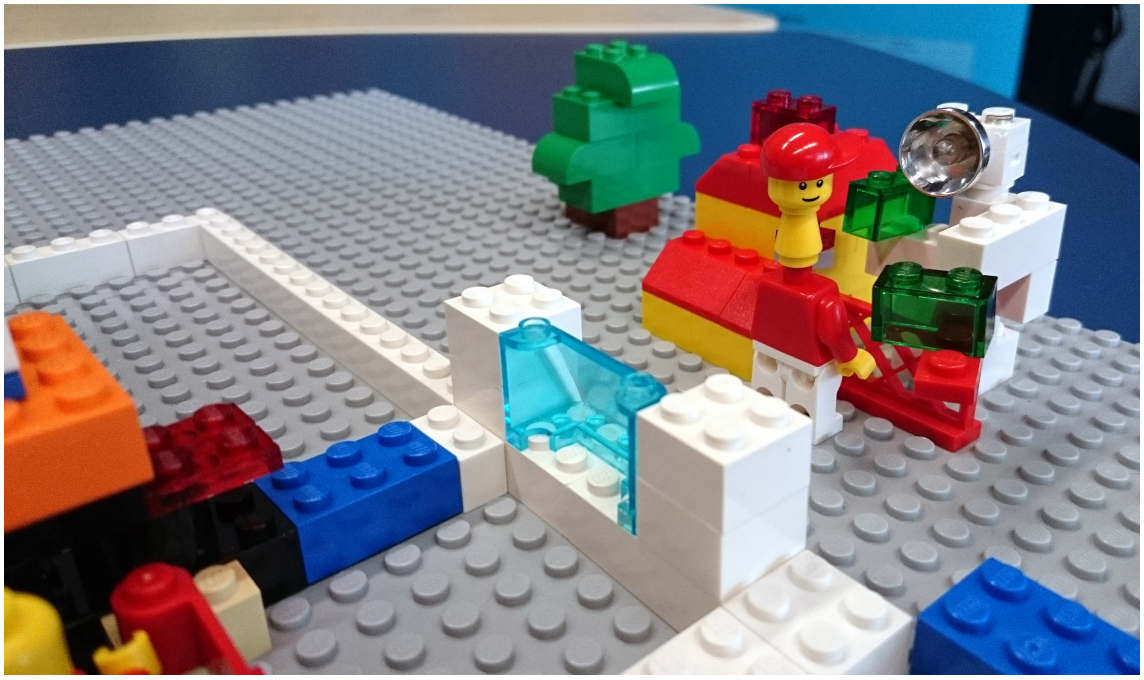


Figure 8.2.G2.1: Perspective from inside to outside the premises where the attacker (white body, silver head) hands Cleo (red shirt, long neck) an USB stick containing malware. Cleo's son stands in the background (green bricks).

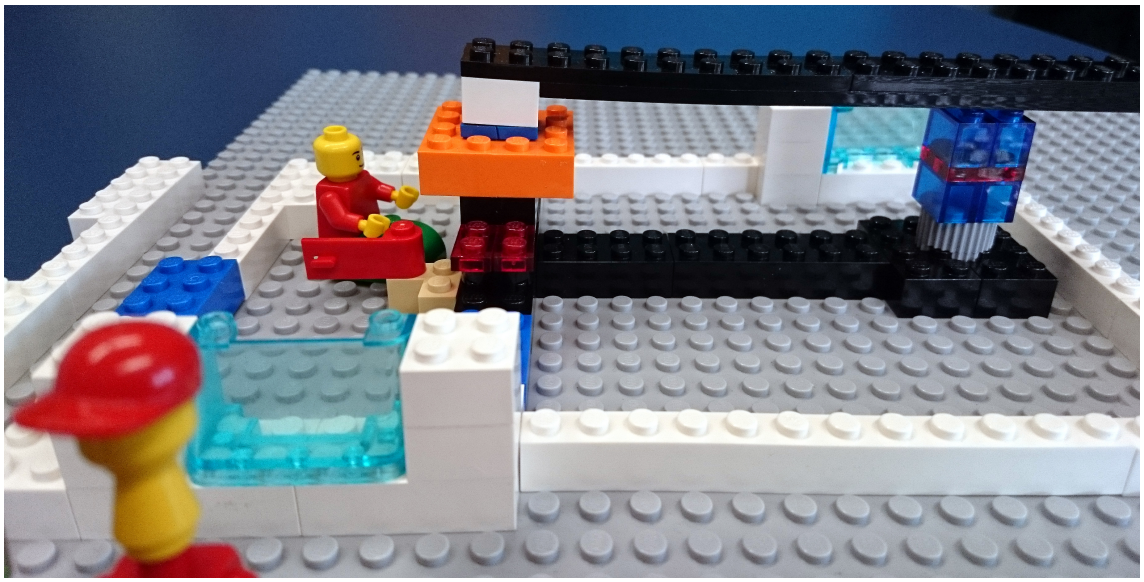


Figure 8.2.G2.2: Seeing Sydney (systems administrator) from the outside while he is working; in the front Cleo's head

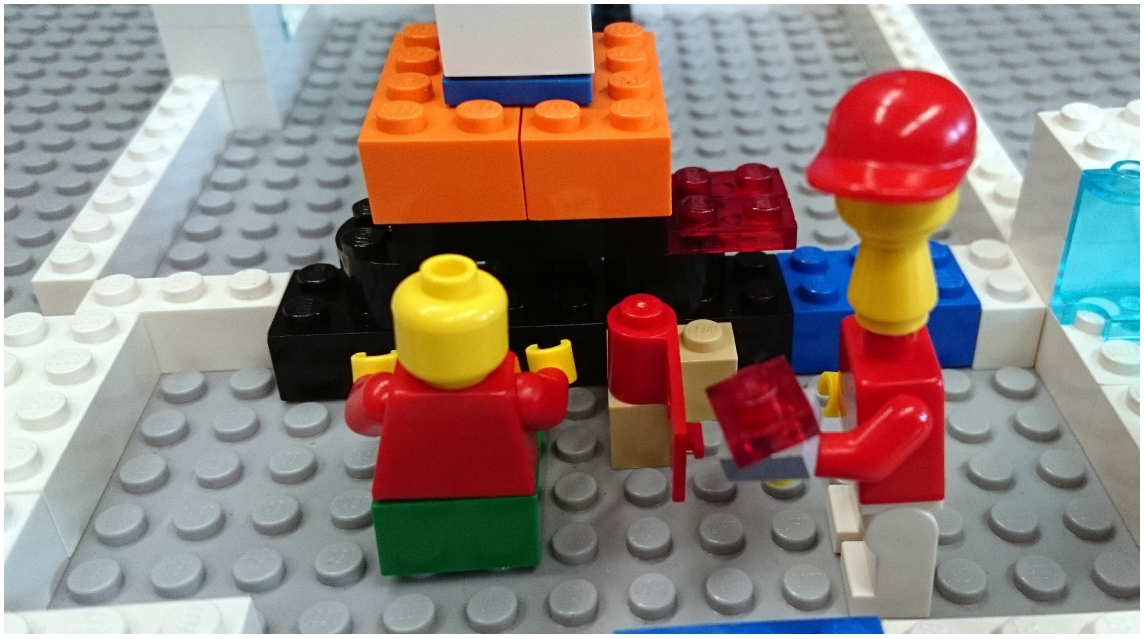


Figure 8.2.G2.3: Cleo enters the office with the malicious USB stick (red, transparent brick) awaiting Sydney to leave

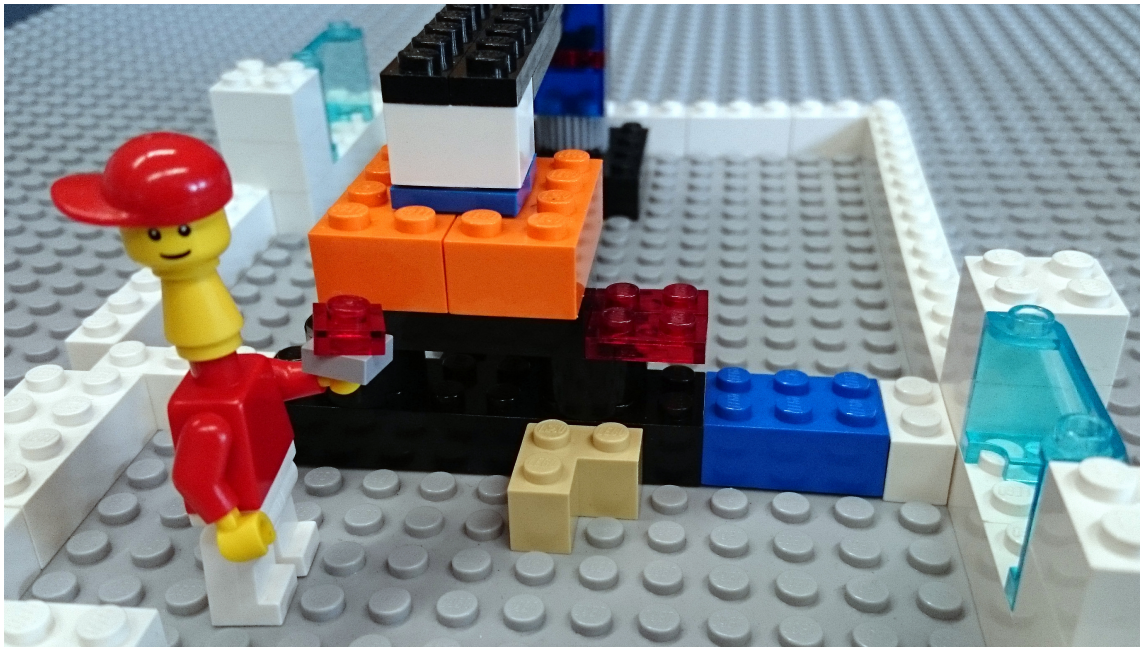


Figure 8.2.G2.4: Sydney left and Cleo inserts the USB stick into the computer

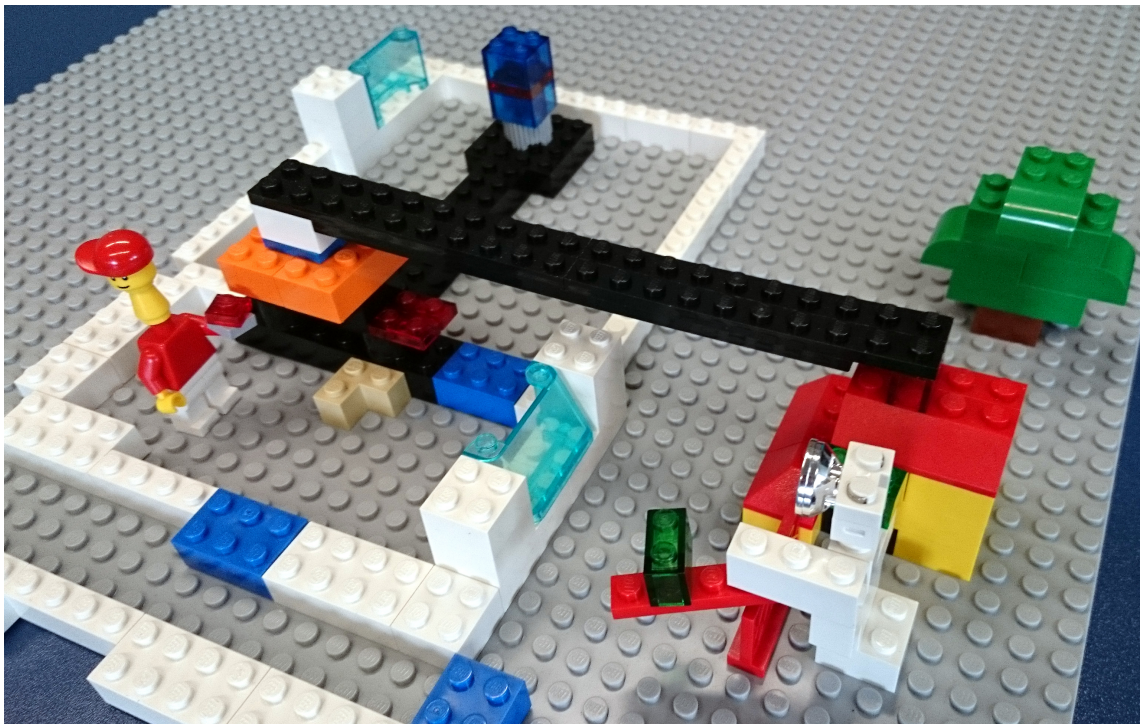


Figure 8.2.G2.5: The malware creates a backdoor on the server (network connection depicted as a long black brick to the outside) for the attacker to obtain the file

9. Social Engineering Poetry Slam

The 'Social Engineering Poetry Slam' is a format created by the author as a prototype for the 31st Chaos Communication Congress (31C3) in 2014. After the initial presentation of that idea, a group was formed to enhance the poetry slam concept. At the 33rd Chaos Communication Congress (2016) the poetry slam was held as a self-organised session [UR16]. This chapter describes the conceptual ideas behind the Social Engineering (SE) Poetry Slam in Section 9.1 and documents how the 2016 poetry slam was executed (Section 9.2). It finishes with a short reflection of 'Lessons Learned' in Section 9.3.

9.1 Social Engineering Poetry Slam Concept

Similar to a poetry slam, the SE Poetry Slam is moderated on stage by the poetry slam hosts. Speakers, the so-called slammers, present their anecdotes within a given time frame. The slammers and the audience come from the ethical hacker community. Hence, the poetry slam can be organised for suitable hacker conferences. As the name suggests, the main poetry slam topic should address SE. Slammers are urged to focus on stories covering SE. The major motivation is to enable the ethical hacker community to talk about SE in an entertaining environment. Again, the understanding of what comprises SE differs. A brief introduction to SE can focus the contributions by using the SE indicators (Section 3.1) or by providing examples.

Slammers can be recruited during the session. It is handy to have a few slammers registered before the event to start smoothly. Usually, poetry slams have a final round of the best three slammers who present additional contributions different from their initial ones. Many slammers will not have a second story prepared if they decide to participate spontaneously. It is thus recommended to avoid a second round to allow as many slammers as possible to participate. The winner of a typical poetry slam receives an award. This can be one motivation to join. A present for each slammer can motivate, too. To offer this platform to

talk about SE can be an entertaining factor. Another motivation can be the aspect of 'being heard' or 'telling a story'. Hutchings and Holt [HH18] identified 'to safely have their voice heard' as one motivation for offenders to participate in interviews (Section 6.1.1). Although slammers are urged not to talk about criminal activities, this motivation seems to fit here as well. So far, no evaluation of the motivation has been conducted.

9.1.1 Audience becomes Jury

The competitive nature is comparable with the TRE$_S$PASS SE award, but more interactive between slammers and audience. The SE award[1] was granted to the authors of the best SE attack scenarios, including a presentation and panel discussion at the CPDP conference in 2015. Here, the authors submitted their written SE scenarios. A jury of experts from different disciplines chose the best scenarios and invited the author of the best contribution. Authors were motivated to submit scenarios to win the award endowed with prize money. In comparison, all presentations in the poetry slam are public to the audience and can be used for later analysis, even if they were downvoted. The slammer's contribution is not submitted to a jury previously selected. That is, the audience becomes the jury like in other poetry slams and the organisers do not have any clue which specific story will be presented.[2] This is not a disadvantage: although the slammers and audience receive SE information and recommended evaluation criteria, such as presentation style, novelty, creativity, feasibility of attack, feasibility of real-life experiments or scalability of attack. The evaluation via scoring can result in anecdotes that do not express SE at the end. Recorded sessions can mitigate this issue and researchers can evaluate the material afterwards according to their SE definition and criteria. This leaves the session interactive and avoids hindering the slammers. Hence, presenting even non-SE stories becomes more important in such settings than the pre-selection of contributions by a jury.

The audience needs to vote for their champion. This procedure can be manifold and should correspond to the environment. Possible approaches include online voting, noise level per slammers etc. Typically, the presentation time is restricted. It can be extended by the audience through initially agreed signals, such as shouting 'more' or stomping.

9.1.2 Pseudonymity of Slammers

To protect the participants' identity as well as to provide a sufficiently safe feeling to present, three styles of presentation can be offered:

- slammer presents
- slammer presents in a protective suit or with a mask (Figure 9.3(a))
- slammer uses a 'medium' who reads the written slam contribution aloud (uebelhacker as a medium presenting agnepix' poem in Figure 9.3(b))

Slammers can enter by a stage name. Organisers of a poetry slam usually do not know what will presented. When talking publicly about SE precautions are needed. Slammers are advised not to incriminate themselves and not to talk about planned attacks. Although an *ethical* hacker community should be chosen because of the expectation of more responsible

[1] The winner of the 'Cybercrime Social Engineering Analysis Challenge' was Demetris Antoniou (youtube)
[2] A registered slammer's title may reveal some clues, but not about the content in detail.

actions, one cannot anticipate if some contribution can be seen by someone as an offence now or in the future. Pseudonymity offers a limited protection. This precaution resembles the safety measure for offender interviews to avoid discussions of future activities [HH18] (Section 6.1.1).

Because the poetry slam should be recorded for later evaluation, slammers are given the option to be removed in the final cut prior to publishing. For various reasons, a slammer may become hesitant in hindsight. This opt-out should be communicated before slammers tread the boards.

9.1.3 Post-Session Evaluation

A recorded session is valuable for later evaluation. It can be transcribed and transformed into anecdotes upon which further research can be conducted. If the recording is released under the previously mentioned constraints to protect slammers, the entertainment factor can help to convey awareness against SE attacks. For future SE Poetry Slams, interested slammers or organisers can also inform themselves about how this specific slam can be performed. In fact, fellow researchers can confirm or refute conclusions drawn. Hence, shareability is achievable to enable constructing a transdisciplinary knowledge base (Section 5.3) on SE (see general shareable knowledge in Spring and Illari [SI18]). A shareable license applicable for Open Data publications is highly recommended, e.g., Creative Commons[3] licenses. Content removed from the recording should not be considered for evaluation as the contribution cannot be transcribed and analysed thoroughly without sufficient time. It would rely on memorised stories and hinder fellow researchers to re-evaluate the anecdotes.

9.2 Social Engineering Poetry Slam @ 33C3

At the 33rd Chaos Communication Congress (2016) the SE Poetry Slam was held as a self-organised session by uebelhacker, Anna Fuchs, and ysf. The session was recorded and published in the TUHH Open Research (TORE) portal under Creative Commons License (CC BY-SA 4.0) as OpenData [UR16]. kolAflash cut and transcoded the recording into videos (webm, mp4; 720p, 1080p) and audio only (mp3, opus/ogg). kolAflash also created the opening credits (Figure 9.2.1) and removed two contributors upon request.

9.2.1 Advertisement & Pre-Slam Communication

The main information about the procedure was announced on the conference wiki (see Appendix C.5) with the keywords ‘social, art, game, hacking, security’. An abstract of the wiki version was added to the c3nav app which is used to manage all events, see Figure 9.2.2. Posters at the event location linked to the conference wiki. Additionally, the poetry slam was advertised on the conference screens. On the wiki page SE was shortly introduced as: “a human interaction needs to be present to enable the attack, i.e., dumpster diving is not social engineering, it’s just gathering pre-attack information, but (spear) phishing or scams like the ‘Enkeltrick’ are. The community discusses some persuasion principles of why people succumb to these attacks. Presentations can base on the principles

[3] Creative Commons: https://creativecommons.org/

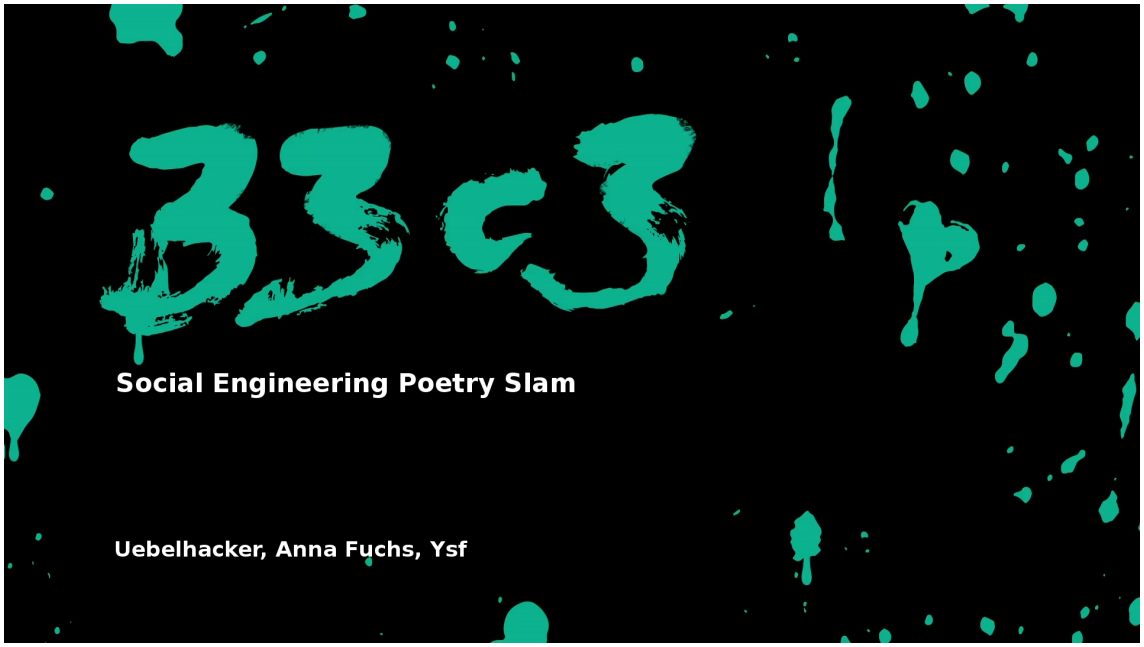


Figure 9.2.1: Opening credits of recording created by kolAflash [UR16]

of Cialdini or Stajano/Wilson." It links to the publications of Cialdini [Cia07], Stajano and Wilson [SW09], and Mitnick and Simon [MS02], and furthermore to the TV series 'The Real Hustle'[4].

A dedicated DECT phone number (`7526` — originating from the letters on the dial pad of '`SLAM`') was registered to work as one communication channel. The DECT telephone system is commonly used during these conferences. The phone was handled between the slam organisers. One false phone call was received. An e-mail address (`33c3-slam@datapirate.de`) including a PGP key (`0xD42C10B0E28B80DC`) was provided. Incentives were not announced before the event. At the event, it was mentioned that each contributor will receive a gift and the best slammer will win a bottle of vodka.

9.2.2 Results

No slammer registered before the event. Fourteen participants presented their SE story spontaneously. More would have liked to participate, but the session's time slot was limited. It is assumed that as the format was new, after the first slammers more wanted to join. The audience consisted of around 300 hackers. Around 50 more people tried to enter, but the workshop room was already crowded and the entrance blocked. Table 9.1 lists all contributions. Slammers 'no name' and 'Walther' were removed from the recording upon request. Of these twelve, one wore a mask (Figure 9.3(a)) and one used a medium (Figure 9.3(b)) for pseudonymisation, i.e., ten slammers presented without any pseudonymisation. No one requested a protective suit. After five minutes a rooster sound was played to indicate to finish each presentation. In one case the audience wanted to hear more signalled by a wild applause (Slammer Björn (#14), Section 9.2.2). The audience voted via a webtool.

[4] 'The Real Hustle', BBC, 2006–2012, https://www.bbc.co.uk/programmes/b006m8mf
[5] https://f-droid.org/app/de.c3nav.droid

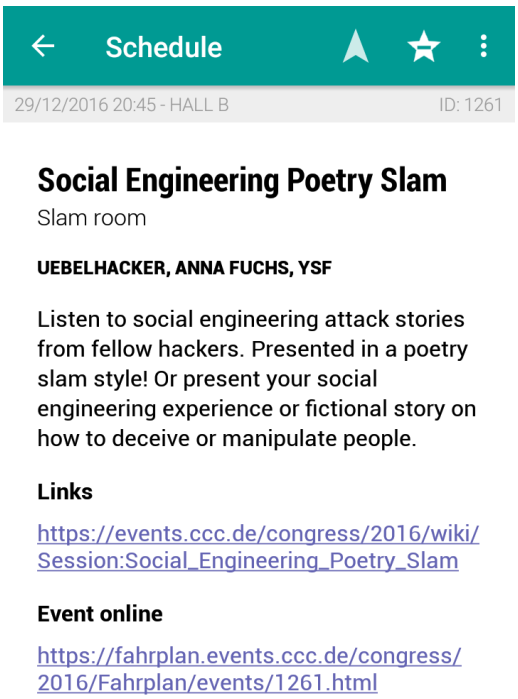


Figure 9.2.2: Android app c3nav[5] displaying the event

(a) slammer used a mask to hide the identity

(b) medium uebelhacker presented "My 1st German Poem" written by slammer agnepix

Figure 9.2.3: Two types to achieve pseudonymity (mask, medium)

#	Slammer	Score	Title	Pseudonymisation
#1	Jacob	283	Beginners Engineering	mask
#2	Chris van't Hop	314	mail from non-locked computers	none
#3	Spip	174	spit	none
#4	Lyndis	257	Writers' Group (Anecdote 9.2)	none
#5	Ulrich	235	Access Control (Anecdote 9.3)	none
#6	kolAflash	357	Handling Hotlines	none
#7	cyremur	232	Project Inception	none
#8	agnepix	349	My 1st German Poem	medium
#9	Daan	319	You just need to sign off on it.	none
#10	~~no name~~[6]	235	~~no title~~	n/a
#11	~~Walther~~[6]	323	~~ice bear on stage~~	n/a
#12	Oliver	334	A story about my friend	none
#13	Sorry (Marc)	**457**	Fuckhochschule (Anecdote 9.1)	none
#14	Björn	315	Start-up Social Engineering (Anecdotes 9.4 & 9.5)	none

Table 9.1: Participants plus scores and pseudonymisation type at the 33C3 Social Engineering Poetry Slam

The contributions ranged from sending e-mails from unlocked computers to raise awareness and other non-malicious use of SE techniques over tailgating to SE penetration testing. The level of detail differed. Many stories were not expressing SE with respect to malicious intent (SE Indicator 3.4) according to Definition 3.17. Relevant contributions are discussed here and checked against the five SE indicators of Section 3.1 to identify SE (Research Question 1.2), starting with the winning slammer.

Slammer Sorry (#13): "Fuckhochschule"

Slammer #13 Sorry (Marc) won with his talk called "Fuckhochschule" by 457 votes (Anecdote 9.1). Without being able to ask the voters, a plausible explanation can be the unintentional mispronunciation of the German 'Fachhochschule'[7] by a non-native speaker. With this hilarious start and a real story about a large-scale e-mail attack to gather passwords, the prize was well awarded. The contribution was transcribed in Anecdote 9.1.

For sure, the targeted persons who received the e-mail to provide their username and password enabled successful SE by their actions (SE Indicator 3.1). The attacker communicated via a distribution list writing to a lot of targeted persons directly and intentionally (SE Indicator 3.2). The targeted persons were unaware of the attacker's intentions (SE

[6] Participants 'no name' (#10) and 'Walther' (#11) were removed from the recording upon request. The pseudonymisation type cannot be determined and re-evaluated by fellow researchers anymore.

[7] in English: University of Applied Sciences

Indicator 3.3). They were deceived (SE Indicator 3.5) in the hope to get help with a problem (unsubscription). The likely deceptive technique was pretending to be able to solve this issue as an IT administrator, thus, impersonation of an authority figure (Section 4.5.2, 'Authority by Hierarchy' because of administrative privileges). An experienced time scarcity might also lead to a success: the targeted persons wants this issue solved as soon as possible. The attacker did not create this pressure, just used an already existing issue. This is related to the 'Scarcity of Time' principle (Section 4.5.2). The attacker may have exploited gullible persons and their trust in the goodness of mankind (Section 4.3.1). Also, because this anecdote is strongly communication-focused, the targeted person may be biased according to the Pollyanna principle (Section 2.2.2).

> **Anecdote 9.1 — Slammer Sorry: "Fuckhochschule".** "Some years ago I went to a bachelor university which fell under the same organisation in multiple locations in the country; this was the Netherlands. The German word for this type of this is, I believe, Fuckhochschule. My German is horrible, but I assume you know what I mean. In another location someone sent an e-mail to some group, some group mates. But she accidentally sent it to the entire faculty. And someone replied and people started adding more and more faculties and more and more locations of this school around the country. And it became one giant cluster fuck. Everyone was replying to e-mails. You got at least 150 e-mail a day. And people were asking to be unsubscribed from this e-mail list. Which was not possible because this was an e-mail which was given to you by your university. So, first I'd like to explain in a reply-to-all obviously: 'This was not possible'. You have unsubscribe from your education you're following to remove yourself from the e-mail list. And people kept complaining about it. And at a certain point I was sick of it because I got 250 e-mails a day. And I couldn't find any sensitive e-mails anymore. And I said, 'Please, send me your username, password, and motivation to get removed from this list.' And I got literally more than 100 replies with an actual motivation and an username and password. As a side note people already said: 'This mailing list is annoying, please, remove me'. And no one even checked."
> [UR16, #13]

If this anecdote is based on a real event, over 100 revealed username password combinations is a lot. Regarding the remaining SE Indicator 3.4 (Attacker: Malicious Intent with Goal) it is assumed that the slammer had no malicious intent at all, but mimicked it. This anecdote presents a surprisingly feasible attack vector using SE. One can think of a hacked e-mail account via which the attack is executed to avoid mimicking malicious intent. Similar situations can offer similar attack vectors. This anecdote expresses SE. Since the attacker jumped onto the bandwagon of unsubscription requests, the attacker did not create or plan the users' wish, but abused the situation. This is a case of 'Opportunistic SE' (Definition 4.5).

Slammer Lyndis (#4): Writers' Group

Anecdote 9.2 of slammer #4 (Lyndis) expressed definitely good intentions and not mimicking malicious ones (non-malicious, ~~SE Indicator 3.4~~) to get people join her writers' group. The success depended on the targeted persons decision to join (SE Indicator 3.1) while being unaware of this deception (SE Indicator 3.3, SE Indicator 3.5). The later

honesty towards the targeted persons about the deception (ethical 'debriefing') was not counterproductive. The initial communication of the 'attacker' was bidirectional and intentional (SE Indicator 3.2). The targeted persons were committed due to their own interest prior to the attempt. Then stayed consistent when talked about. The principle of 'Liking' can be assumed here, too, due to similar interests. By revealing personal information ('opening up') the attacker may have triggered a reciprocal response when the targeted persons revealed something as well, consolidating a bond. In summary, a few of Cialdini's principles (Section 4.5.1) may be involved here: Commitment & Consistency, Liking, and Reciprocity. However, because neither malicious intent nor mimicking one was identified (~~SE Indicator 3.4~~), this anecdote does not express SE.

> **Anecdote 9.2 — Slammer Lyndis: Writers' Group.** "My story is to do good things with Social Engineering. I am a writer and when I came to my university I thought I could get to know other people who write 'cause it's a university. There might be some group that formed, but it didn't because it's a really small university. So, I founded my own group. And finding the first two people were really really easy. But expanding the group was really hard because we were a group, we had our insiders, we knew each other and writers often are shy and introverted. And didn't get along so well with the group that was already there. So, I started my own technique to deceive them to stay. [audience: you deceived them all!] And it's really really simple because they only need a person who they can connect with. So, every time someone came to our group. A few days later I catched them at the campus and I talked to them. And I said a bit about me, really personal stuff, so they could connect to me. And they opened up to me and said something personal about them. And if this little bond is formed, suddenly it was really easy to made them stay because they liked me, they could connect to the group, and to they stayed and now I have something about 12 people in my group. And we all get really really good along with each other. And the nice thing about this is, I said to them 'Hey, I deceived you to be here.' I explained to them how I did it and they laughed, called me a psychopath and stayed." [UR16, #4]

Slammer Ulrich (#5): Access Control

Anecdote 9.3 of Slammer Ulrich (#5) stands out as it is narrated as the targeted person. Tailgating is a well-known attack to enter premises protected by physical access control system or visitor policies, see Anecdote 1.2 occurred at Hamburg University of Technology (TUHH). Typically, a genuine employee (targeted person) opens a door, enabling the attacker to follow close behind (SE Indicator 3.1). The targeted person may notice it and does not intervene because of being a nice person. It can be speculated whether the targeted person is aware of breaking an organisational access policy or of granting access to an unprivileged person (both according to SE Indicator 3.3). Probably, the targeted, deceived person assumes that the attacker is a colleague (SE Indicator 3.5) — eventually wearing a fake, non-functional badge. Not in this anecdote apparently, but in general an attacker would enter with malicious intent (SE Indicator 3.4). It is not clear whether an intentional communication initiated by the attacker is needed; non-verbal communication could be imagined here.

> **Anecdote 9.3 — Slammer Ulrich: Access Control.** "What I am talking about is something all the people working for bigger companies have seen in their daily life. It's something about access control. If you have something like a access card to enter your building and you arrive by your bike and somebody else also arrives with his bike. And maybe say 'Hi' and when you come along the entrance, he says 'Hi, sorry, I forgot my card. Can you please let me in.' And if this person is really nice, maybe really charismatic or something like this. You would never say 'no'. And this is something, I wouldn't say 'no'. Who would say 'no'? Please, raise your hands! Even some people. But doesn't that feel a little bit harsh? I don't know. Because if something is really empathic, really nice, really charismatic, how tough is it to deny him of letting in? And this is something even if I am aware of this, it's really hard for me to do this in reality. And so I was compromised already." [UR16, #5]

Slammer Björn (#14): Start-Up Social Engineering

Slammer Björn (#14) presented three different SE attack stories they performed in a penetration test for a company.[8] Two of which exploit human traits to succeed. The first one (Anecdote 9.4) offered the penetration testers to enter the premises by distracting security personnel and employees with a cute puppy. The second one (Anecdote 9.5) took place directly after when they already had access to the building. They then used the cleaning personnel workwear found to access the restricted server room copying files, cf. Anecdote 8.2 and Section 8.2.2 about suspecting cleaning personnel. The slammer was able to present this long story because the audience decided to applaud after the official time was over.

In Anecdote 9.4, the malicious intent (SE Indicator 3.4) must be *mimicked* in SE penetration tests, partly because of ethical reasons (Section 5.2.1). Without letting the penetration testers pass the entrance, the act would not have been successful (SE Indicator 3.1). However, it is not quite clear of 'how' they got around the security personnel, probably using another entrance. A communication was planned, but comprises more of indirect elements (SE Indicator 3.2). The puppy let employees start the conversation. The intentional distraction and focus on the dog deceived the employees (SE Indicator 3.5). It left them unaware of an ongoing act of SE (SE Indicator 3.3). Similar to the classical tailgating technique (Anecdotes 1.2 & 9.3), this one adds the element of distraction. With a more precisely elaborated story part on how the premises are entered, this anecdote can express SE easily.

Once they entered the building, they checked rooms and found the workwear of cleaning personnel (Anecdote 9.5). They impersonated the cleaning personnel and in a group effort even trespassed into the server room. The technician enabled the successful access (SE Indicator 3.1). The penetration testers communicated intentionally and bidirectionally under the pretext that they needed access to the server room (SE Indicator 3.2). The technician was unaware of SE (SE Indicator 3.3) and tricked by the impersonation of cleaning personnel (SE Indicator 3.5). Like above, the penetration testers mimicked malicious intent (SE Indicator 3.4). They provided proof of their successful attack by copying files.

[8] more on ethical SE penetration testing and methodologies in Section 5.2.1 and Dimkov et al. [Dim+10]

Anecdote 9.4 — Slammer Björn: Start-Up Social Engineering (Puppy). "[...] So, we ended up going there with a whole group of people. And the first thing was rather simple. It was a big office building and they have a, you are not allowed to bring any pets there, of course. So we tried out something simple. We just went to the next animal shelter and there was this, yeah, cute, little, fluffy puppy. And it was really cute with its eyes and tipsy and everything is nice. So, we just took the dog on a leash and went for a walk and we went in there like we would have guessed that the first security would go up and like 'Yeah, you are not allowed to bring that dog in here'. Okay, nothing happened. So we went around and then we were in the middle of all the secretaries there. Instead of anybody saying, we should get out. Everyone was like 'hey, there is a fluffy dog, just get there.' So, we were three people on the that part. First one going through, everyone was like going nuts running after them. And then we were in the cafeteria, just drinking coffee. And like there was this dog and the dog was playing around and doing 'sit' and 'give your paw'. And like yeah, it was funny. Second guy was just staying at the printer trying to 'print' something. And the third guy had all the time in the world to going through the offices because in all the haste you get to the dog and play. Well, nobody locked their computer, e-mail programs were open, and my really favourite, my friend had told me: 'Yeah, there is this one paranoid guy. He uses a password manager with secure passwords.' Well, it is nice to have secure passwords, it's nice if you cycle them, but, shit, if you don't lock that account. [...]" [UR16, #14]

Here, by chance the penetration testers were left alone in the server room. They did not plan to let the technician spill his coffee or have an urgent meeting as it seems. They improvised suiting the situation. But one may assume that sooner or later the technician would have left because the cleaning process would have taken some time. Nevertheless, the anecdote expresses SE. It is not clear how the technician contacted the fake cleaning personnel. The penetration testers already had access to the premises as this was a subsequent attack.

Anecdote 9.5 — Slammer Björn: Start-Up Social Engineering (Janitor). "[...] So, we just looked around and we saw there is one open door. Ah, okay, look in and a bit looking around. There is this nice uniform of the cleaning staff in here. [...] Cleaning stuff is like important, yeah, they clean the stuff you don't want to clean it; you're just wanted to work. But you let them in, because they should [clean] the stuff and you can continue working [...] But we took the cleaning outfit and then just told the technician guy, go to, yeah, we wanted to make it count, go to the server room. And yeah, he accidentally spilled his coffee. So he had to [run] into an important meeting. So, he just called the cleaning staff and then he directly called us. And then the other guys should have let us in. And like for cleaning up the coffee. They never saw our faces before, never ask for a name or for anything. We just were allowed to roam in there for free and they never checked if we went out. We actually were so funny to copy all the contents of the file servers there and left with that. [...]" [UR16, #14]

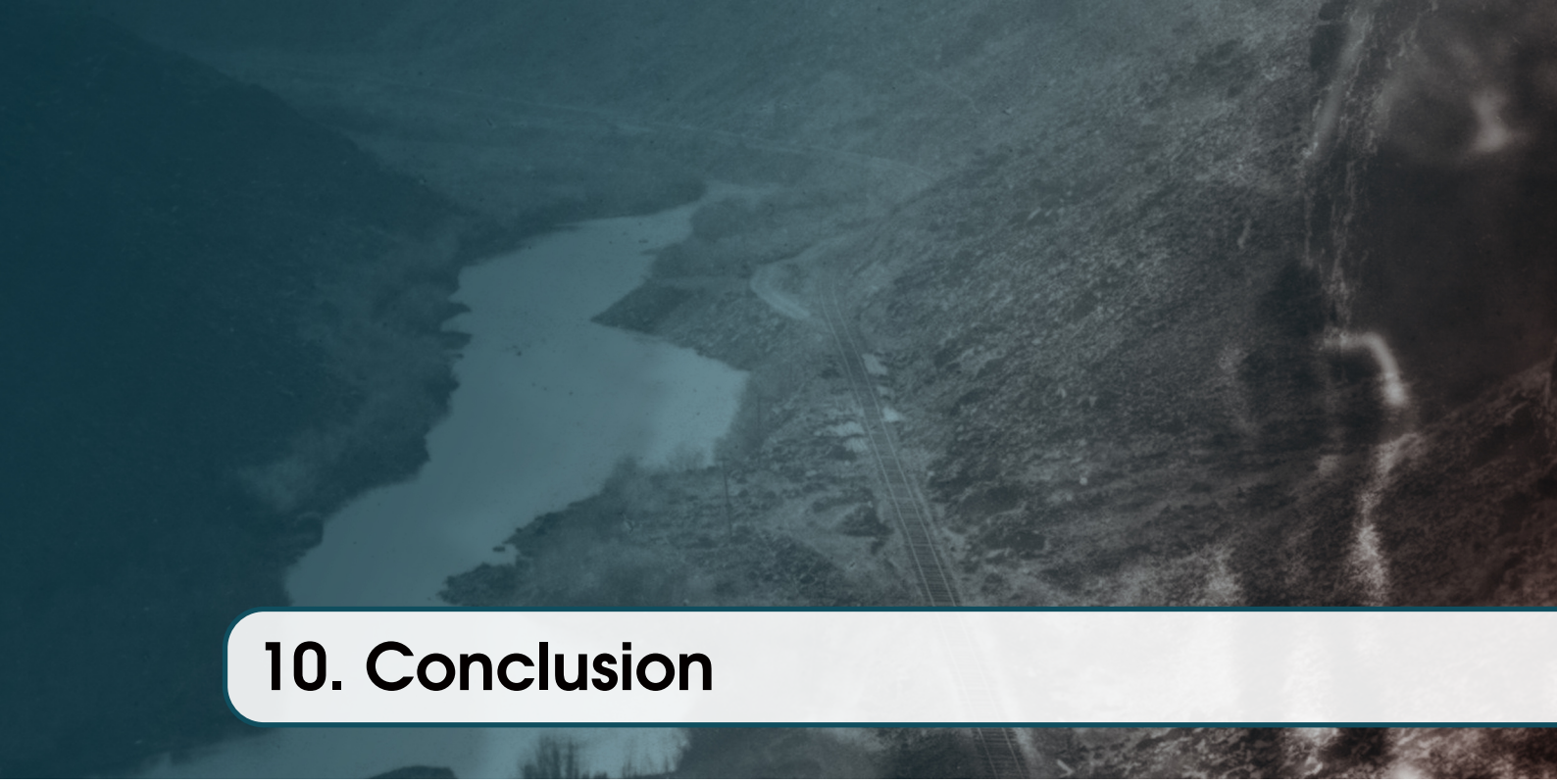
9.3 Lessons Learned

The introduction of this new format to the ethical hacker community revealed the challenge to find slammers before the event started. Pseudonymisation efforts were not required as often as expected. But helped kickstarting the first talks. The organisers found one fellow hacker beforehand who was willing to present under some circumstances. It is important to have that ace in the hole.

It was difficult to impose guidelines about what comprises SE to the contributions as well as audience scoring. The wiki (Appendix C.5) contained sufficient input on SE, but was probably not considered by all. Hence, a brief introduction was given at the slam. The challenge is to provide good guidelines and simultaneously prevent too strict formulations hindering participation. The organisers cannot control the audience's scoring behaviour which is good. Therefore, it was crucial to have a recording for later evaluation. The used one-dimensional scoring system could be enhanced by other aspects like the five SE indicators (Section 3.1). The audience would have to decide how well each indicator was present — which can end in a more burdensome voting process.

Some presentations were lacking some details needed to see the full SE picture. There is no efficient backchannel for post-slam interviews if organisers cannot contact the slammers later. To guarantee pseudonymity organisers should not be allowed to gather contact details. In summary, the SE Poetry Slam is a usable format and can motivate ethical hackers to contribute their ideas on SE. Novel insights can be gained, e.g., Anecdote 9.4 showed an interesting distraction effect by using a puppy.

Conclusion



10. Conclusion

10.1 A SE Definition Fit For Security Research

As a foundation for this thesis, five Social Engineering (SE) indicators were developed in Chapter 3: firstly, to find a definition of SE appropriate for SE research in multiple disciplines and, secondly, to identify evidence that contains SE in alignment with the SE definition (following the Research Question 1.2 in Part II). As SE research is interdisciplinary, a suitable SE definition usable in various disciplines is needed that can foster a common knowledge base. With this interdisciplinary prerequisite, the SE indicators were developed and existing SE definitions were evaluated whether and in which modality they express SE according to the SE indicators: explicitly or implicitly, partially, too general or not at all. The SE indicators comprise the following components (Table 10.1): the main *enabler* of a *successful* SE attack relies on the reaction of the targeted person (SE Indicator 3.1). The attacker uses deceptive techniques (SE Indicator 3.5) with malicious intent and a goal (SE Indicator 3.4) to attack an unaware targeted person (SE Indicator 3.3). To achieve that, the attacker must communicate intentionally with the targeted person in indirect, unidirectional or bidirectional form (SE Indicator 3.2). "A targeted person becomes a *Social Engineering victim* iff the SE attack targeting that person succeeds benefiting the attacker's intentions." (Definition 1.4)

The SE indicators are also independent from a changing threat landscape. They can identify SE in classic grandparent scams, but also in new attack vectors, such as spear phishing for carbon emission certificates. Attackers can choose recent threats from the news that are perceived as more vivid for targeted persons, making use of the availability bias. Such type of SE were called 'Opportunistic SE' (Definition 4.5). The COVID-19 pandemic is one example where attackers tried to lure self-employed targeted persons into using fake corona relief fund websites (Hamburger Corona Soforthilfe) or impersonating health care workers to interview targeted persons as a distraction. Similarly, when the General

SE Indicator 3.1	Targeted Person: Human Enabler
SE Indicator 3.2	Attacker: Intentional Communication
SE Indicator 3.3	Targeted Person: Unawareness
SE Indicator 3.4	Attacker: Malicious Intent with Goal
SE Indicator 3.5	Attacker: Deceptive Techniques

Table 10.1: Overview of SE indicators from Section 3.1 (copy of Table 3.2)

Data Protection Regulation (GDPR) was introduced in the EU and European Economic Area (EEA) and prevalent in the news, phishing e-mails appeared, e.g., urging targeted persons to enter banking details (Figure 4.3). Anecdote 9.1 originating from the Social Engineering Poetry Slam (SEPS) showed a case of Opportunistic SE as well. The SE indicators can support researcher here reliably and are usable in longitudinal studies and even if the opportunities for attackers change.

Regarding Research Question 1.1, none of the analysed SE definitions expressed all postulated SE indicators explicitly. Hence, a suitable SE definition (Definition 3.17) was created that incorporates all SE indicators explicitly. The five SE indicators and the corresponding SE definition are designed to be reusable for fellow researchers of other disciplines. The SE indicators define the scope between a very broad view of 'everything is SE' and restrictive ones like demanding the involvement of technology. Some existing SE definitions *restrict* their view on SE to have digital technology involved, however, the understanding of SE in this thesis is broader: the grandparent scam ('Enkeltrick') can occur in-person at the door of targeted, often elderly, persons — no technology involved, but interpreted as SE if all SE indicators match. Some SE definitions draw their scope *too general*, e.g., bribery of a targeted person describes an act where that person is aware of it, thus, is interpreted here as not expressing SE. Another analysed SE definition can be interpreted as that an 'attacker' may be a benevolent actor using SE. Other SE definitions present aspects *implicitly* that are seen as crucial for SE and are made explicit in this thesis, e.g., the intentional communication between attacker and the targeted person initiated by the attacker. The explicitness of the SE indicators incorporated in the SE definition fosters a precise identification of SE and avoids interpretation. As a result the SE definition becomes longer.

Deceptive techniques can be based on one or more persuasion principles such as scarcity. Predominant principles in the SE literature are the six basic principles by Cialdini originating from marketing and sales as well as the principles by Stajano and Wilson focusing on susceptibility to scams. The former shows that insights from other disciplines (how to sell goods) can fit very well into the multidisciplinary research of SE. These principles can explain how SE attacks succeed. For instance in marketing and sales, a limited offer or an artificial shortage of a specific good can motivate consumers to buy now to avoid losing that opportunity (loss aversion; scarcity). This perceived time pressure can also be used to lure targeted persons into complying to an attacker's request. Having the SE indicators in mind, principles re-used from marketing and sales can be applied to create malicious actions such as SE.

10.2 Data Collection and Identification of SE

The SE indicators are of dual-use: they are applicable for identifying SE evidence of multidisciplinary origin. That is, Research Question 1.2 can be answered to find SE by satisfying all five SE indicators. These SE indicators are a more suitable approach compared to existing publications consisting of mainly definitions. The only approach, resembling the SE indicators, was developed as an ontological model for SE *attacks* [Mou+14a]. Chapter 6 grouped possible sources of SE evidence into interviews, experiments, and literature in the broadest sense. All of which contain advantages and disadvantages regarding the evidences' novelty, veracity or verifiability. That is why the term 'anecdote' was chosen for all types of SE evidence as explained in the introduction (Section 1.7). The SE indicators showed whether an anecdote expressed SE. Regarding SE experiments, the targeted persons (subjects) must face a *fictitious* malicious intent (SE Indicator 3.5) where no major harm to the subjects is involved (ethical considerations in Section 5.2.1).

The following sources were examined in greater detail: Chapter 7 analysed court documents as evidence source ('literature'); the author conducted a Lego modelling session ('interview'; Chapter 8); the author created, organised and moderated events at ethical hacker conferences in poetry slam style to let hackers tell their SE anecdotes ('interview'; Chapter 9).

Court Documents

Although the main functions of court documents (Chapter 7) are communication, control, and legitimation for the judicial process and the public, the truth finding process to revise, e.g., previous decisions, shapes a reliable source of real events for evidence-focused SE research. Furthermore, court documents are often accessible to the public, depending on each country's publication practice. The author and fellow researchers examined court documents iteratively with respect to phishing as one deceptive technique and presented the predominant workflow found in phishing cases. Challenges for SE research were categorised and presented when consulting court documents. Besides the typical roles of attacker and targeted person (Section 1.6), the money mule role remained nonassignable to these roles in the court documents: is a money mule a gullible or naïve targeted person who fell for the trick to become a financial agent for the attacker (phisher)? Or is the money mule well aware of laundering money and pleads innocent in court? Despite being historical documents published after the court ruling, a novel attack vector for research was found: a successful spear phishing attack against organisations that are trading carbon emission certificates in Germany.

Lego Modelling

Efforts to create cross-functional groups to model SE attacks can hint to novel SE insights. Modelling with Lego bricks offers a feasible approach because the handling of the bricks is well-known to almost all participants. Based on a cloud environment scenario and defined personas from the TRE_{S}PASS project, TUHH students from different academic disciplines created two scenarios (Chapter 8). Regarding the SE indicators defined afterwards, none of the scenarios expressed SE. For future modelling sessions, these SE indicators can guide participants to create better fitting scenarios regarding SE. Although this was an academic setting, Lego modelling can be accomplished in an organisational context with cross-functional teams of various professions.

Social Engineering Poetry Slam

The developed SEPS concept in Chapter 9 showed a new approach to create a platform for ethical hackers to present SE anecdotes. The audience (and researchers) can be kept in the dark about whether any anecdote was fictitious or not. Because the audience becomes the jury, a validation during the event cannot be achieved easily whether SE was involved from a researcher's perspective. Therefore, a recording is crucial for research to transcribe and examine the anecdotes whether they match the SE indicators. Moreover, fellow researchers can later re-evaluate the results if the recording is available. To motivate more hackers to participate, optional pseudonymisations were introduced (mask, protective suit, presentation via a medium). This novel SEPS was conceptualised by the author and conducted together with fellow ethical hackers. An interesting contribution talked about physical penetration testing to enter an organisation's building: a puppy was used by the penetration testers to distract the (security) personnel and enter the premises successfully.

Summary

The multidisciplinary nature of SE research dictated to delve into applicable fields, but also to focus on cybersecurity. Sources of evidence differ in the level of detail, their original purpose or expressing real or fictitious acts. Novel insights can be gained by subsuming the stories of various sources under the term 'anecdote'. The use of the developed SE indicators ensured an interdisciplinary identification of SE. The exemplary Lego modelling, the conducted court document analyses as well as the conceptualisation of the SEPS showed where evidence-focused SE research can commence.

10.3 Susceptibility of Targeted Persons

Chapter 2 showed what kind of *general* factors on which level of "human mental programming" (term coined by Geert Hofstede) may influence the behaviour when dealing with cybersecurity. The three levels comprise universal human nature, culture specific to groups, and an individual's personality. This thesis focused mainly on the targeted persons and less on the attacker. Some targeted persons may become victims more likely than others. For instance, they are reacting differently on loss aversion or scale higher in overconfidence (human nature). The cultural background may influence a group's value system towards more individualistic or collectivistic goals. The personality, unique for each targeted person, can express what motivates an individual, such as a subject high on the extraversion trait is looking for excitement and social attention. Most of these summarised human factors are measurable in one form or another. An attacker can exploit these general factors to craft attacks whether based on universal, group-related or individual aspects of a targeted person.

To better understand how SE works, Chapter 4 elaborated on its components alongside SE examples. The general human factors from Chapter 2 were complemented by aspects more present in SE such as gullibility which relates to an unaware targeted person (SE Indicator 3.3). Deceptive techniques were discussed focusing on phishing, impersonation, and the apple road attack. According to SE Indicator 3.5, deceptive techniques incorporate at least one persuasion principle. Persuasion principles, especially those developed by Cialdini and their refinements, are commonly used in SE research to explain underlying mechanisms. They were summarised and complemented with the mode of thinking of

targeted persons why they may fall for that principle. Each principle were appended with consulted best practice defences including self-questioning phrases for targeted persons. Finally, the developed Social Engineering Personality Framework (SEPF) based on literature review explained possible correlations between the Five-Factor Model (FFM) personality traits and Cialdini's persuasion principles regarding the susceptibility towards SE. This framework was created by a TUHH student supervised and then adjusted in this thesis by the author.

Research to characterise whether a targeted person is more or less susceptible in particular relies on many factors and still lacks many insights. Addressing Research Question 1.3, the need for more as well as reproduced or corroborated experimental approaches arises to grasp these multidisciplinary human factors. It will be then possible to examine data of cross-cultural or longitudinal studies. The latter does not exist in cybersecurity research yet [Leu17].

10.4 SE Anecdotes to Transport SE Situations

SE anecdotes can transport the intended content to the desired audience as the research field narrative psychology reveals (Section 1.8). Stories resemble the mode of thinking like an easy-to-swallow pill and are processed with less mental effort. Hence, SE situations can be communicated using anecdotes that complement organisational policies, awareness trainings and other educational campaigns (Research Question 1.4). For instance, the German IT-Grundschutz Catalogues [BSI13] of the German Federal Office for Information Security (BSI) contains example stories for the categorised security threats to foster a better understanding. And vice-versa, as this mode of thinking is a universal human component, SE anecdotes can be collected for research from different data sources. Such as court documents can tell stories of phishing cases in this thesis.

10.5 Science of Security for SE Research

Chapter 5 elaborated on the question of how science can be practised in security, starting with the philosophy of science. The same discussion about science of security applies also to science of security for SE research of Socio-Technical Systems (STS). Criticism and challenges for finding evidence in security research were discussed such as claiming security experiments are untenable and impossible to reproduce. The former claim contained the challenge of how to deal with ethical and privacy aspects in experiments that apply deceptive techniques (SE Indicator 3.5). The latter differentiated between the types of reproducibility and their possible benefits for SE research. Due to the lack of a Research Ethics Board (REB) at the Hamburg University of Technology (TUHH), the author refrained from deceptive experiments. These general challenges were addressed and applied to SE research with respect to how to build an interdisciplinary knowledge base. An inductive-deductive modelling cycle was consulted to create this knowledge base based on structured observations. This approach was applied, e.g., to create the SE indicators and the SE definition. The chapter closed with the statement that, as a first step, *focusing* on the various facets of interdisciplinary SE evidence (evidence-focused SE research) is more feasible rather than to restrict the list of SE evidence sources (evidence-based SE research). That is, the focus on where evidence may be found can commence finding

plausible SE anecdotes as a first step — even ones that are fictitious, hard to verify their veracity or of unknown origin. Some anecdotes may come from hearsay because of the well-motivated secrecy in information security to share evidence [SI18]. In a second step, further research such as conducting experiments can then foster evidence-based research. Thus, some anecdotes expressing the SE indicators may lose the attribute of being labelled as fictitious.

The SE definitions examined are shorter and do not break down in explicit indicators. The merit of the explicitness is its applicability to a security context within STS. Concise definitions may omit aspects that lead to varying interpretations in an interdisciplinary context. Hence, the precise SE indicators were favoured to be able to create an interdisciplinary knowledge base. The knowledge base on SE can be filled by novel insights or by results from reproduced experiments. The latter is important to show the development over time (longitudinal studies), to compare differences and similarities between cultures, professions, age groups etc. For instance, the Lego modelling session was reconducted with TUHH students where a tech-savvy mindset can be assumed and the participants were assumed coming largely from Western, Educated, Industrialised, Rich and Democratic (WEIRD) societies. Further sessions in different settings are needed for more profound conclusions.

10.6 Outlook

SE anecdotes can lead to SE experiments for further research. Also, anecdotes not expressing all SE indicators explicitly may be transformed to express usable (fictitious) SE anecdotes. The author normalised SE anecdotes into SE scenarios (unpublished). The author conducted two questionnaires with security experts to identify which persuasion principles may be more prevalent in each SE scenario. These SE scenarios could be translated into a Natural Semantic Metalanguage (NSM) [GW04] to become more modifiable to address specific organisational or cultural aspects. Furthermore, the author added the Ten Item Personality Inventory (TIPI) questions to the questionnaires to differentiate between personality traits regarding what kind of similar scenarios were already experienced or perceived as threat (susceptibility). One use case for peer-reviewed SE scenarios are awareness trainings where the content is based on SE scenarios expressing the persuasion principles desired for the training. McBride et al. [MCW12] also used the FFM to propose adjusted Security Education, Training, and Awareness (SETA) programs. They also created shortened ‘scenarios’ in a survey about security policies and the FFM. Situational factors were added how they were perceived by the subjects [WCM11]. This may complement questionnaires with SE scenarios with questions about perceived ‘threat severity’, ‘self-efficacy’ or ‘threat vulnerability’.

Opportunistic SE applies the same deceptive techniques and principles, e.g., using scarcity of time (time pressure), with malicious intent. However, the deceptive context varies with respect to recent events. While the SE indicators can identify SE sufficiently, it challenges awareness campaigns and trainings if they are based on specific SE anecdotes. That is, the use of narrative psychology for transporting SE situations needs to advance. Besides supporting trainees with general self-questions regarding persuasion principles, a similar approach as the NSM must be developed for SE anecdotes expressing Opportunistic SE. Recent events could be implemented and tackle a changing SE context.

"Culture determines and limits strategy" [Sch04]. Previous research of the author covered security-aware organisational cultures [Ueb13b]. Besides the cultural background of national cultures of Section 2.7, "organisational cultures differ mainly at the level of practices" [Hof14b]. They can be learned and unlearned more easily than national cultures [Hof14b]. Future research may focus more on the organisational culture regarding SE. Bate [Bat97] discussed how organisational cultures may be changed in general. Van Niekerk and Solms [VS05] addressed two related dimensions of human factors (attitude and information security knowledge) by creating a holistic framework for an information security sub-culture. Furthermore, research of fostering an organisational cybersecurity learning culture may look into transactive memory (Section 2.1.1) and shadow security [KPS15].

Research about psychological safety [Edm99] as part of a cybersecurity culture can be relevant: although the original goal was to achieve high performance teams via team learning, the aspects of admitting and talking about mistakes (similar to the agile 'failebration') can improve the cybersecurity culture. One 'prime directive' of psychological safety consists of the assumption that employees act in the best interest of an organisation which resembles the SE indicators about targeted persons and that users (employees) are not the enemy [AS99].

In related research, the author with collaborators improved the eXtensible Abuse Report Format (X-ARF) [BÜ11; HUV12]. Metzner [Met17] and Uebelacker and Metzner [UM17] developed a procedure for enabling end-users to report suspicious e-mails directly from their e-mail client to an internal or external handler. End-users can submit e-mails where they are unsure about how to proceed. Security experts can then analyse these e-mails and report back with recommendations maintaining confidentiality as well as enabling them to become aware and mitigate new threats. With this feedback channel to the end-users, end-users may improve their own detection abilities as a 'human firewall'. Additionally, the confidentiality of the reporting assures that end-users do not feel to be in the pillory or face disciplinary measures. Furthermore, the incident handlers get a glimpse of what is happening in the field and may adapt existing awareness trainings accordingly. The main technical advantage of X-ARF is the human- *and* machine-readable format.

Appendices

List of Figures

List of Tables

C. Appendix Content

C.1 German Version of Human Mental Programming

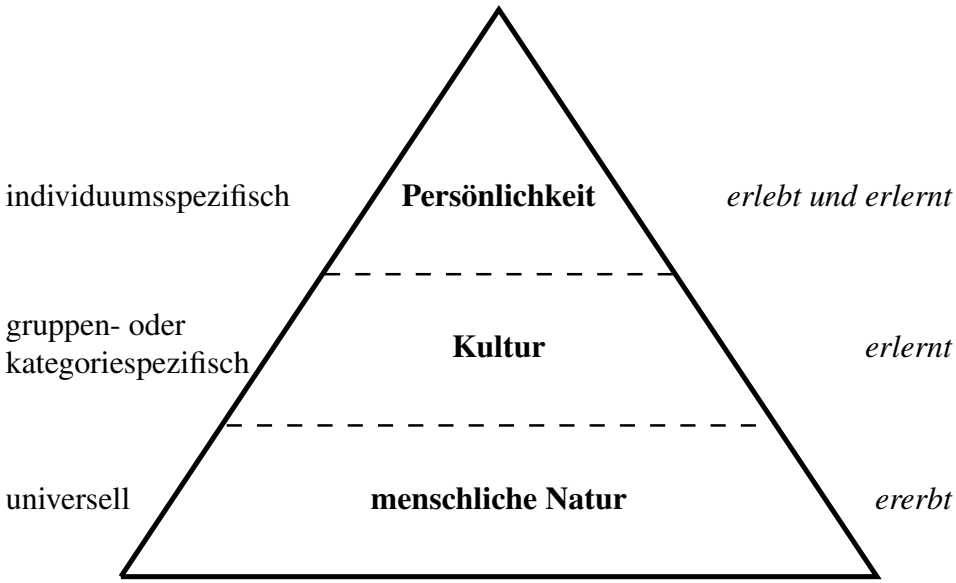


Figure C.1.1: German version of Figure 2.2 “three levels of uniqueness in human mental programming” [Hof01] (see also Übelacker [Übe02])

C.2 Original Diagram of the Human Error Classification

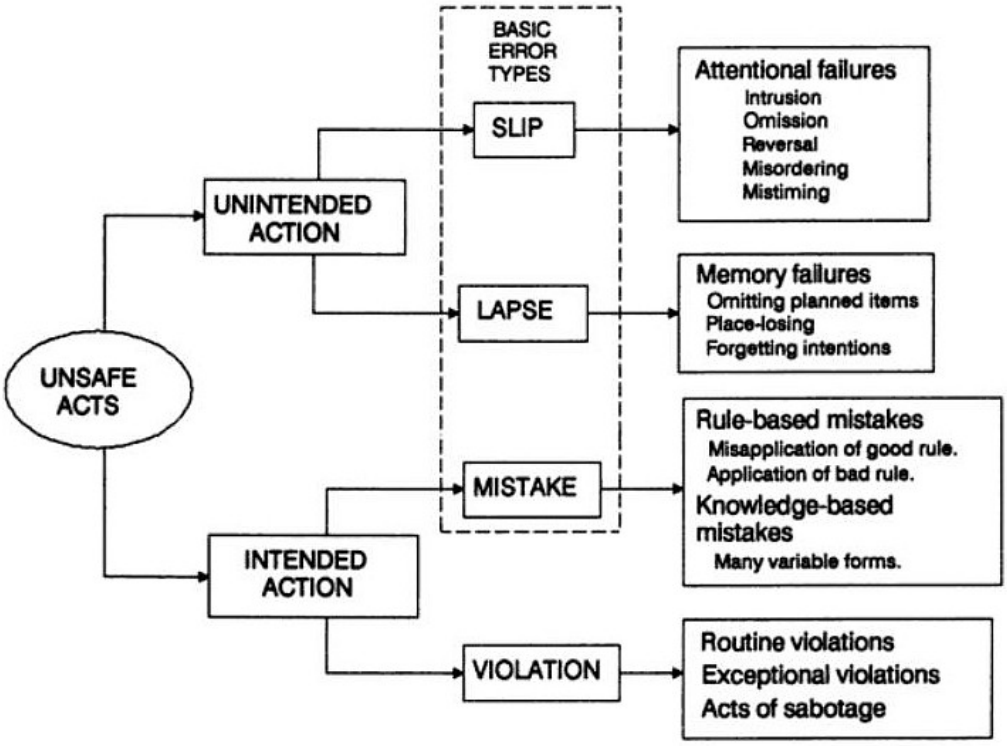


Figure C.2.2: Human error classification of unsafe acts as defined by Reason [Rea90] (original version of Figure 2.1 which was redrawn by Sofia Morais)

C.3 Original Diagram of the Persuasion Knowledge Model

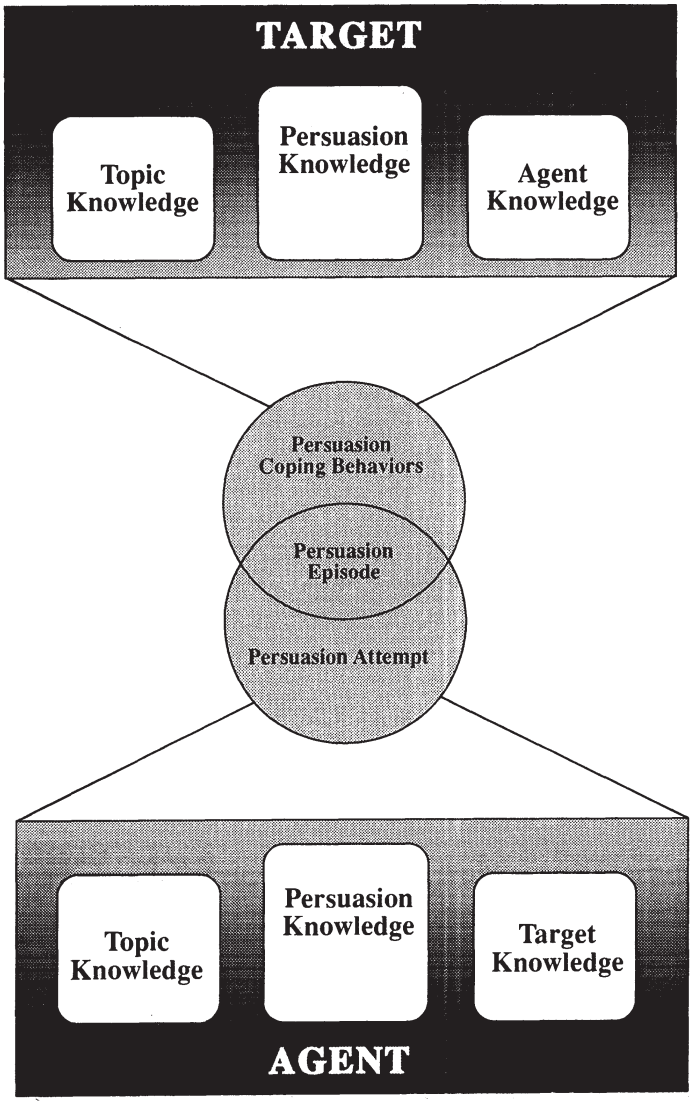


Figure C.3.3: The Persuasion Knowledge Model (PKM) by Friestad and Wright [FW94] (original version of Figure 4.1 which was redrawn by Sofia Morais)

C.4 Enkeltrick Anecdote 4.1 in German

Anecdote C.1 — Enkeltrick/Neffentrick per Telefon (Leh13). transcribed phone call from corresponding SpiegelTV report of Spiegel Online article [Leh13]: suspect Sylwia K. calls from Germany an elderly woman in Poland trying to persuade her to 'borrow' money.

Opfer:	Hallo?
Täterin:	Hallo Tante!
Opfer:	Ruft denn da die Jadwiga an?
Täterin:	Ja genau.
Opfer:	Ich erkenne Deine Stimme gar nicht.
Täterin:	Ich habe Halsschmerzen, weisst Du.
Opfer:	Du Arme!
Täterin:	Ich habe ein Riesenproblem seit heute morgen.
(. . .)	
Täterin:	Kannst Du mir bis morgen Geld leihen?
Opfer:	Wieviel?
Täterin:	20.000!
Opfer:	Ich leih Dir nichts — ich habe nichts.
Täterin:	Wieviel könntest Du mir denn leihen?
Opfer:	Ich weiss nicht. . . ich erkenne Deine Stimme gar nicht. . .
Täterin:	Na, weil ich in der Bank bin.

C.5 Social Engineering Poetry Slam @ 33C3 Wiki Content

Listen to social engineering attack stories from fellow hackers. Presented in a poetry slam style! A poetry slam can be a novel research approach to find stories of social engineering attacks, fictional or experienced. This slam will give us a new platform to discover and discuss social engineering.

Or present your social engineering experience or fictional story on how to deceive or manipulate people in the attacker's malicious interest. How did you get social engineered? Did you hear from a social engineering incident or know someone who managed to detect and mitigate it?

There are many definitions of social engineering in the wild, in short: a human interaction needs to be present to enable the attack, i.e., dumpster diving is not social engineering, it's just gathering pre-attack information, but (spear) phishing or scams like the "Enkeltrick" are. The community discusses some persuasion principles of why people succumb to these attacks. Presentations can base on the principles of Cialdini or Stajano/Wilson (links below).

How can I participate as a slammer?

If you want to slam the hack out of our minds, you have 5min to present your attack experience, 3min longer if the audience wants more. You can decide in which form and style, but you can use your voice only: no beamer for slides; no direct interaction with the audience (and do no harm), no other "tools" like for magicians.

Because we have to plan the event, we would like you to **register as slammer in advance**. Please, write an e-mail to: 33c3-slam@datapirate.de (PGP: 0xD42C10B0 E28B80DC) (Hint: You could use an extra e-mail address if you want to contact us anonymously.) Your e-mail must include the following information:

- your (stage) name or pseudonym
- title of your slam
- any anonymisation needed?

Please, appear 30 minutes before the event in our org room Hall C.1 (that is: 2016–12–29 20:30 CET/GMT+1).

Please, respect the privacy of others if you talk about sensitive information and do not incriminate yourself!

Please, pay attention that our event is in English, if you will need any assistance with translation, please let us know and give us some time in advance to help you with that.

If you have any further questions you can write us: 33c3-slam@datapirate.de or you can call us via DECT during the congress under: 7526 (SLAM). For example, if you will have any concerns or questions, if you'll need any help, if you'll have any feedback or ideas, if you would like to cancel your participation, if you would like to communicate to our "medium" (see below).

Anonymisation Methods

This event is going to be recorded with help from the VOC Angels. If this does not suit you in any way, tell us beforehand and may shut it down for your presentation.

We understand that not everyone likes to present hacking stories in public. That's why we offer different methods of anonymisation if wanted by the slammer. Slammers can choose between following options:

- Wear a mask
- Wear a mask and a protective suit
- Use our "medium" (another person) to present your story.
  - Please, pay attention that the entertainment value has to be given by content and style of the written story. Amusement value will not be added by the medium.

How will the winner be chosen?

Every slam has to have a winner. Due to the short time, we will conduct one round only without a typical final round. The voting will happen after each slammer has presented. Every slammer says again the (stage) name and the title of the contribution, afterwards voting starts. We will choose the winner by the results of the audience's votes.

The audience can assess (1–10, and hopefully audience is not social engineered;) the presented stories based on e.g., presentation style, novelty, creativity, feasibility of attack, feasibility of real-life experiments or scalability of attack. We will use a certain voting tool whereby everyone in the room can give one's vote.

Links

- Robert Cialdini and his six principles of influence: https://en.wikipedia.org/wiki/Robert_Cialdini
- The Real Hustle (BBC three series, Wilson et al.): https://en.wikipedia.org/wiki/The_Real_Hustle
- Stajano/Wilson: Understanding Scam Victims — Seven Principles for Systems Security: https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-754.pdf
- Kevin Mitnick: The Art of Deception: https://en.wikipedia.org/wiki/The_Art_of_Deception

FROM : +34961125606 - DATE : 20:01:16 14:15 - TO : +49[redacted] D 1/1

DIEGO ABOGADOS
"CONFIANZA & EFICIENCIA"

DIEGO ABOGADOS
Calle Fernando el Santo,
Madrid 28010. Spain
Tel: +34 663 476 579
Fax: +34 961 125 606
Email: willyferrarid@yahoo.es

REF JP/ESP/P01/AJ-787/01-16 ***Date: 20.01.2016***

Sehr geehrter Uebelacker
Fax No.[redacted]

Zunächst muss ich Sie bitten Ihr Vertrauen in dieser Transaktion, als absolut vertraulich und streng geheim zu halten. Obwohl ich weiß, dass eine Transaktion dieser Größenordnung jemand besorgt und beunruhigt, versichere ich Ihnen, dass alles gut sein wird am Ende des Tages.

Lassen Sie mich die Einführung an Sie richtig beginnen. Es mag Sie überraschen dass Sie diesen Brief von mir erhalten, da es keine früheren Schriftwechsel zwischen uns gab. ***Mein Name ist Willy Ferrari Diego*** *ein persönlicher Rechtsanwalt von* ***Engr. Lucas Uebelacker.***

Mein Zweck der Kontaktaufnahme mit Ihnen ist,Ihnen zu helfen dass Sie die Gelder die von meinen verstorbenen Klient hinterlassen wurde zu sichern, um zu vermeiden, dass die Gelder von der Bank beschlagnahmt oder fuer unbrauchbar erklärt wird. Dieser Fond wird auf 9,500,000,00 Euro (Nine Millionen fünf hundert tausend Euro) geschaetzt.

Die Bank hat mir Mittgeteilt, dass ich einen nächsten Angehörigen kontaktieren muss oder das Konto wird fuer unbrauchbar erklärt und den Fond in einen Bankschatz abgelegt. Bisher waren all meine Bemühungen jemandem zu finden gescheitert. Daher habe ich Sie kontaktiert, ich wollte Sie fragen, um Ihre Zustimmung, Sie auf der Bank als nächsten Angehörigen / Empfänger von meinem verstorbenen Kunden präsentieren zu duerfen, da Sie den gleichen Nachnamen haben, so dass der Erlös aus diesem Konto auf Ihr Konto gezahlt werden kann.

Alle rechtlichen Dokumente zur Sicherung Ihrer Ansprüche als mein Mandant/ nächsten Angehörigen werde ich Ihnen zu kommen lassen. Alles, was ich verlange, ist Ihre ehrliche Zusammenarbeit, damit wir diese Transaktion erfolgreich erreichen.

Ich möchte darauf hinweisen, dass ich 10% dieses Geldes, einer karitativen Organisationen gemeinsam spenden moechte, während die restlichen 90% zu gleichen Teilen zwischen uns aufgeteilt wird. Diese Transaktion ist ohne Risiko. Ich werde meine Position als Anwalt des Kunden, die erfolgreiche Ausführung der Transaktion garantieren. Wenn Sie interessiert sind, bitte kontaktieren Sie mich per ***Tel: +34663476579 E-Mail: willyferrarid@yahoo.es FAX: +34961125606***

Bitte Ich spreche kein Deutsch, ist es besser Sie mir eine E-Mail oder Fax zu senden.
Wenn Sie Englisch sprechen, dann können Sie mich anrufen.
Wenn Sie mich brauchen, um Sie auf Anruf, dann meine private Dolmetscher rufen Sie

Ich warte auf Ihre Antwort. Ich werde Ihnen dann mehr Details und informationen die Ihnen helfen werden diese Transaktion zu verstehen zu schicken.

Die beabsichtigte Transaktion wird unter einer legitimen Anordnung, die Sie und mich aus einer Verletzung des Gesetzes schützt. Allerdings, wenn dieser Geschäftsvorschlag Ihre moralische Ethik verstößt,dann bitte ich Sie meine aufrichtige Entschuldigung anzunehmen.

Wenn Sie dieses Ziel mit mir erreichen wollen, dann bitte kontaktieren Sie mich mit Ihrem Interesse für weitere Erläuterungen.

Mit freundlichen Grüßen,

Willy Ferrari Diego

Figure C.5.4: Facsimile scam received by the author's parents on 2016-01-20

MITTEILUNG IHRER POLIZEI

Warnung vor falschen Polizisten

DIE POLIZEI WARNT VOR BETRÜGERN, DIE SICH ALS POLIZEIBEAMTE AUSGEBEN

SO GEHEN DIE BETRÜGER VOR:

Am Telefon meldet sich eine Person bei Ihnen und gibt sich als ermittelnder Polizeibeamter aus. Die Person am Telefon sagt, dass bei einem festgenommenen Einbrecher ein Notizzettel mit Ihrem Namen und Ihrer Anschrift gefunden wurde. Nun wolle die Polizei weitere Straftaten verhindern und andere Komplizen festnehmen. Dazu sei Ihre Mitarbeit erforderlich.

Der Täter ist in diesem Gespräch sehr geschickt und wird versuchen, Informationen über Bankkonten, Wertanlagen, Schmuck, Vermögensverhältnisse und vorhandene Wertgegenstände zu erlangen.

Auch wird ggf. angeboten, natürlich zu Ihrem Schutz, vorbeizukommen und Geld, Schmuck sowie weitere Wertgegenstände abzuholen, um es für einen bestimmten Zeitraum bei der Polizei sicher aufzubewahren.

Mancher Täter ist sogar so dreist und erzählt, dass man Hinweise auf eine angebliche Mittäterschaft von Bankmitarbeitern habe. Ziel ist es, auch das Vertrauen in die Bank und deren Mitarbeiter zu erschüttern.

Es ist auch schon vorgekommen, dass während eines Telefonats ein angebliches Gespräch von Tätern vorgespielt wird, in dem deutlich Stimmen von Personen zu hören sind, die sich verabreden, Geld von Ihrem Konto abzuheben.

ACHTUNG: HIERBEI HANDELT ES SICH NICHT UM POLIZEIBEAMTE!

Seien Sie auch misstrauisch, wenn Sie während des Gesprächs mit dem Handy zur Bank gehen und Geld von Ihrem Konto abheben sollen. Da Sie ja immer noch mit dem Täter telefonieren, kann dieser sicher sein, dass Sie keine Person Ihres Vertrauens befragen oder über die Telefonnummer 110 die richtige Polizei anrufen.

ACHTUNG: DIE TELEFONNUMMER DER POLIZEI IN DER TELEFONANZEIGE!

Durch technische Manipulation können die Täter die echte Telefonnummer der Polizei (auch 110) im Display Ihres Telefons anzeigen.Dazu der ausdrückliche Hinweis: die Notrufnummer 110 wird nicht übertragen!

Präventionstipps: So können Sie sich schützen

- **Die „echte" Polizei fordert Sie niemals auf, Banküberweisungen oder Bargeldabhebungen durchzuführen, um Ermittlungen zu unterstützen.**
- **Seien Sie misstrauisch. Gesundes Misstrauen ist keine Unhöflichkeit. Sie haben immer Zeit für eine Rücksprache mit Angehörigen und Vertrauenspersonen!**
- **Lassen Sie sich nicht unter Druck setzen, auch nicht durch angeblich dringende Ermittlungen zu einem Einbruch in der Nähe!**
- **Polizisten in ziviler Kleidung weisen sich mit einem Dienstausweis aus und haben Verständnis dafür, dass man bei der Polizeizentrale nachfragt. Suchen Sie selber die Telefonnummer der Polizei heraus.**
- **Rufen Sie nie über die am Telefon angezeigte Nummer zurück - legen Sie auf! Verständigen Sie bei verdächtigen Vorfällen umgehend die 110!**

Noch ein Hinweis: In letzter Zeit wurden auch Fälle bekannt, in denen sich die Betrüger als Staatsanwälte, Bankmitarbeiter oder andere Amtspersonen ausgegeben haben.

Ihre Polizeidienststelle

Polizeiinspektion Harburg
Präventionsteam
Schützenstraße 17
21244 Buchholz/N.

Telefon: 04181 / 285 - 0

Impressum LKA Niedersachsen, Dezernat 32, 30169 Hannover

Figure C.6.5: Warning poster distributed by the State Office of Criminal Investigation (LKA) of Lower Saxony about criminals impersonating police officers (Anecdote 4.10)

D. Acronyms

BCP	Best Current Practice
BIP	Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung
BKA	Federal Criminal Police Office
BSI	German Federal Office for Information Security
CEO	Chief Executive Officer
CISO	Chief Information Security Officer
CISSP	Certified Information Systems Security Professional
CSIRT	Computer Security Incident Response Team
DDoS	Distributed Denial of Service
DT	Dark Triad
DEHSt	Deutsche Emissionshandelsstelle
ELM	Elaboration Likelihood Model
EU	Expected Utility
EEA	European Economic Area
FAIR	Findability, Accessibility, Interoperability, Reusability
FFM	Five-Factor Model
FMSE	Financially-Motivated Social Engineering
FOCA	Fingerprinting Organizations with Collected Archives
GDPR	General Data Protection Regulation
HCS	Hamburger Corona Soforthilfe
HUMINT	Human Intelligence
IDN	Internationalised Domain Name
ICT	Information and Communication Technology
IPIP	International Personality Item Pool
IRC	Internet Relay Chat
ISMS	Information Security Management System

ISP	Internet Service Provider
ITL	International Transaction Log
LKA	State Office of Criminal Investigation
MOM	Motive, Opportunity, and Means
NIST	National Institute of Standards and Technology
NSM	Natural Semantic Metalanguage
OPSEC	Operations Security
ORCID	Open Researcher & Contributor ID
PCS	Police Crime Statistics
PERSEC	Personnel Security
PGP	Pretty Good Privacy
PI	Predictive Index
PII	Personally Identifiable Information
PKM	Persuasion Knowledge Model
REB	Research Ethics Board
RFC	Request for Comments
RCT	Randomised Controlled Trial
RHUL	Royal Holloway University London
SE	Social Engineering
SEPF	Social Engineering Personality Framework
SEPS	Social Engineering Poetry Slam
SETA	Security Education, Training, and Awareness
SMTP	Simple Mail Transfer Protocol
SSH	Secure Shell
STS	Socio-Technical Systems
TARA	Threat Assessment and Remediation Analysis
TIPI	Ten Item Personality Inventory
TLD	Top Level Domain
TORE	TUHH Open Research
TRE$_S$PASS	Technology-supported Risk Estimation by Predictive Assessment of Socio-technical Security
TTP	Tactics, Techniques, and Procedures
TUHH	Hamburg University of Technology
UML	Unified Modelling Language
UNFCCC	United Nations Framework Convention on Climate Change
WEIRD	Western, Educated, Industrialised, Rich and Democratic
XACML	eXtensible Access Control Markup Language
X-ARF	eXtensible Abuse Report Format

E. Bibliography

[§249 StPO] Deutscher Bundestag. *§249 StPO: Furnishing of documentary evidence by reading out of documents; taking cognisance of wording of documents*. German Code of Criminal Procedure (Strafprozessordnung). July 2019. URL: https://www.gesetze-im-internet.de/englisch_stpo/englisch_stpo.html#p1777 (visited on 09/29/2020) (cited on page 119).

[ACL05] Nava Ashraf, Colin F Camerer, and George Loewenstein. "Adam Smith, Behavioral Economist". In: *Journal of Economic Perspectives* 19.3 (2005), pages 131–145. DOI: 10.1257/089533005774357897 (cited on pages 34, 35, 38, 90).

[Ada08] Scott Adams. *Das Dilbert-Prinzip: Die endgültige Wahrheit über Chefs, Konferenzen, Manager und andere Martyrien*. Redline Wirtschaft. Verlag Moderne Industrie, 2008. ISBN: 9783636015921 (cited on page 39).

[Ada15] Scott Adams. *Dilbert Comic 2015-05-18: I Hate Mondays More Than Garfield*. 2015. URL: https://dilbert.com/strip/2015-05-18 (visited on 04/28/2021) (cited on page 22).

[AG KS, 2850 Js 26209/14] Amtsgericht Kassel. *Hinweis auf Veränderung des rechtlichen Gesichtspunktes im Eröffnungsbeschluss; Schadensgleiche Vermögensgefährdung und Fälschung beweiserheblicher Daten bei internen Bestellungen unter falschen Namen; Verklammerung von Bestellung und Warenentgegennahme unter falschem Namen durch zugleich verwirklichtes*

Vermögensdelikt. Az 243 Ds - 2850 Js 26209/14, Urteil vom 28.05.2015, juris KORE215742015. May 2015. URL: https://oj.is/830415 (cited on page 121).

[AL07] Michael C. Ashton and Kibeom Lee. "Empirical, Theoretical, and Practical Advantages of the HEXACO Model of Personality Structure". In: *Personality and Social Psychology Review* 11.2 (2007). PMID: 18453460, pages 150–166. DOI: 10.1177/1088868306294907 (cited on page 44).

[AM06] Ross Anderson and Tyler Moore. "The Economics of Information Security". In: *Science* 314.5799 (Nov. 2006), pages 610–613. ISSN: 0036-8075. DOI: 10.1126/science.1130992 (cited on page 14).

[And08] Ross Anderson. *Security Engineering*. 2nd. John Wiley & Sons, 2008. ISBN: 978-0-470-06852-6 (cited on pages 11, 14, 49, 104).

[And37] Hans Christian Andersen. *Keiserens nye Klæder*. translated by Jean Hersholt and published in 1949. University of Southern Denmark, The Hans Christian Andersen Centret. 1837. URL: https://andersen.sdu.dk/vaerk/hersholt/TheEmperorsNewClothes_e.html (visited on 02/26/2021) (cited on pages 73, 74).

[AS15] George A Akerlof and Robert J Shiller. *Phishing for Phools: The Economics of Manipulation and Deception*. Princeton University Press, 2015 (cited on pages 21, 23, 84, 87–91, 93, 115).

[AS18] H. Aldawood and G. Skinner. "Educating and Raising Awareness on Cyber Security Social Engineering: A Literature Review". In: *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*. 2018, pages 62–68. DOI: 10.1109/TALE.2018.8615162 (cited on page 16).

[AS99] Anne Adams and Martina Angela Sasse. "Users are not the enemy". In: *Communications of the ACM* 42.12 (1999), pages 40–46. DOI: 10.1145/322796.322806 (cited on pages 12, 159).

[Bar14] Sean Barnum. "Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX)". In: *MITRE Corporation* (2014). Version 1.1, Revision 1, pages 1–22 (cited on pages 52, 71).

[Bat97] Paul Bate. *Cultural Change – Strategien zur Änderung der Unternehmenskultur*. Gerling Akademie Verlag, 1997 (cited on page 159).

[Bay. VGH, 16a D 12.2519] Bayerischer Verwaltungsgerichtshof. *Disziplinarrecht; Polizeihauptmeister (BesGr. A9); außerdienstliche und innerdienstliches Dienstvergehen; Beihilfe zum Computerbetrug; leichtfertige Geldwäsche; Nebentätigkeit ohne Genehmi-*

gung. Az 16a D 12.2519, Urteil vom 23.07.2014, juris JURE140015753, Berufungsverfahren von VG München, 09.10.2012, Az M 13 DK 12.3091. July 2014. URL: https://oj.is/739392 (cited on pages 72, 125).

[Ben+15] Zinaida Benenson, Gabriele Lenzini, Daniela Oliveira, Simon Parkin, and Sven Uebelacker. "Maybe Poor Johnny Really Cannot Encrypt – The Case for a Complexity Theory for Usable Security". In: *New Security Paradigms Workshop (NSPW)*. University of Twente, The Netherlands: Association for Computing Machinery (ACM), 2015. ISBN: 978-1-4503-3754-0. DOI: 10.1145/2841113.2841120. URL: https://www.nspw.org/papers/2015/nspw2015-benenson.pdf (cited on pages 12, 19, 31, 37, 38, 85).

[BfV16] Bundesamt für Verfassungsschutz. *Verfassungsschutzbericht 2016*. BMI17006. Bundesministerium des Innern, 2017. URL: https://www.verfassungsschutz.de/download/vsbericht-2016.pdf (cited on pages 61, 79, 80, 82).

[Bha07] B. Bhagyavati. "Social Engineering". In: *Cyber Warfare and Cyber Terrorism*. Edited by Lech J. Janczewski and Andrew M. Colarik. Idea Group Inc (IGI), 2007. Chapter XXIII, pages 182–190. ISBN: 9781591409922. DOI: 10.4018/978-1-59140-991-5 (cited on page 52).

[Bis+10] Matt Bishop, Sophie Engle, Deborah A Frincke, Carrie Gates, Frank L Greitzer, Sean Peisert, and Sean Whalen. "A Risk Management Approach to the "Insider Threat"". In: *Insider Threats in Cyber Security*. Edited by Christian W Probst, Jeffrey Hunker, Dieter Gollmann, and Matt Bishop. Volume 49. Boston, MA: Springer US, 2010, pages 115–137. ISBN: 978-1-4419-7132-6. DOI: 10.1007/978-1-4419-7133-3_6 (cited on page 69).

[Bis03] Matt Bishop. *Computer Security: Art and Science*. Library of Congress Number QA76.9.A25 B56 2002. Addison Wesley Professional, 2003. ISBN: 0-201-44099-7 (cited on pages 11, 12).

[BMV10] M. Bezuidenhout, F. Mouton, and H.S. Venter. "Social Engineering Attack Detection Model: SEADM". In: *Information Security for South Africa (ISSA), 2010*. 2010, pages 1–8. DOI: 10.1109/ISSA.2010.5588500 (cited on page 56).

[Bra96] Allan M. Brandt. "Recruiting Women Smokers: The Engineering of Consent". In: *American Medical Women's Association* (1996). ISSN: 0098-8421. URL: http://nrs.harvard.edu/urn-3:HUL.InstRepos:3372908 (cited on pages 75, 76).

[Bru90] Jerome S Bruner. *Acts of meaning*. Volume 3. Harvard University Press, 1990 (cited on page 23).

[BS06] Carlo Batini and Monica Scannapieca. *Data Quality. Concepts, Methodologies and Techniques*. Springer, 2006. ISBN: 978-3-540-33172-8. DOI: 10.1007/3-540-33173-5 (cited on pages 21, 122, 123).

[BSI13] BSI. *IT-Grundschutz Catalogues*. Volume 13. Germany's Federal Office for Information Security (BSI), 2013. URL: https://www.bsi.bund.de/EN/Topics/ITGrundschutz/ITGrundschutzCatalogues/itgrundschutzcatalogues_node.html (cited on pages 24, 116, 157).

[BSI15] BSI. *IT-Grundschutz Catalogues*. Volume 15. draft. Germany's Federal Office for Information Security (BSI), 2015. URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/International/GSK_15_EL_EN_Draft.html (cited on pages 23, 24, 63, 116).

[BSI19a] BSI für Bürger. *Schadsoftware Emotet: Wie kann man sich schützen?* 2019. URL: https://www.bsi-fuer-buerger.de/BSIFB/DE/Service/Aktuell/Informationen/Artikel/emotet.html (visited on 01/31/2021) (cited on page 80).

[BSI19b] Bundesamt für Sicherheit in der Informationstechnik (BSI). *IT-Grundschutz-Kompendium*. Bundesanzeiger Verlag GmbH, 2019. ISBN: 978-3-8462-0906-6 (cited on page 63).

[BSW08] Adam Beautement, M. Angela Sasse, and Mike Wonham. "The Compliance Budget: Managing Security Behaviour in Organisations". In: *Proceedings of the 2008 New Security Paradigms Workshop*. Edited by Matt Bishop, Christian W Probst, Angelos D. Keromytis, and Anil Somayaji. NSPW '08. Lake Tahoe, California, USA: Association for Computing Machinery, 2008, pages 47–58. ISBN: 9781605583419. DOI: 10.1145/1595676.1595684 (cited on page 37).

[BÜ11] Matthias Bräck and Sven Übelacker. *X-ARF: Network Abuse Reporting*. talk given by Matthias Bräck and Sven Übelacker at the Easterhegg. Chaos Computer Club e.V., 2011. URL: https://eh11.easterhegg.eu/fahrplan/events/4293.de.html (cited on pages 19, 159).

[Buj02] L. McM. Bujold. *Diplomatic Immunity*. Volume 14. Baen Books, 2002 (cited on page 89).

[Bul+15] Jan-Willem Hendrik Bullée, Lorena Montoya, Wolter Pieters, Marianne Junger, and Pieter H. Hartel. "The Persuasion and Security Awareness Experiment: Reducing the Success of Social Engineering Attacks". English. In: *Journal of Experimental Criminology* (2015), pages 1–19. ISSN:

1573-3750. DOI: 10.1007/s11292-014-9222-7 (cited on pages 16, 71, 87–90, 93).

[Bul+18] Jan-Willem Hendrik Bullée, Lorena Montoya, Wolter Pieters, Marianne Junger, and Pieter Hartel. "On the anatomy of social engineering attacks – A literature-based dissection of successful attacks". In: *Journal of Investigative Psychology and Offender Profiling* 15.1 (2018), pages 20–45. DOI: 10.1002/jip.1482 (cited on pages 87, 89, 90).

[BV07] Roy F. Baumeister and Kathleen D. Vohs. *Encyclopedia of Social Psychology*. 2007. DOI: 10.4135/9781412956253 (cited on pages 33, 43, 44, 75, 84).

[Cam18a] Cambridge Dictionary. *anecdote*. 2018. URL: https://dictionary.cambridge.org/dictionary/english/anecdote (visited on 08/30/2018) (cited on page 21).

[Cam18b] Cambridge Dictionary. *social engineering*. 2018. URL: https://dictionary.cambridge.org/dictionary/english/social-engineering (visited on 11/09/2018) (cited on page 49).

[Cam20] Cambridge Dictionary. *gullibility*. 2020. URL: https://dictionary.cambridge.org/dictionary/english/gullibility (visited on 12/21/2020) (cited on pages 71, 73).

[Cam21a] Cambridge Dictionary. *credulity*. 2021. URL: https://dictionary.cambridge.org/dictionary/english/credulity (visited on 06/28/2021) (cited on page 73).

[Cam21b] Cambridge Dictionary. *impersonation*. 2021. URL: https://dictionary.cambridge.org/dictionary/english/impersonation (visited on 04/02/2021) (cited on page 78).

[Cam21c] Cambridge Dictionary. *imposture*. 2021. URL: https://dictionary.cambridge.org/dictionary/english/imposture (visited on 04/02/2021) (cited on page 78).

[Cam99] Colin Camerer. "Behavioral economics: Reunifying psychology and economics". In: *Proceedings of the National Academy of Sciences* 96.19 (1999), pages 10575–10577. ISSN: 0027-8424. DOI: 10.1073/pnas.96.19.10575 (cited on pages 34, 35).

[Cap+14] Deanna D Caputo, Shari Lawrence Pfleeger, Jesse D Freeman, and M Eric Johnson. "Going spear phishing: Exploring embedded training and awareness". In: *IEEE Security & Privacy* 12.1 (Jan. 2014), pages 28–38. ISSN: 1540-7993. DOI: 10.1109/MSP.2013.106 (cited on pages 16, 82).

[Car07] Nancy Cartwright. "Are RCTs the Gold Standard?" In: *BioSocieties* 2.1 (Mar. 2007), pages 11–20. ISSN: 1745-

8560. DOI: 10.1017/S1745855207005029 (cited on page 102).

[CBG13] Gokul Chittaranjan, Jan Blom, and Daniel Gatica-Perez. "Mining Large-Scale Smartphone Data for Personality Studies". In: *Personal and Ubiquitous Computing* 17.3 (2013), pages 433–450. URL: http://infoscience.epfl.ch/record/192373/files/Chittaranjan%5C_PUC%5C_2012.pdf (cited on page 43).

[Cia07] Robert B. Cialdini. *Influence: The Psychology of Persuasion*. HarperCollins, 2007 (cited on pages 36, 84, 87, 88, 94, 118, 121, 142).

[Cro02] Michele L. Crossley. "Introducing Narrative Psychology". In: *Narrative, Memory and Life Transitions*. Edited by Christine Horrocks, Kate Milnes, Brian Roberts, and David Robinson. Huddersfield: University of Huddersfield, Apr. 2002, pages 1–13. URL: http://eprints.hud.ac.uk/id/eprint/5127/ (cited on page 23).

[D2.5.1] Margaret Ford, Dieter Gollmann, Claude Heath, Mariëlle Stoelinga, Sven Uebelacker, and Ahmed Seid Yesuf. *TRE$_S$PASS Information Testing and Degradation Tools*. Edited by Sven Uebelacker. Deliverable D2.5.1. 2016 (cited on pages 110, 111).

[D4.1.2] The TRE$_S$PASS Project, D4.1.2. *Final requirements for visualisation processes and tools*. Deliverable D4.1.2. 2015 (cited on page 129).

[D4.2.2] The TRE$_S$PASS Project, D4.2.2. *Methods for visualization of information security risks*. Deliverable D4.2.2. 2016 (cited on page 129).

[D4.3.3] The TRE$_S$PASS Project, D4.3.3. *Visualizations of socio-technical dimensions of information-security risks*. Deliverable D4.3.3. 2016 (cited on page 129).

[Daw06] Richard Dawkins. *The selfish gene*. Oxford University Press, 2006, 360 p. ISBN: 978-0-19-929115-1. URL: http://books.google.de/books?isbn=0199291144 (cited on page 41).

[Dim+10] Trajce Dimkov, André van Cleeff, Wolter Pieters, and Pieter Hartel. "Two Methodologies for Physical Penetration Testing Using Social Engineering". In: *Proceedings of the 26th Annual Computer Security Applications Conference*. ACSAC '10. Austin, Texas, USA: ACM, 2010, pages 399–408. ISBN: 978-1-4503-0133-6. DOI: 10.1145/1920261.1920319 (cited on pages 54, 104, 106–109, 147).

[DiM+19] Mariano Di Martino, Pieter Robyns, Winnie Weyts, Peter Quax, Wim Lamotte, Ken Andries, and Flanders Make. "Personal Information Leakage by Abusing the GDPR 'Right of Access'". In: *Fifteenth Symposium on Usable*

Privacy and Security (SOUPS 2019). Santa Clara, CA: USENIX Association, Aug. 2019 (cited on page 79).

[Dim12] Trajce Dimkov. "Alignment of Organizational Security Policies – Theory and Practice". PhD thesis. Enschede: University of Twente, Feb. 2012. DOI: 10.3990/1.9789036533317 (cited on page 42).

[Dod+15] Peter Sheridan Dodds, Eric M. Clark, Suma Desu, Morgan R. Frank, Andrew J. Reagan, Jake Ryland Williams, Lewis Mitchell, Kameron Decker Harris, Isabel M. Kloumann, James P. Bagrow, Karine Megerdoomian, Matthew T. McMahon, Brian F. Tivnan, and Christopher M. Danforth. "Human Language Reveals a Universal Positivity Bias". In: *Proceedings of the National Academy of Sciences* 112.8 (2015), pages 2389–2394. ISSN: 0027-8424. DOI: 10.1073/pnas.1411678112 (cited on page 34).

[Döl84] Dieter Dölling. "Probleme der Aktenanalyse in der Kriminologie". In: *Methodologische Probleme in der kriminologischen Forschungspraxis*. Edited by Helmut Kury. Interdisziplinäre Beiträge zur kriminologischen Forschung. C. Heymann, 1984, pages 265–286. ISBN: 9783452199041. URL: https://books.google.de/books?id=waOaAAAACAAJ (cited on pages 119, 120).

[DsiN20] Deutschland sicher im Netz. *DsiN-Sicherheitsindex 2020*. Technical report. Deutschland sicher im Netz e.V., June 2020. URL: https://www.sicher-im-netz.de/dsin-sicherheitsindex-2020 (cited on page 71).

[DZA12] A. Darwish, A.E. Zarka, and F. Aloul. "Towards Understanding Phishing Victims' Profile". In: *Computer Systems and Industrial Informatics (ICCSII), 2012 International Conference on*. 2012, pages 1–5. DOI: 10.1109/ICCSII.2012.6454454 (cited on pages 70, 74).

[Edm99] Amy Edmondson. "Psychological Safety and Learning Behavior in Work Teams". In: *Administrative Science Quarterly* 44.2 (1999), pages 350–383. DOI: 10.2307/2666999 (cited on page 159).

[Ege+13] Serge Egelman, Andreas Sotirakopoulos, Ildar Muslukhov, Konstantin Beznosov, and Cormac Herley. "Does My Password Go Up to Eleven?: The Impact of Password Meters on Password Selection". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI 2013. Paris, France: ACM, 2013, pages 2379–2388. ISBN: 978-1-4503-1899-0. DOI: 10.1145/2470654.2481329 (cited on pages 104, 117).

[EH16] Svea Eckert and Peter Hornung. *Osnabrück: Sechseinhalb Jahre Haft für "Phishing"*. July 2016. URL: https://www.ndr.de/nachrichten/niedersachsen/

osnabrueck_emsland/Osnabrueck-Sechseinhalb-Jahre-Haft-fuer-Phishing,postbank302.html (visited on 05/04/2017) (cited on page 123).

[Fag07] Isabelle J. Fagnot. "Behavioral Information Security". In: *Cyber Warfare and Cyber Terrorism*. Edited by Lech J. Janczewski and Andrew M. Colarik. Idea Group Inc (IGI), 2007. Chapter XXV, pages 199–205. ISBN: 9781591409922. DOI: 10.4018/978-1-59140-991-5 (cited on pages 30, 60).

[Fei15] Dror G. Feitelson. "From Repeatability to Reproducibility and Corroboration". In: *ACM SIGOPS Operating Systems Review* 49.1 (Jan. 2015), pages 3–11. ISSN: 0163-5980. DOI: 10.1145/2723872.2723875 (cited on pages 103, 108, 109).

[FJ07] Peter Finn and Markus Jakobsson. "Designing Ethical Phishing Experiments". In: *IEEE Technology and Society Magazine* 26.1 (2007), pages 46–58. ISSN: 0278-0097. DOI: 10.1109/MTAS.2007.335565 (cited on pages 107, 117).

[FRS05] Ivan Flechais, Jens Riegelsberger, and M. Angela Sasse. "Divide and Conquer: The Role of Trust and Assurance in the Design of Secure Socio-technical Systems". In: *Proceedings of the 2005 Workshop on New Security Paradigms*. NSPW 2005. Lake Arrowhead, California: ACM, 2005, pages 33–41. ISBN: 1-59593-317-4. DOI: 10.1145/1146269.1146280 (cited on pages 14, 95).

[FW94] Marian Friestad and Peter Wright. "The Persuasion Knowledge Model: How People Cope with Persuasion Attempts". In: *Journal of Consumer Research* 21.1 (June 1994), pages 1–31. ISSN: 0093-5301. DOI: 10.1086/209380 (cited on pages 20, 55, 67, 68, 104, 125, 169).

[GC02] Rosanna E. Guadagno and Robert B. Cialdini. "Online Persuasion – An Examination of Gender Differences in Computer-Mediated Interpersonal Influence". In: *Group Dynamics: Theory, Research, and Practice* 6.1 (2002), pages 38–51. DOI: 10.1037/1089-2699.6.1.38 (cited on page 74).

[GC05] Rosanna E. Guadagno and Rpbert B. Cialdini. "Online Persuasion and Compliance: Social Influence on the Internet and Beyond". In: *The Social Net: Human Behavior in Cyberspace* (2005), pages 91–113 (cited on pages 37, 74, 87).

[GGF17] Paul A Grassi, Michael E Garcia, and James L Fenton. "NIST Secial Publication 800-63-3: Digital Identity Guidelines". In: *NIST Special Publication* (2017). DOI: 10.6028/NIST.SP.800-63-3 (cited on pages 13, 20, 54, 60, 69).

[GMK19] Florian Gondesen, Matthias Marx, and Ann-Christine Kycler. "A shoulder-surfing resistant image-based authenti-

cation scheme with a brain-computer interface". In: *International Conference on Cyberworlds, CW 2019*. 2019, pages 336–343. ISBN: 978-172812297-7. DOI: 10.1109/CW.2019.00061 (cited on page 52).

[Gol06] Dieter Gollmann. "Why Trust is Bad for Security". In: *Electronic Notes in Theoretical Computer Science* 157.3 (2006), pages 3–9. ISSN: 1571-0661. DOI: 10.1016/j.entcs.2005.09.044 (cited on page 95).

[Gra02] David Gragg. "A Multi-Level Defense against Social Engineering". In: *SANS Reading Room* 13 (Dec. 2002). URL: https://www.sans.org/reading-room/whitepapers/engineering/multi-level-defense-social-engineering-920 (cited on pages 62, 85).

[Gre09] Stephen Greenspan. *Annals of Gullibility: Why We Get Duped and How to Avoid It*. Non-Series. Praeger, 2009. ISBN: 9780313362163 (cited on page 73).

[GRS03] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. "A Very Brief Measure of the Big-Five Personality Domains". In: *Journal of Research in Personality* 37.6 (2003), pages 504–528. ISSN: 0092-6566. DOI: 10.1016/S0092-6566(03)00046-1 (cited on page 46).

[GS00] T. Grandison and M. Sloman. "A survey of trust in internet applications". In: *IEEE Communications Surveys Tutorials* 3.4 (2000), pages 2–16. DOI: 10.1109/COMST.2000.5340804 (cited on page 95).

[GW04] Cliff Goddard and Anna Wierzbicka. "Cultural scripts: What are they and what are they good for?" In: *Intercultural Pragmatics* 1.2 (2004), pages 153–166. DOI: 10.1515/iprg.2004.1.2.153 (cited on pages 22, 41, 158).

[Had10] C. Hadnagy. *Social Engineering: The Art of Human Hacking*. Wiley, 2010 (cited on pages 61, 62).

[Har13] Shon Harris. *CISSP All-in-One Exam Guide, Sixth Edition*. 6th. McGraw-Hill Education Group, 2013. ISBN: 9780071781718 (cited on pages 16, 52, 59, 119, 120).

[HB16] Martin Haase and Kai Biermann. *BigBrotherAward 2016 Neusprech: Datenreichtum*. 2016. URL: https://bigbrotherawards.de/2016/neusprech-datenreichtum (visited on 05/07/2021) (cited on page 14).

[HCH14] Claude P. Heath, Lizzie Coles-Kemp, and Peter A. Hall. "Logical Lego? Co-constructed Perspectives on Service Design". In: *Proceedings of the NordDesign Conference 2014 in Melbourne*. 2014 (cited on page 129).

[Her09] Cormac Herley. "So Long, and No Thanks for the Externalities: The Rational Rejection of Security Advice by Users". In: *Proceedings of the 2009 Workshop on New Se-*

curity Paradigms Workshop. NSPW '09. Oxford, United Kingdom: Association for Computing Machinery, 2009, pages 133–144. ISBN: 9781605588452. DOI: 10.1145/1719030.1719050 (cited on page 11).

[HH18] Alice Hutchings and Thomas J Holt. "Interviewing Cybercrime Offenders". In: *Journal of Qualitative Criminal Justice & Criminology* (Oct. 2018), pages 75–94. DOI: 10.21428/88de04a1.1fdab531 (cited on pages 114, 120, 140, 141).

[HHN10] Joseph Henrich, Steven J Heine, and Ara Norenzayan. "The weirdest people in the world?" In: *Behavioral and brain sciences* 33.2-3 (Apr. 2010), pages 61–135. DOI: 10.1017/S0140525X0999152X (cited on pages 109, 117, 118).

[HKB12] J. B. Hirsh, S. K. Kang, and G. V. Bodenhausen. "Personalized Persuasion: Tailoring Persuasive Appeals to Recipients' Personality Traits". In: *Psychological Science* 23.6 (2012), pages 578–581. DOI: 10.1177/0956797611436349 (cited on pages 45, 94–96).

[HM04] Geert Hofstede and Robert R. McCrae. "Personality and Culture Revisited: Linking Traits and Dimensions of Culture". In: *Cross-Cultural Research* 38.1 (2004), pages 52–88. DOI: 10.1177/1069397103259443 (cited on page 30).

[Hoe15] Thomas Hoeren. *Internet Law Scriptorium*. Apr. 2015. URL: https://www.uni-muenster.de/Jura.itm/hoeren/lehre/materialien (cited on page 121).

[Hof01] Geert Hofstede. *Lokales Denken, globales Handeln*. 2. Beck-Wirtschaftsberater im dtv, 2001. ISBN: 3-423-50807-8 (cited on pages 30, 31, 41, 43, 67, 70, 94, 167).

[Hof14a] Hofstede Center. *National Cultural Dimensions*. 2014. URL: http://geert-hofstede.com/national-culture.html (visited on 04/27/2014) (cited on pages 41–43, 118).

[Hof14b] Hofstede Center. *Organisational Culture & Change Management*. 2014. URL: http://geert-hofstede.com/organisational-culture.html (visited on 04/27/2014) (cited on page 159).

[Hol07] Tilmann Holst. "Automatic Correlation, Rating and Analyzing of Heterogeneous Network and Incident Data". diploma thesis. University of Hamburg, 2007. URL: http://books.google.de/books?id=IcDitgAACAAJ (cited on pages 19, 20).

[Hop+18] Sabrina Hoppe, Tobias Loetscher, Stephanie A. Morey, and Andreas Bulling. "Eye Movements During Everyday Behavior Predict Personality Traits". In: *Frontiers in Human Neuroscience* 12 (2018), page 105. ISSN: 1662-5161. DOI: 10.3389/fnhum.2018.00105 (cited on pages 43, 46).

[Hos13] Hossiep, Rüdiger et al. *Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP)*. 2013. URL: http://www.testentwicklung.de/testverfahren/BIP/index.html.de (visited on 05/07/2013) (cited on page 44).

[HP04] R. Hossiep and M. Paschen. "Rezension der 2. Auflage des Bochumer Inventars zur berufsbezogenen Persönlichkeitsbeschreibung (BIP)". In: *Zeitschrift für Arbeits-und Organisationspsychologie A&O* 48.2 (2004), 79–86 (cited on page 44).

[HS14] Eric Hatleback and Jonathan M Spring. "Exploring a mechanistic approach to experimentation in computing". In: *Philosophy & Technology* 27.3 (2014), pages 441–459. DOI: 10.1007/s13347-014-0164-9 (cited on pages 117, 118).

[HSS06] Ute R Hülsheger, Elke Specht, and Frank M Spinath. "Validität des BIP und des NEO-PI-R". In: *Zeitschrift für Arbeits- und Organisationspsychologie A&O* 50.3 (2006), 135–147. URL: http://www.psycontent.com/content/835n015520161172/ (cited on page 47).

[Hub+09] M. Huber, S. Kowalski, M. Nohlberg, and S. Tjoa. "Towards Automating Social Engineering Using Social Networking Sites". In: *Computational Science and Engineering, 2009. CSE '09. International Conference on*. Volume 3. 2009, pages 117–124. DOI: 10.1109/CSE.2009.205 (cited on page 92).

[HUV12] Tilmann Haak, Sven Uebelacker, and Torsten Voss. *X-ARF – End-to-End Security with S-MIME and PGP-MIME*. slides of a talk given by Tilmann Haak at the 37th meeting of the TF-CSIRT. TERENA, 2012. URL: http://www.terena.org/activities/tf-csirt/meeting37/voss-x-arf.pdf (cited on pages 19, 159).

[Ira+11] Danesh Irani, Marco Balduzzi, Davide Balzarotti, Engin Kirda, and Calton Pu. "Reverse Social Engineering Attacks in Online Social Networks". In: *Detection of Intrusions and Malware, and Vulnerability Assessment*. Edited by Thorsten Holz and Herbert Bos. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pages 55–74. ISBN: 978-3-642-22424-9. DOI: 10.1007/978-3-642-22424-9_4 (cited on page 62).

[Iva+15] Marieta Georgieva Ivanova, Christian W Probst, René Rydhof Hansen, and Florian Kammüller. "Attack Tree Generation by Policy Invalidation". In: *Information Security Theory and Practice*. Edited by Raja Naeem Akram and Sushil Jajodia. Springer International Publishing, 2015, pages 249–259. ISBN: 978-3-319-24018-3. DOI: 10.1007/978-3-319-24018-3_16 (cited on page 116).

[Iye10] Sheena Iyengar. *The Art of Choosing*. Little, Brown Book Group Limited, 2010. ISBN: 978-0-349-12142-0 (cited on page 118).

[JG13] Monique Janneck and Sascha R. Guczka. "The Resigned, the Confident, and the Humble: A Typology of Computer-Related Attribution Styles". In: *Human Factors in Computing and Informatics*. Edited by Andreas Holzinger, Martina Ziefle, Martin Hitz, and Matjaž Debevc. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pages 373–390. ISBN: 978-3-642-39062-3. DOI: 10.1007/978-3-642-39062-3_24 (cited on page 32).

[JMO17] M. Junger, A. L. Montoya Morales, and F. J. Overink. "Priming and Warnings are not Effective to Prevent Social Engineering Attacks". In: *Computers in Human Behavior* 66 (2017), pages 75–87. ISSN: 0747-5632. DOI: 10.1016/j.chb.2016.09.012 (cited on page 16).

[JS06] Iris Junglas and Christiane Spitzmüller. "Personality Traits and Privacy Perceptions: An Empirical Study in the Context of Location-Based Services". In: *Mobile Business, 2006. ICMB '06. International Conference on*. 2006. DOI: 10.1109/ICMB.2006.40 (cited on page 95).

[JS17] Debora Jeske and Paul van Schaik. "Familiarity with internet threats: beyond awareness". In: *Computers & Security* (2017) (cited on pages 60, 70, 79).

[Kah11] Daniel Kahneman. *Thinking, Fast and Slow*. Penguin Books, 2011 (cited on pages 34, 36, 40, 85, 87).

[KD12] Erin E. Kenneally and David Dittrich. *The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research*. Technical report. US Department of Homeland Security, Aug. 2012. DOI: 10.2139/ssrn.2445102 (cited on pages 107, 108).

[KD99] Justin Kruger and David Dunning. "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments". In: *Journal of Personality and Social Psychology* 77.6 (1999), page 1121 (cited on page 39).

[KLD07] Stefan Kiltz, Andreas Lang, and Jana Dittmann. "Taxonomy for Computer Security Incidents". In: *Cyber Warfare and Cyber Terrorism*. Edited by Lech J. Janczewski and Andrew M. Colarik. Idea Group Inc (IGI), 2007. Chapter XLVIII, pages 412–417. ISBN: 9781591409922. DOI: 10.4018/978-1-59140-991-5 (cited on pages 59, 60).

[KLG06] Bradley Kirkman, Kevin Lowe, and Cristina Gibson. "A Quarter Century of Culture's Consequences: A Review of Empirical Research Incorporating Hofstede's Cultural Values Framework". In: *Journal of International Business*

Studies 37 (Feb. 2006), pages 285–320. DOI: 10.1057/palgrave.jibs.8400202 (cited on pages 41, 42).

[Klo06] John Klossner. *Human Nature: Data Security and Dave*. 2006. URL: http://www.jklossner.com/humannature (visited on 09/14/2020) (cited on page 13).

[Kor+13] Barbara Kordy, Piotr Kordy, Sjouke Mauw, and Patrick Schweitzer. "ADTool: Security Analysis with Attack-Defense Trees (Extended Version)". In: *arXiv preprint arXiv:1305.6829* (2013). URL: http://arxiv.org/pdf/1305.6829.pdf (cited on page 115).

[Kor+16] Oliver Korn, Martin Hauptmeier, Anke Frieling, Frank Steinhoff, Dawid Bekalarczyk, and Ulrich Schweiger. *Metakognition – ein Schlüssel für mentale Leistungsfähigkeit und psychische Gesundheit*. 2016. URL: https://www.addisca.org/images/veroeffentlichungen/Metakognitive-Techniken-in-der-Gesundheitspraevention-2016.pdf (visited on 11/10/2020) (cited on page 39).

[KPS15] Iacovos Kirlappos, Simon Parkin, and M. Angela Sasse. ""Shadow Security" as a Tool for the Learning Organization". In: *SIGCAS Comput. Soc.* 45.1 (Feb. 2015), pages 29–37. ISSN: 0095-2737. DOI: 10.1145/2738210.2738216 (cited on pages 11, 37, 159).

[Kre18] Brian Krebs. *Sextortion Scam Uses Recipient's Hacked Passwords*. 2018. URL: https://krebsonsecurity.com/2018/07/sextortion-scam-uses-recipients-hacked-passwords/ (visited on 05/11/2021) (cited on page 82).

[KSR17] G. Kunjadić, M. Savković, and S. Radović. "Social Engineering Attack Method on ICT Systems Using USB Stick". In: *Proceedings of Sinteza 2017 - International Scientific Conference on Information Technology and Data Related Research*. 2017, pages 35–39. DOI: 10.15308/Sinteza-2017-35-39 (cited on page 76).

[Kuh02] Markus G Kuhn. "Optical time-domain eavesdropping risks of CRT displays". In: *Proceedings 2002 IEEE Symposium on Security and Privacy*. IEEE. 2002, pages 3–18. ISBN: 0-7695-1543-6 (cited on page 52).

[Kus13] David Kushner. *The Real Story of Stuxnet – How Kaspersky Lab tracked down the malware that stymied Iran's nuclear-fuel enrichment program*. 2013. URL: http://spectrum.ieee.org/telecom/security/the-real-story-of-stuxnet (visited on 03/23/2014) (cited on page 77).

[Las+13] E. Lastdrager, L. Montoya, P. Hartel, and M. Junger. "Applying the Lost-Letter Technique to Assess IT Risk Behaviour". In: *Socio-Technical Aspects in Security and Trust*

(STAST), 2013 Third Workshop on. June 2013, pages 2–9. DOI: 10.1109/STAST.2013.15 (cited on page 76).

[Leh13] Roman Lehberger. *MEK-Einsatz gegen Enkeltrickbetrüger*. Apr. 2013. URL: http://www.spiegel.de/panorama/justiz/a-892830.html (visited on 12/21/2020) (cited on pages 72, 170).

[Leu17] Eric Rutger Leukfeldt. *Research Agenda: the Human Factor in Cybercrime and Cybersecurity*. Eleven International Publishing, 2017. ISBN: 978-94-6236-753-1 (cited on pages 13, 14, 17, 96, 105, 157).

[LG BN, 3 O 236/06] Landgericht Bonn. *Stornobuchung bei einer Kontogutschrift infolge einer mit ausgespähter PIN und TAN*. Az 3 O 236/06, Urteil vom 29.12.2006, juris KORE222342007. Dec. 2006. URL: https://oj.is/121461 (cited on pages 121, 122).

[LG K, 3 O 390/13] Landgericht Köln. *Online-Banking: Haftung des Zahlungsdienstleisters für nicht autorisierte Zahlungsvorgänge nach sog. Phishing*. Az 3 O 390/13, Urteil vom 26.08.2014, juris JURE140016284. Aug. 2014. URL: https://oj.is/733266 (cited on page 121).

[Lin+19] Tian Lin, Daniel E. Capecci, Donovan M. Ellis, Harold A. Rocha, Sandeep Dommaraju, Daniela S. Oliveira, and Natalie C. Ebner. "Susceptibility to Spear-Phishing Emails: Effects of Internet User Demographics and Email Content". In: *ACM Trans. Comput.-Hum. Interact.* 26.5 (July 2019), 32:1–32:28. ISSN: 1073-0516. DOI: 10.1145/3336141 (cited on pages 70, 74).

[LJ16] Rutger Leukfeldt and Jurjen Jansen. "Cyber Criminal Networks and Money Mules: An Analysis of Low-Tech and High-Tech Fraud Attacks in the Netherlands". In: *International Journal of Cyber Criminology* 9.2 (June 2016), pages 173–184. DOI: 10.5281/zenodo.56210 (cited on page 125).

[LKA18] State Office of Criminal Investigation of Lower Saxony. *Bewerbungsmail mit Schadsoftware im Anhang*. 2018. URL: https://www.polizei-praevention.de/aktuelles/bewerbungsmail-mit-schadsoftware-im-anhang.html (visited on 08/30/2018) (cited on page 116).

[Lon08] Johnny Long. *No Tech Hacking: a Guide to Social Engineering, Dumpster Diving, and Shoulder Surfing*. Edited by Kevin David Mitnick. Safari Books Online. Oxford: Elsevier Science, 2008, page 285. ISBN: 9780080558752 (cited on page 117).

[Luc01] J.A. Lucy. "Sapir–Whorf Hypothesis". In: *International Encyclopedia of the Social & Behavioral Sciences*. Edited by Neil J. Smelser and Paul B. Baltes. Oxford: Pergamon,

2001, pages 13486–13490. ISBN: 978-0-08-043076-8. DOI: 10.1016/B0-08-043076-7/03042-4 (cited on page 22).

[MAP18] David Modic, Ross Anderson, and Jussi Palomäki. "We Will Make You Like Our Research: The Development of a Susceptibility-to-Persuasion Scale". In: *PLOS ONE* 13.3 (Mar. 2018), pages 1–21. DOI: 10.1371/journal.pone.0194119 (cited on page 85).

[Mar74] Odo Marquard. "Inkompetenzkompensationskompetenz? Über Kompetenz und Inkompetenz der Philosophie". In: *Philosophisches Jahrbuch* 81.2 (1974), pages 341–349. URL: https://philosophisches-jahrbuch.de/ (cited on page 39).

[Mas15] Ronald Robert Mason. *Loose Wheel Nut Indicator*. US Patent D733,511 S. July 2015. URL: https://patents.google.com/patent/USD733511S1/en (cited on page 36).

[May14] Philipp Mayring. *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Klagenfurt, 2014, page 143 (cited on page 120).

[MCW12] Maranda McBride, Lemuria Carter, and Merrill Warkentin. *One Size Doesn't Fit All: Cybersecurity Training Should Be Customized*. Technical report. Institute for Homeland Security Solutions, 2012. URL: http://sites.duke.edu/ihss/files/2011/12/CyberSecurity_2page-summary_mcbride-2012.pdf (cited on page 158).

[Med14] Betty Medsger. *The Burglary: The Discovery of J. Edgar Hoover's Secret FBI*. 2014. ISBN: 978-0307962959 (cited on page 15).

[Met17] Adrian Metzner. "End user reporting of suspicious E-mails using X-ARF". master thesis. Hamburg University of Technology, 2017. DOI: 10.15480/882.1396 (cited on pages 19, 46, 159).

[MGM14] Nicole L. Muscanell, Rosanna E. Guadagno, and Shannon Murphy. "Weapons of Influence Misused: A Social Influence Analysis of Why People Fall Prey to Internet Scams". In: *Social and Personality Psychology Compass* 8.7 (2014), pages 388–396. DOI: 10.1111/spc3.12115 (cited on pages 87–91).

[Mil56] George A. Miller. "The Magical Number Seven, Plus or Minus Two". In: *Psychological Review* 63.2 (Mar. 1956), pages 81–97 (cited on pages 37, 85).

[Mil65] Stanley Milgram. "Some Conditions of Obedience and Disobedience to Authority". In: *Human relations* 18.1 (1965), pages 57–76. DOI: 10.1177/001872676501800105 (cited on pages 14, 93).

[MJ92] R. R. McCrae and O. P. John. "An Introduction to the Five-Factor Model and Its Applications". In: *Journal of Personality* 60.2 (1992), pages 175–215. ISSN: 1467-6494. DOI: 10.1111/j.1467-6494.1992.tb00970.x (cited on pages 43, 45, 46).

[ML12] David Modic and Stephen E.G. Lea. "How neurotic are scam victims, really? the big five and internet scams". In: *Law & Humanities eJournal* (Sept. 2012). DOI: 10.2139/ssrn.2448130 (cited on pages 94–96).

[ML13] David Modic and Stephen E.G. Lea. "Scam Compliance and the Psychology of Persuasion". In: *Journal of Applied Social Psychology* (2013). DOI: 10.2139/ssrn.2364464 (cited on page 85).

[MLV15] Francois Mouton, Louise Leenen, and HS Venter. "Social engineering attack detection model: SEADMv2". In: *Cyberworlds (CW), 2015 International Conference on*. IEEE. 2015, pages 216–223 (cited on page 56).

[MLV16] Francois Mouton, Louise Leenen, and H.S. Venter. "Social Engineering Attack Examples, Templates and Scenarios". In: *Computers & Security* 59 (2016), pages 186–209. ISSN: 0167-4048. DOI: 10.1016/j.cose.2016.03.004 (cited on pages 51, 56, 59, 115).

[MMV13] F. Mouton, M. M. Malan, and H. S. Venter. "Social engineering from a normative ethics perspective". In: *2013 Information Security for South Africa*. IEEE. Aug. 2013, pages 1–8. DOI: 10.1109/ISSA.2013.6641064 (cited on pages 56, 106, 107).

[Mon+13] Yves-Alexandre de Montjoye, Jordi Quoidbach, Florent Robic, and Alex (Sandy) Pentland. "Predicting Personality Using Novel Mobile Phone-Based Metrics". In: *Social Computing, Behavioral-Cultural Modeling and Prediction*. Edited by Ariel M. Greenberg, William G. Kennedy, and Nathan D. Bos. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pages 48–55. ISBN: 978-3-642-37210-0. DOI: 10.1007/978-3-642-37210-0_6 (cited on page 43).

[Mou+14a] Francois Mouton, Louise Leenen, Mercia M. Malan, and H.S. Venter. "Towards an Ontological Model Defining the Social Engineering Domain". English. In: *ICT and Society*. Edited by Kai Kimppa, Diane Whitehouse, Tiina Kuusela, and Jackie Phahlamohlaka. Volume 431. IFIP Advances in Information and Communication Technology. Springer Berlin Heidelberg, 2014, pages 266–279. ISBN: 978-3-662-44207-4. DOI: 10.1007/978-3-662-44208-1_22 (cited on pages 52, 56–59, 64, 67, 74, 75, 84, 91, 107, 115, 155).

[Mou+14b] Francois Mouton, Mercia M Malan, Louise Leenen, and Hein S Venter. "Social engineering attack framework". In:

Information Security for South Africa (ISSA), 2014. IEEE. 2014, pages 1–9 (cited on page 56).

[MS02] Kevin D. Mitnick and William L. Simon. *The Art of Deception: Controlling the Human Element of Security*. Wiley, 2002. ISBN: 0471237124 (cited on pages 20, 22, 54, 62, 63, 70, 75, 79, 115, 142).

[MS05] Kevin D. Mitnick and William L. Simon. *The Art of Intrusion: The Real Stories Behind the Exploits of Hackers, Intruders & Deceivers*. Wiley, 2005. ISBN: 0764569597 (cited on pages 62, 115).

[MS17a] Don A. Moore and Derek Schatz. "The three faces of overconfidence". In: *Social and Personality Psychology Compass* 11.8 (2017). DOI: 10.1111/spc3.12331 (cited on page 38).

[MS17b] Steven J. Murdoch and M. Angela Sasse. *Should you really phish your own employees?* Aug. 2017. URL: https://tech.newstatesman.com/business/phishing-employees (visited on 09/03/2019) (cited on page 106).

[NATO19] NATO Standardization Office. *AAP-06 – NATO Glossary of terms and definitions (Edition 2019)*. 2019. URL: https://nso.nato.int/nso/ZPUBLIC/_BRANCHINFO/TERMINOLOGY_PUBLIC/NON-CLASSIFIED%20NATO%20GLOSSARIES/AAP-6.PDF (visited on 10/06/2020) (cited on page 52).

[NDR20a] NDR. *Betrugsversuch: Hamburger Corona-Hilfen gestoppt*. Apr. 2020. URL: https://www.ndr.de/betrugsversuch100.html (visited on 12/30/2020) (cited on page 79).

[NDR20b] NDR. *Polizei warnt vor schamlosen Corona-Betrügern*. Mar. 2020. URL: https://www.ndr.de/trickbetrueger134.html (visited on 12/30/2020) (cited on page 79).

[NDR20c] 3 Aktuell NDR 90. *Von Rentner ausgetrickst: Falsche Polizisten vor Gericht*. May 2020. URL: https://www.ndr.de/nachrichten/hamburg/Von-Rentner-ausgetrickst-Falsche-Polizisten-vor-Gericht,betrueger148.html (visited on 06/16/2020) (cited on page 63).

[Nid+15] Michael Nidd, Marieta Georgieva Ivanova, Christian W Probst, and Axel Tanner. "Tool-Based Risk Assessment of Cloud Infrastructures as Socio-Technical Systems". In: *The Cloud Security Ecosystem*. Edited by Ryan Ko and Kim-Kwang Raymond Choo. Boston: Syngress, 2015. Chapter 22, pages 495–517. ISBN: 978-0-12-801595-7. DOI: 10.1016/B978-0-12-801595-7.00022-7 (cited on page 130).

[OAS10] OASIS. *eXtensible Access Control Markup Language (XACML) Version 3.0*. 2010. URL: https://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-cs-01-en.pdf (visited on 09/14/2021) (cited on page 20).

[Ole+17] Victoria C. Oleynick, Colin G. DeYoung, Elizabeth Hyde, Scott Barry Kaufman, Roger E. Beaty, and Paul J. Silvia. "Openness/Intellect". In: *The Cambridge Handbook of Creativity and Personality Research*. Edited by Gregory J. Feist, Roni Reiter-Palmon, and James C.Editors Kaufman. Cambridge Handbooks in Psychology. Cambridge University Press, 2017, pages 9–27. DOI: 10.1017/9781316228036.002 (cited on page 36).

[OLG D, 2 Ss 437/97] Oberlandesgericht Düsseldorf. *Absprachewidrige Geldabhebung vom Bankautomaten mit Scheckkarte und Geheimnummer als Computerbetrug*. Az 2 Ss 437/97 - 123/97 II, Urteil vom 05.01.1998, juris KORE502629800. Jan. 1998 (cited on page 121).

[OLG HAM, 31 U 31/15] Oberlandesgericht Hamm. *Haftung des Zahlungsdienstleisters für nicht autorisierte Zahlungsvorgänge: Einwand der unzulässigen Rechtsausübung bei nicht autorisiertem Zahlungsvorgang durch sog. Pishing*. Az 31 U 31/15, Urteil vom 16.03.2015, juris KORE522842015. Mar. 2015. URL: https://oj.is/2148640 (cited on page 122).

[Ols19] Isaac Olsen. *Addicted to Fortnite? Montreal law firm says video game company should pay up*. 2019. URL: https://www.cbc.ca/news/canada/montreal/fortnite-lawsuit-calex-l%C3%A9gal-montreal-1.5308625 (visited on 10/07/2019) (cited on pages 55, 75).

[Orw95] George Orwell. *1984*. 16. Verlag Ullstein GmbH, 1995. ISBN: 3548225624 (cited on page 14).

[Ovid] P. Ovidius Naso. *Epistulae Heroidum – Phyllis Demophoonti*. 25. URL: https://www.thelatinlibrary.com/ovid/ovid.her2.shtml (visited on 09/23/2019) (cited on page 106).

[PBC09] James L Parrish Jr, Janet L Bailey, and James F Courtney. "A Personality Based Model for Determining Susceptibility to Phishing Attacks". In: *Little Rock: University of Arkansas* (2009). URL: http://www.swdsi.org/swdsi2009/Papers/9J05.pdf (cited on pages 74, 94, 96).

[PC86] Richard E Petty and John T Cacioppo. "The Elaboration Likelihood Model of Persuasion". In: *Advances in Experimental Social Psychology*. Edited by Leonard Berkowitz. Volume 19. Academic Press, 1986, pages 123–205. DOI: 10.1016/S0065-2601(08)60214-2 (cited on page 87).

[PDP13] Wolter Pieters, Trajce Dimkov, and Dusko Pavlovic. "Security Policy Alignment: A Formal Approach". In: *IEEE*

Systems Journal 7.2 (2013), pages 275–287. DOI: 10.1109/JSYST.2012.2221933 (cited on pages 52, 76, 77).

[Pfl+10] S. L. Pfleeger, J. B. Predd, J. Hunker, and C. Bulford. "Insiders Behaving Badly: Addressing Bad Actors and Their Actions". In: *IEEE Transactions on Information Forensics and Security* 5.1 (2010), pages 169–179. DOI: 10.1109/TIFS.2009.2039591 (cited on page 69).

[PH13] Christian W Probst and René Rydhof Hansen. "Reachability-based Impact as a Measure for Insiderness". In: *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications* 4.4 (Dec. 2013), 38–48. URL: http://eprints.eemcs.utwente.nl/24198/01/jowua-v4n4-3.pdf (cited on pages 69, 70).

[PH69] Laurence J Peter and Raymond Hull. *The Peter Principle – Why Things Always Go Wrong*. HarperCollins e-books, 1969 (cited on page 39).

[Pha15] Ngoc-Minh Michal Pham. "Court Rulings as Evidence for Social Engineering Research". bachelor thesis. Hamburg University of Technology, 2015. DOI: 10.15480/882.1271 (cited on pages 19, 121).

[PR78] United States National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. *The Belmont report: ethical principles and guidelines for the protection of human subjects of research*. Volume 2. Department of Health, Education, and Welfare, National Commission for the, 1978, pages 1–18 (cited on page 107).

[Qui13] Susanne Quiel. "Social Engineering in the Context of Cialdini's Psychology of Persuasion and Personality Traits". bachelor thesis. Hamburg University of Technology, 2013. DOI: 10.15480/882.1124 (cited on pages 19, 94).

[RB96] Bruce Rind and Prashant Bordia. "Effect on Restaurant Tipping of Male and Female Servers Drawing a Happy, Smiling Face on the Backs of Customers' Checks". In: *Journal of Applied Social Psychology* 26.3 (Feb. 1996), pages 218–225. DOI: 10.1111/j.1559-1816.1996.tb01847.x (cited on pages 74, 84, 117).

[Rea08] James Reason. *The Human Contribution: Unsafe Acts, Accidents and Heroic Recoveries*. Ashgate Burlington, VT, 2008. ISBN: 9780754674023 (cited on page 29).

[Rea90] James Reason. *Human Error*. Cambridge University Press, 1990 (cited on pages 29, 30, 54, 168).

[RFC2119] Scott O. Bradner. *Key words for use in RFCs to Indicate Requirement Levels*. RFC 2119. Mar. 1997. DOI: 10.17487/RFC2119 (cited on page 64).

[RFC4949] R. Shirey. *Internet Security Glossary, Version 2*. RFC 4949. Aug. 2007. DOI: 10.17487/RFC4949 (cited on page 49).

[RFC5890] J. Klensin. *Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework*. RFC 5890. Aug. 2010. DOI: 10.17487/RFC5890 (cited on page 82).

[RFC8174] Barry Leiba. *Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words*. RFC 8174. May 2017. DOI: 10.17487/RFC8174 (cited on page 64).

[RFC821] Jonathan B Postel. *Simple Mail Transfer Protocol*. RFC 821. Aug. 1982. DOI: 10.17487/RFC821 (cited on page 79).

[RM10] Thomas Ryan and G Mauch. "Getting in bed with Robin Sage". In: *Black Hat Conference*. 2010. URL: http://media.blackhat.com/bh-us-10/whitepapers/Ryan/BlackHat-USA-2010-Ryan-Getting-In-Bed-With-Robin-Sage-v1.0.pdf (cited on pages 41, 89, 91).

[Rol02] Jean-Pierre Rolland. "The Cross-Cultural Generalizability of the Five-Factor Model of Personality". English. In: *The Five-Factor Model of Personality Across Cultures*. Edited by Robert R. McCrae and Jüri Allik. International and Cultural Psychology Series. Springer US, 2002, pages 7–28. ISBN: 978-0-306-47355-5. DOI: 10.1007/978-1-4615-0763-5_2 (cited on page 41).

[RWB12] Emilee Rader, Rick Wash, and Brandon Brooks. "Stories as Informal Lessons about Security". In: *Proceedings of the Eighth Symposium on Usable Privacy and Security*. SOUPS '12. Washington, D.C.: Association for Computing Machinery, 2012. ISBN: 9781450315326. DOI: 10.1145/2335356.2335364 (cited on page 24).

[Sag+02] Brad J Sagarin, Robert B Cialdini, William E Rice, Sherman B Serna, et al. "Dispelling the Illusion of Invulnerability: The Motivations and Mechanisms of Resistance to Persuasion". In: *Journal of Personality and Social Psychology* 83.3 (2002), pages 526–541. DOI: 10.1037/0022-3514.83.3.526 (cited on page 87).

[SBW01] Martina Angela Sasse, Sacha Brostoff, and Dirk Weirich. "Transforming the 'Weakest Link' — a Human/Computer Interaction Approach to Usable and Effective Security". In: *BT Technology Journal* 19.3 (2001), pages 122–131. DOI: 10.1023/A:1011902718709 (cited on page 12).

[SC09] The Science Council. *Definition of Science*. 2009. URL: https://sciencecouncil.org/about-science/our-definition-of-science/ (visited on 11/09/2021) (cited on page 103).

[Sch+14] Guillaume Schaff, Carlo Harpes, Matthieu Aubigny, Marianne Junger, and Romain Martin. "RISK-DET: ICT Secu-

rity Awareness Aspect Combining Education and Cognitive Sciences". In: *ICCGI 2014, The Ninth International Multi-Conference on Computing in the Global Information Technology*. 2014, pages 51–53. URL: http://www.thinkmind.org/download.php?articleid=iccgi_2014_3_10_10035 (cited on page 16).

[Sch+17] Paul van Schaik, Debora Jeske, Joseph Onibokun, Lynne Coventry, Jurjen Jansen, and Petko Kusev. "Risk Perceptions of Cyber-Security and Precautionary Behaviour". In: *Computers in Human Behavior* 75 (2017), pages 547–559. ISSN: 0747-5632. DOI: 10.1016/j.chb.2017.05.038 (cited on pages 70, 79).

[Sch00] Bruce Schneier. *Secrets & Lies – Digital Security in a Networked World*. Wiley Computer Publishing, 2000. ISBN: 0-471-25311-1 (cited on pages 12, 13, 79).

[Sch04] Edgar H. Schein. *Organizational Culture and Leadership*. 3. Jossey-Bass, 2004. ISBN: 0-7879-6845-5 (cited on page 159).

[Sch08a] Jamison W. Scheeres. "Establishing the Human Firewall: Reducing an Individual's Vulnerability to Social Engineering Attacks". thesis. Air Force Institute of Technology, 2008 (cited on page 87).

[Sch08b] Bruce Schneier. "The Psychology of Security". In: *Progress in Cryptology – AFRICACRYPT 2008*. Edited by S. Vaudenay. Volume 5023. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2008, 50–79. ISBN: 978-3-540-68159-5. DOI: 10.1007/978-3-540-68164-9_5 (cited on pages 11, 12, 14, 24, 29, 34–36, 39–41, 71).

[Sch14] Bruce Schneier. *1971 Social Engineering Attack*. 2014. URL: https://www.schneier.com/blog/archives/2014/02/1971%5C_social%5C_eng.html (visited on 05/11/2021) (cited on page 15).

[Sch99] Daniel L. Schacter. "The Seven Sins of Memory: Insights from Psychology and Cognitive NeuroScience". In: *American Psychologist* 54.3 (1999), pages 182–203. DOI: 10.1037/0003-066X.54.3.182 (cited on page 31).

[Ser15] IBM Security – Managed Security Services. *IBM 2015 Cyber Security Intelligence Index*. Technical report. IBM Corporation, May 2015 (cited on page 69).

[SFS15] Bahareh Shojaie, Hannes Federrath, and Iman Saberi. "The Effects of Cultural Dimensions on the Development of an ISMS Based on the ISO 27001". In: *2015 10th International Conference on Availability, Reliability and Security*. 2015, pages 159–167. ISBN: 978-1-4673-6590-1. DOI: 10.1109/ARES.2015.25 (cited on page 43).

[Sho18] Bahareh Shojaie. "Implementation of information security management systems based on the ISOIEC 27001 standard in different cultures". PhD thesis. University of Hamburg, 2018 (cited on pages 41, 42).

[Shr+06] J. Shropshire, M. Warkentin, A.C. Johnston, and M.B. Schmidt. "Personality and IT Security: An Application of the Five-Factor Model". In: *Proceedings of the Americas Conference on Information Systems*. 2006, pages 3443–3449 (cited on pages 44–46).

[SI18] Jonathan M Spring and Phyllis Illari. "Building General Knowledge of Mechanisms in Information Security". In: *Philosophy & Technology* 32.4 (2018), pages 627–659. ISSN: 2210-5441. DOI: 10.1007/s13347-018-0329-z (cited on pages 14, 105, 124, 126, 141, 158).

[Smi20] Graham Smith. *Don't Invest in Bitcoin Code, Bitcoin Doubler or Bitcoin Trader – They Are All Scams*. 2020. URL: https://news.bitcoin.com/bitcoin-code-doubler-autotrader-scam/ (visited on 09/14/2021) (cited on page 55).

[SMP17] Jonathan M Spring, Tyler Moore, and David Pym. "Practicing a Science of Security: A Philosophy of Science Perspective". In: *New Security Paradigms Workshop*. Santa Cruz, CA, USA, Oct. 2, 2017. DOI: 10.1145/3171533.3171540 (cited on pages 21, 101–105, 109, 110, 120).

[Spo62] Heinrich Spoerl. *Die Feuerzangenbowle – Eine Lausbüberei in der Kleinstadt*. 2584. Bertelsmann Lesering, 1962 (cited on page 77).

[Spr08] Springer. "Social Engineering". In: *Encyclopedia of Genetics, Genomics, Proteomics and Informatics*. Dordrecht: Springer Netherlands, 2008, pages 1838–1838. ISBN: 978-1-4020-6754-9. DOI: 10.1007/978-1-4020-6754-9_15826 (cited on pages 49, 54).

[SS16] Roland Schilling and Frieder Steinmetz. *USB devices phoning home*. Technical report. Hamburg University of Technology, Feb. 2016. DOI: 10.15480/882.1279 (cited on page 77).

[Str09] Strand Consult. *How will psychologists describe the iPhone syndrome in the future?* 2009. URL: https://strandconsult.dk/how-will-psychologists-describe-the-iphone-syndrome-in-the-future/ (visited on 11/16/2020) (cited on page 40).

[SU19] Christian Sillaber and Sven Uebelacker. "Phishing in höchstgerichtlicher Judikatur". In: *Innsbrucker Beiträge zur Rechtstatsachenforschung*. Edited by Michael Ganner and Caroline Voithofer. Volume 10. innsbruck university press, 2019, pages 145–153. ISBN: 978-3-903187-65-8. URL:

http://d-nb.info/1189476916 (cited on pages 19, 120, 121, 124).

[Sun18] Jonny Sun. *Twitter: Tweet about Pavlovian Conditioning*. 2018. URL: https://twitter.com/jonnysun/status/997800789736394752 (visited on 11/09/2020) (cited on page 14).

[SW09] Frank Stajano and Paul Wilson. *Understanding Scam Victims: Seven Principles for Systems Security*. Technical report 754. Technical Report UCAM-CL-TR-754. Computer Laboratory, University of Cambridge, UK, 2009. URL: http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-754.pdf (cited on pages 19, 75, 85, 86, 88, 121, 142).

[Sys15] IBM Security Systems. *IBM X-Force Threat Intelligence Quarterly, 2Q 2015*. Technical report. IBM Corporation, June 2015 (cited on page 69).

[TAB30] Christoph Bogenstahl. *Dark Patterns – Mechanismen (be)trügerischen Internetdesigns*. Technical report. Themenkurzprofil Nr. 30. Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB), Nov. 2019. URL: https://www.tab-beim-bundestag.de/de/publikationen/themenprofile/Themenkurzprofil-030.html (cited on pages 55, 75).

[Thi16] Harold Thimbleby. "Human Error in Safety-Critical Programming". In: *Developing Safe Systems, Proceedings of the 24nd Safety-Critical Systems Symposium*. Edited by Mike Parsons and Tom Anderson. Brighton, UK: Center for Software Reliability, 2016, pages 183–202. ISBN: 9781519420077 (cited on pages 29, 30, 40, 41, 91).

[Tho+17] Daniel R. Thomas, Sergio Pastrana, Alice Hutchings, Richard Clayton, and Alastair R. Beresford. "Ethical issues in research using datasets of illicit origin". In: *Proceedings of the Internet Measurement Conference (IMC)*. ACM. London, UK, Nov. 2017, pages 445–462. DOI: 10.1145/3131365.3131389 (cited on page 107).

[Tho99] Suzanne C. Thompson. "Illusions of Control: How We Overestimate Our Personal Influence". In: *Current Directions in Psychological Science* 8.6 (1999), pages 187–190. DOI: 10.1111/1467-8721.00044 (cited on pages 40, 71).

[TJ15] Sophie E Tait and Debora Jeske. "Hello Stranger!: Trust and Self-Disclosure Effects on Online Information Sharing". In: *International Journal of Cyber Behavior, Psychology and Learning (IJCBPL)* 5.1 (2015), pages 42–55. DOI: 10.4018/ijcbpl.2015010104 (cited on page 95).

[TK74] Amos Tversky and Daniel Kahneman. "Judgment under Uncertainty: Heuristics and Biases". In: *Science* 185.4157 (1974), pages 1124–1131. ISSN: 0036-8075. DOI: 10.

1126/science.185.4157.1124 (cited on pages 24, 31, 35, 36, 42).

[Übe02] Sven Übelacker. "IT-Sicherheit, Unternehmenskulturen und wirtschaftsbedrohende Kriminalität". diploma thesis. University of Ulm, 2002. DOI: 10.15480/882.1117 (cited on pages 18, 31, 167).

[Ueb13a] Sven Uebelacker. *Security-Aware Organisational Cultures – A Starting Point for Mitigating Socio-Technical Risks*. Slides of a talk presented at the RiskKom workshop at the GI Conference INFORMATIK 2013, September 16th 2013, University of Koblenz-Landau, Koblenz, Germany. Hamburg University of Technology, Sept. 2013 (cited on page 16).

[Ueb13b] Sven Uebelacker. "Security-Aware Organisational Cultures as a Starting Point for Mitigating Socio-Technical Risks". In: *INFORMATIK 2013*. Edited by Gesellschaft für Informatik e.V. (GI). Volume 220. Lecture Notes in Informatics (LNI). RiskKom workshop. University of Koblenz-Landau, Koblenz, Germany: Matthias Horbach, Sept. 2013, pages 2046–2057. ISBN: 978-3-88579-614-5. DOI: 10.13140/2.1.1389.6000 (cited on pages 18, 31, 41–43, 46, 159).

[UM17] Sven Uebelacker and Adrian Metzner. "Privacy-Respecting End-User Reporting of Suspicious E-Mails Using X-ARF". Poster presented by Adrian Metzner at 10th International Conference on IT Security Incident Management & IT Forensics (IMF2017). 2017 (cited on pages 19, 159).

[UQ14] Sven Uebelacker and Susanne Quiel. "The Social Engineering Personality Framework". In: *2014 Workshop on Socio-Technical Aspects in Security and Trust, STAST 2014, Vienna, Austria, July 18, 2014*. Edited by Giampaolo Bella and Gabriele Lenzini. IEEE. Vienna University of Technology, Vienna, Austria: IEEE Computer Society, July 2014, pages 24–30. DOI: 10.1109/STAST.2014.12 (cited on pages 19, 45, 46, 61, 70, 88–90, 94–97, 109).

[UR16] Sven Uebelacker and Youssef Rebahi-Gilbert. *Social Engineering Poetry Slam @ 33C3*. video recording on 33rd Chaos Communication Congress, Hamburg. Dec. 2016. URL: https://doi.org/10.15480/336.2707 (cited on pages 139, 141, 142, 145–148).

[Ver19] Verizon RISK Team. *2019 Data Breach Investigations Report*. Accessed: 2019-05-08. 2019 (cited on pages 54, 59, 126).

[Vet20] Alexander Vetterl. "Honeypots in the Age of Universal Attacks and the Internet of Things". PhD thesis. University of Cambridge, Computer Laboratory, Feb. 2020. URL:

https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-944.html (cited on page 45).

[VG B, 10 K 333.10] Verwaltungsgericht Berlin. *Ersatz von durch Phishing-Angriff abhanden gekommener Emissionszertifikate*. German. Az 10 K 333.10, Urteil vom 13.09.2013, juris JURE130016503. Sept. 2013. URL: https://oj.is/651966 (cited on pages 15, 124, 126, 127).

[VG M, M 13 DK 12.3091] Verwaltungsgericht München. *Geldwäsche; Betrug (bes. schwerer Fall); Nebentätigkeit*. Az M 13 DK 12.3091, Urteil vom 09.10.2012. Oct. 2012. URL: https://oj.is/561517 (cited on pages 72, 125).

[VL17] Steve GA Van de Weijer and E Rutger Leukfeldt. "Big Five Personality Traits of Cybercrime Victims". In: *Cyberpsychology, Behavior, and Social Networking* 20.7 (2017), pages 407–412. DOI: 10.1089/cyber.2017.0028 (cited on page 46).

[VM02] John Viega and Gary R McGraw. *Building Secure Software: How to Avoid Security Problems the Right Way*. Addison-Wesley, 2002 (cited on page 12).

[Voy94] Voyager. "Janitor Privileges". In: *The Hacker Quarterly*. Edited by Emmanuel Goldstein. Volume 11. 4. 4q94_36, Eric Gordon Corley. 2600, 1994, page 36. URL: https://store.2600.com/collections/pdf/products/the-hacker-digest-volume-11-pdf (cited on pages 133, 134).

[VS05] Johan Van Niekerk and Rossouw von Solms. "An holistic framework for the fostering of an information security sub-culture in organizations". In: *Information Security South Africa (ISSA)*. Volume 1. 13. 2005. URL: https://digifors.cs.up.ac.za/issa/2005/Proceedings/Full/041_Article.pdf (cited on page 159).

[WCM11] Merrill Warkentin, Lemuria Carter, and Maranda McBride. "Exploring the Role of Individual Employee Characteristics and Personality on Employee Compliance with Cybersecurity Policies". In: *The 2011 Dewald Roode Workshop on Information Systems Security Research*. 2011 (cited on pages 46, 69, 158).

[Weg87] Daniel M. Wegner. "Transactive Memory: A Contemporary Analysis of the Group Mind". In: *Theories of Group Behavior*. Edited by Mullen B. and Goethals G.R. Springer Series in Social Psychology. Springer, New York, NY, USA, 1987, pages 185–208. ISBN: 978-1-4612-4634-3. DOI: 10.1007/978-1-4612-4634-3_9 (cited on page 32).

[Wei80] Neil D Weinstein. "Unrealistic optimism about future life events". In: *Journal of personality and social psychology*

39.5 (1980), page 806. DOI: 10.1037/0022-3514.39.5.806 (cited on page 40).

[Wes08] Ryan West. "The Psychology of Security – Why do good users make bad decisions?" In: *Commun. ACM* 51.4 (Apr. 2008), pages 34–40. ISSN: 0001-0782. DOI: 10.1145/1330311.1330320 (cited on pages 34, 90).

[Wik18] Wikipedia contributors. *The Emperor's New Clothes — Wikipedia, The Free Encyclopedia*. 2018. URL: https://en.wikipedia.org/w/index.php?title=The_Emperor%27s_New_Clothes&oldid=874567812 (visited on 12/22/2018) (cited on page 73).

[Wik19] Wikipedia contributors. *Wilhelm Voigt — Wikipedia, The Free Encyclopedia*. 2019. URL: https://en.wikipedia.org/w/index.php?title=Wilhelm_Voigt&oldid=873207336 (visited on 01/29/2019) (cited on page 50).

[Wil+16] Mark D Wilkinson et al. "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific data* 3 (Mar. 2016), page 160018. ISSN: 2052-4463. DOI: 10.1038/sdata.2016.18 (cited on page 128).

[Wod07] Krzysztof Woda. "The Analysis of Money Laundering Techniques". In: *Cyber Warfare and Cyber Terrorism*. Edited by Lech J. Janczewski and Andrew M. Colarik. Idea Group Inc (IGI), 2007. Chapter XVIII, pages 138–145. ISBN: 9781591409922. DOI: 10.4018/978-1-59140-991-5 (cited on page 53).

[Wös15] Hans-Christian Wöste. *Kriegsende 1945: Willi Herold – In falscher Uniform vom Schornsteinfeger zum Henker*. 2015. URL: https://www.welt.de/article140055856/ (visited on 02/04/2019) (cited on pages 51, 94, 116).

[WT99] Alma Whitten and J Doug Tygar. "Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0." In: *Usenix Security*. Volume 1999. 1999 (cited on pages 12, 36).

[Wyn+11] Jackson Wynn, Joseph Whitmore, Geoff Upton, Lindsay Spriggs, Dan McKinnon, Richard McInnes, Richard Graubart, and Lauren Clausen. *Threat Assessment and Remediation Analysis (TARA) – Methodology Description Version 1.0*. Technical report. MITRE Technical Report MTR110176. MITRE, Oct. 2011. URL: https://www.mitre.org/sites/default/files/pdf/11_4982.pdf (cited on page 69).

[Zuc67] Carl Zuckmayer. *Der Hauptmann von Köpenick: ein deutsches Märchen in drei Akten*. Fischer Bücherei KG, 1967 (cited on pages 50, 77, 116).

Index

D

E

F

G

H

I

J

L

M

N

O

P

Q

R

S

T

U

V

W

DOI: 10.15480/882.9595

ISBN 978-3-947051-31-1

9 783947 051311