

# Defining and Fortifying Against Cognitive Vulnerabilities in Social Engineering

Yomna Abdelrahman<sup>\*1</sup>, Florian Alt<sup>\*2</sup>, Tilman Dingler<sup>\*3</sup>,  
Christopher Hadnagy<sup>\*4</sup>, Abbie Maroño<sup>\*5</sup>, and Verena Distler<sup>†6</sup>

- 1 University of the Bundeswehr – Munich, DE. [yomna.abdelrahman@unibw.de](mailto:yomna.abdelrahman@unibw.de)
- 2 University of the Bundeswehr – Munich, DE. [florian.alt@unibw.de](mailto:florian.alt@unibw.de)
- 3 Delft University of Technology, NL. [t.dingler@tudelft.nl](mailto:t.dingler@tudelft.nl)
- 4 Social-Engineer – Orlando, US. [chris@social-engineer.com](mailto:chris@social-engineer.com)
- 5 Social-Engineer – Orlando, US. [abbie@social-engineer.com](mailto:abbie@social-engineer.com)
- 6 University of the Bundeswehr – Munich, DE. [verena.distler@unibw.de](mailto:verena.distler@unibw.de)

---

## Abstract

Social engineering has become the main vector for human-centered cyber attacks, resulting from an unparalleled level of professionalization in the cybercrime industry over the past years. Hereby, through manipulation, criminals seek to make victims take actions that compromise security, such as revealing credentials, issuing payments, or disclosing confidential information. Little effective means for protection exist today against such attacks beyond raising awareness through education. At the same time, the proliferation of sensors in our everyday lives – both in personal devices and in our (smart) environments – provides an unprecedented opportunity for developing solutions assessing the cognitive vulnerabilities of users and serves as a basis for novel means of protection.

This report documents the program and the outcomes of the Dagstuhl Seminar 23462 “Defining and Fortifying Against Cognitive Vulnerabilities in Social Engineering”. This 3-day seminar brought together experts from academia, industry, and the authorities working on social engineering. During the seminar, participants developed a common understanding of social engineering, identified grand challenges, worked on a research agenda, and identified ideas for collaborations in the form of research projects and joint initiatives.

**Seminar** November 12–15, 2023 – <https://www.dagstuhl.de/23462>

**2012 ACM Subject Classification** Security and privacy–Human and societal aspects of security and privacy

**Keywords and phrases** Social Engineering, Cognitive Vulnerabilities, Phishing, Vishing

**Digital Object Identifier** 10.4230/DagRep.13.11.103

---

\* Editor / Organizer

† Editorial Assistant / Collector



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Defining and Fortifying Against Cognitive Vulnerabilities in Social Engineering, *Dagstuhl Reports*, Vol. 13, Issue 11, pp. 103–129

Editors: Yomna Abdelrahman, Florian Alt, Tilman Dingler, Christopher Hadnagy, and Abbie Maroño



DAGSTUHL Dagstuhl Reports

REPORTS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

## 1 Executive Summary


*Yomna Abdelrahman (University of the Bundeswehr – Munich, DE)*

*Florian Alt (University of the Bundeswehr – Munich, DE)*

*Tilman Dingler (Delft University of Technology – Delft, NL)*

*Christopher Hadnagy (Social Engineer – Orlando, US)*

*Abbie Maroño (Social Engineer – Orlando, US)*

**License**  Creative Commons BY 4.0 International license  
© Yomna Abdelrahman, Florian Alt, Tilman Dingler, Christopher Hadnagy, and Abbie Maroño

Social engineering which is defined as “any act that influences a person to take an action that may or may not be in their best interests”. In regards to when social engineering is being used by threat actors it is used as psychological manipulation of people into performing actions or disclosing confidential information. Sadly, this form of attack has existed for almost as long as mankind itself. With the advent of AI tools, this form of attack reached a new quality, posing a threat to any online user. Prominent forms of social engineering are phishing attacks and their various subforms (vishing, twishing, QRishing, etc.), physical attacks (dumpster diving, tailgating), and, more recently, deep fakes.

This three-day Dagstuhl Seminar on “Defining and Fortifying Against Cognitive Vulnerabilities in Social Engineering” brought together experts in (user-centered) security, psychology, HCI, computer science, and ethics to identify grand challenges and identify a research roadmap for mitigating social engineering threats. Over the course of the seminar, participants developed an in-depth understanding of the seminar topic. This was achieved by focusing on different aspects of social engineering, discussing how it links to the users’ vulnerabilities, namely cognitive vulnerabilities, and how mitigation approaches can be developed.

Day 1 began by introducing the seminar topic, focus, and goals. Afterwards, all participants introduced themselves and their areas of expertise. Each participant contributed and described reading material related to the seminar topic. The material was made accessible to all seminar participants and is attached as a reading list to this report. Following the introductions, day one featured a keynote by Prof. Angela Sasse, entitled “Manipulation, Deception, and Self-Deceit – Broadening Our Perspective of Social Engineering”. It highlighted how and why the digital environment makes us so susceptible to social engineering. It took a critical perspective on state-of-the-art approaches to address social engineering. The second talk of day one was given by Chris Hadnagy, who presented important and practical insights into the strategies of modern hackers. Both talks gave a compelling overview of social engineering attacks, an understanding of the most commonly targeted vulnerabilities, and a sense of why it is difficult to mitigate them. Participants then worked in groups to identify grand challenges in social engineering from both researchers’ and practitioners’ perspectives. Dr. Thomas Kosch and Dr. Yomna Abdelrahman jointly led the last session of day one. It focused on detecting cognitive vulnerabilities and provided an overview of sensing technologies and users’ internal states to be inferred, e.g., fatigue, cognitive load, etc. Day one concluded with a group work activity on what we can learn from modern sensors and how to design systems and methods to help mitigate social engineering attacks.

Day two started with a keynote by Mary D’Angelo, addressing the complex topic of understanding and tracing threat actors and social engineers on the dark web. It highlighted the need for collaborative efforts to understand this evolving threat better. Mary D’Angelo and Chris Hadnagy led an open discussion: on the one hand, it focused on the role of practitioners and industry in providing realistic data sets and insights from real-life attacks.

On the other hand, the question of how researchers could use those datasets to (a) better understand attacks and (b) design mitigating techniques was discussed. The second activity on day two was a walk to the ruins, during which participants, led by Claude Kirchner, discussed the ethical aspects of the seminar topic. The afternoon of day two was a group work activity led by Dr. Mohamed Khamis in which participants worked towards addressing the previously identified grand challenges. Breakout groups focused on the different attack phases. Day two ended by transforming the proposed solutions into concrete research projects and agendas.

Day three started with a keynote by Alia Saad, which demonstrated different approaches to addressing human-centered security issues from a technical perspective, using examples from current research. Participants followed up on the proposed research projects in the second session of the day, led by Prof. Florian Alt and Prof. Tilman Dingler. They worked together on refining their ideas and identifying potential collaborations.

This Dagstuhl Seminar provided a platform for interdisciplinary collaboration, fostering a deeper understanding of social engineering and its cognitive vulnerabilities. The identified grand challenges and proposed research projects underscore the importance of collaborative efforts between researchers and practitioners in fortifying against evolving social engineering threats. The insights of this seminar lay the foundation for future research and initiatives in the ongoing battle against malicious psychological manipulation in the digital age.

This seminar had several outcomes. First, it established a community of researchers and practitioners with a common understanding of emerging security threats through social engineering. Second, grand challenges were identified that led to a roadmap for social engineering research, including various research questions addressing theoretical, practical, and methodological aspects. Third, ideas for joint research projects emerged, for several of which an initial consortium was established. Among these projects is the idea of establishing a European Research Center on Awareness, Detection, and Mitigation of Social Engineering, the utilization of a dark web dataset that provides insights into the behaviors of threat actors that lead up to an attack, the utilisation of AI to detect sensitive information in unwanted data disclosures (e.g., via social media shares), and an approach to detecting threats in audio conversations based on voice features and conversation behaviours.

## 2 Table of Contents

### Executive Summary

<i>Yomna Abdelrahman, Florian Alt, Tilman Dingler, Christopher Hadnagy, and Abbie Maroño</i> . . . . .	104
--	-----

### Overview of Talks

Manipulation, Deception, and Self-Deceit – Broadening Our Perspective of Social Engineering <i>Angela Sasse</i> . . . . .	107
Physiological Security and Cognitive Vulnerabilities <i>Thomas Kosch, Yomna Abdelrahman</i> . . . . .	107
Threat Actors and Threat Intelligence on the Dark Web <i>Mary D'Angelo</i> . . . . .	108
Biometrics Against Social Engineering <i>Alia Saad</i> . . . . .	109

### Working Groups

Manipulation Mastery – The Strategies of Modern Hackers <i>Christopher Hadnagy, Abbie Marono</i> . . . . .	110
Grand Challenges in Social Engineering <i>Matteo Große-Kampmann, Angela Sasse</i> . . . . .	112
Towards Solutions: Cognitive Vulnerabilities <i>Thomas Kosch, Yomna Abdelrahman</i> . . . . .	117
Social Engineering for Good (A Walk to the Castle Ruins) <i>Tilman Dingler</i> . . . . .	119
Identifying Research Areas and Research Questions <i>Mohamed Khamis</i> . . . . .	120
Towards Collaborations in Social Engineering Research <i>Florian Alt</i> . . . . .	123

<b>Report Summary</b> . . . . .	126
---------------------------------	-----

<b>Reading List</b> . . . . .	127
-------------------------------	-----

<b>Participants</b> . . . . .	129
-------------------------------	-----

## 3 Overview of Talks

### 3.1 Manipulation, Deception, and Self-Deceit – Broadening Our Perspective of Social Engineering

Angela Sasse (*Ruhr-Universität Bochum, DE, [martina.sasse@rub.de](mailto:martina.sasse@rub.de)*)


License  Creative Commons BY 4.0 International license  
© Angela Sasse

The talk examines why the digital environment makes us so susceptible to social engineering – because we have become so used to being manipulated and deceived by others that we don't notice and deceive ourselves that we have choice and control. In fact, ubiquitous tracking of our online activities has created a huge information asymmetry, which Soshanna Zuboff has described as “surveillance capitalism”, and humans have adopted routines to respond to prompts to give our time, attention, and money. While we do not regard them as “attackers” in the traditional cybersecurity sense, they utilize very similar cues and exploit habits. To regain control, we need to engage in regular goal setting and planning of our activities and ration our digital engagements along the lines of Cal Newport’s “digital minimalism”. But we also urgently need reliable trust anchors to enable humans to distinguish friends from foes.

### 3.2 Physiological Security and Cognitive Vulnerabilities

Thomas Kosch (*HU Berlin, DE, [thomas.kosch@hu-berlin.de](mailto:thomas.kosch@hu-berlin.de)*)

Yomna Abdelrahman (*University of the Bundeswehr – Munich, DE, [yomna.abdelrahman@unibw.de](mailto:yomna.abdelrahman@unibw.de)*)

License  Creative Commons BY 4.0 International license  
© Thomas Kosch, Yomna Abdelrahman

Human physiology exerts electric potentials that can be captured by computing devices. Such physiological signals allow the assessment of user states, such as cognitive workload, affect, or stress [2]. While these states can be assessed to allow users to quantify themselves, they are also a gateway for exploiting behaviors in real-time. Social engineering attacks can be tailored depending on the user states, thus increasing the likeliness of social engineering attacks. Furthermore, sensors, such as thermal cameras, are becoming more ubiquitous. Thermal cameras have recently drawn the attention of HCI researchers as a new sensory system enabling novel interactive systems. They are robust to illumination changes, making separating objects from the scene background easy. Far-infrared radiation, however, has another characteristic that distinguishes thermal cameras from their RGB or depth counterparts as it operates in the non-visual spectrum. On the other hand, the visual spectrum, i.e., human visual perception, is limited to only 1 percent of the electromagnetic spectrum. Research has shown that extending visual perception can be beneficial. To investigate the potential of the adoption of thermal imaging, we present the conducted studies to infer users' states, e.g., cognitive load, attention type [5, 6], as well as environmental state, e.g., the presence of recording devices [7], and foot traces [8]. Our findings reflected the potential of thermal imaging to further protect the user by knowing the user's state and nudging them when they are cognitively vulnerable. Yet, our research also explores how thermal imaging might introduce novel attacks, namely thermal attacks[9].

## References

- 1 S. H. Fairclough. 2009. Fundamentals of physiological computing. *Interacting with computers*, 21(1-2), 133-145.
- 2 T. Kosch, J. Karolus, J. Zagermann, H. Reiterer, A. Schmidt & P.W. Woźniak. 2023. A survey on measuring cognitive workload in human-computer interaction. *ACM Computing Surveys*. Association for Computing Machinery, New York, NY, USA
- 3 R.W. Picard. 2000. *Affective computing*. MIT Press.
- 4 N. Sharma & T. Gedeon. 2012. Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer methods and programs in biomedicine*, 108(3), 1287-1301.
- 5 Y. Abdelrahman, E. Velloso, T. Dingler, A. Schmidt & F. Vetere. 2017. Cognitive Heat: Exploring the Usage of Thermal Imaging to Unobtrusively Estimate Cognitive Load. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* <https://doi.org/10.1145/3130898>
- 6 Y. Abdelrahman, A.A. Khan, J. Newn, E. Velloso, S.A. Safwat, J. Bailey, A. Bulling, F. Vetere & A. Schmidt. 2019. Classifying Attention Types with Thermal Imaging and Eye Tracking. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 69 (September 2019). <https://doi.org/10.1145/3351227>
- 7 S. Prange, A. Shams, R. Piening, Y. Abdelrahman & F. Alt. 2021. PriView– Exploring Visualisations to Support Users’ Privacy Awareness. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI ’21)*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3411764.3445067>
- 8 A. Saad, K. Izadi, A. Khan, P. Knierim, S. Schneegass, F. Alt & Y. Abdelrahman. 2023. Hot-Foot: Foot-Based User Identification Using Thermal Imaging. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI ’23)*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3544548.3580924>
- 9 Y. Abdelrahman, M. Khamis, S. Schneegass & F. Alt. 2017. Stay Cool! Understanding Thermal Attacks on Mobile-based User Authentication. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI ’17)*. Association for Computing Machinery, New York, NY, USA, 3751–3763. <https://doi.org/10.1145/3025453.3025461>

### 3.3 Threat Actors and Threat Intelligence on the Dark Web



*Mary D’Angelo (Searchlight Cyber – Washington, DC, US)*

License  Creative Commons BY 4.0 International license  
© Mary D’Angelo

This speech addresses the complex topic of threat actors and social engineering on the dark web, highlighting the need for collaborative efforts to understand this evolving threat. The speaker begins by providing a historical overview of the development of the dark web, from the creation of ARPANET in 1969 to the advent of TOR and Bitcoin, which have facilitated a surge in dark web activities over the last 15 years. Current trends in the dark web, including the rise of malicious social engineering practices, are discussed, with examples such as services for phishing, vishing attacks, and educational resources for threat actors. The MGM hack by Scattered Spider serves as a case study to illustrate the sophisticated nature of these attacks. The speaker emphasizes the urgency of understanding the mechanisms of threat actor communication and transaction on the dark web, the organization of these actors, and their growing capabilities, as evidenced by a significant increase in vishing attacks. The speech concludes with a call for collaborative research between practitioners and the academic community to develop effective defenses against these evolving cyber threats.

### 3.4 Biometrics Against Social Engineering

Alia Saad (University of Duisburg-Essen, DE, [alia.saad@uni-due.de](mailto:alia.saad@uni-due.de))

License  Creative Commons BY 4.0 International license  
 Alia Saad

This talk demonstrated several approaches to mitigating human-centered attacks based on current research examples. The talk was meant to inspire discussion among participants as to how the challenges identified during the seminar can be approached in joint research projects. The first part demonstrated how situations in which users are exposed to a human-centered attack can be studied in detail, using shoulder-surfing as an example [1]. Furthermore, an example was shown of how a user interface can be built that points out risk in-situ [2]. The second part demonstrated how the need for user interaction can be minimized by creating, implementing, and evaluating technical approaches seamlessly running in the background, using behavioral biometrics as an example. The talk demonstrated how mechanisms based on different behaviors can be built, in particular, gait [6] and hand-based interaction [3]. Furthermore, the talk also demonstrates a system for use in everyday life [4].

#### References

- 1 A. Saad, J. Liebers, U. Gruenefeld, F. Alt & S. Schneegass. 2021. *Understanding Bystanders' tendency to shoulder surf smartphones using 360-degree videos in virtual reality*. In Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction (MobileHCI '23). Association for Computing Machinery, New York, NY, USA.
- 2 A. Saad, M. Chukwu & S. Schneegass. 2018. Communicating shoulder surfing attacks to users. In Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia (MUM '18). Association for Computing Machinery, New York, NY, USA.
- 3 A. Saad, M. Pascher, K. Kassem, R. Heger, J. Liebers, S. Schneegass & U. Gruenefeld. 2023. Hand-in-Hand: Investigating Mechanical Tracking for User Identification in Cobot Interaction. In Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia (MUM '23). Association for Computing Machinery, New York, NY, USA.
- 4 A. Saad, K. Izadi, A.A. Khan, P. Knierim, S. Schneegass, F. Alt & Y. Abdelrahman. 2023. HotFoot: Foot-Based User Identification Using Thermal Imaging. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 262, 1–13. <https://doi.org/10.1145/35444548.3580924>
- 5 Mihai Bâce, Alia Saad, Mohamed Khamis, Stefan Schneegass, and Andreas Bulling. *PrivacyScout: Assessing Vulnerability to Shoulder Surfing on Mobile Devices*. Proceedings on Privacy Enhancing Technologies 3 (2022): 650-669.
- 6 A. Saad, N. Wittig, U. Gruenefeld & S. Schneegass. 2022. A Systematic Analysis of External Factors Affecting Gait Identification. In the IEEE International Joint Conference on Biometrics (IJCB), pp. 1-9. IEEE.

## 4 Working Groups


Our seminar brought together participants from academia, industry, and the authorities. We began our seminar with level-setting. Short foundational talks from practitioners and researchers aimed to foster a common level of understanding of the differing perspectives of the various communities as well as a joint language. Based on this, the seminar focused on interactive formats (break-out groups, open discussions) intending to identify grand challenges, a research roadmap, and opportunities for collaboration.



## 4.1 Manipulation Mastery – The Strategies of Modern Hackers

*Chris Hadnagy (Social Engineer – Orlando, FL, USA, chris@social-engineer.com)*

*Abbie Marono (Social Engineer – Orlando, FL, USA, abbie@social-engineer.com)*

License  Creative Commons BY 4.0 International license  
© Christopher Hadnagy, Abbie Marono

**Objective** Due to the very different backgrounds of seminar participants (scientists, practitioners, military), this session aimed to create a common understanding of social engineering and identify state-of-the-art strategies of hackers.

**Methodology** This session followed a two-step approach to develop a common understanding. Firstly, Chris Hadnagy introduced the fundamentals of social engineering. Afterward, participants were split into groups, discussing a set of questions to create a joint understanding of social engineering from different perspectives.

### Step 1: Fundamentals

The talk by Chris Hadnagy aims to define the fundamentals and main vectors used by malicious social engineers in attacking their targets. We define how attackers use phishing, vishing, SMiShing, and impersonation in their attacks. By defining each of these and discussing the advancement in the technology used, we can better understand the minds of the attackers in choosing which method to use. Following this discussion, we went into depth about the stages of social engineering engagement from a practitioner standpoint. The goal was to understand the methodology used by professional social engineers.

#### 4.1.1 Terminology: Social Engineering

Social engineering involves manipulating individuals to give away confidential information or perform actions that compromise security. Techniques include phishing, vishing, and impersonation, where hackers exploit human psychology and vulnerabilities of different kinds.

#### 4.1.2 Attack Vectors

State-of-the-art attack vectors encompass a range of techniques malicious social engineers utilize to exploit vulnerabilities and gain unauthorized access to sensitive information or systems. Prominent attack vectors include:

**Phishing** Phishing involves fraudulent attempts to obtain sensitive information, such as usernames, passwords, and credit card details, by disguising as a trustworthy entity in electronic communication. These attacks commonly occur via email, where users are persuaded to click malicious links or provide confidential information.

**Vishing** Vishing, or voice phishing, is a form of social engineering where attackers use phone calls to deceive individuals into disclosing personal or financial information. The attackers may impersonate legitimate organizations or individuals to manipulate victims into revealing sensitive data or performing certain actions.

**Smishing** Smishing, or SMS phishing, exploits text messaging systems to trick users into revealing personal information or installing malware on their devices. Attackers send deceptive text messages containing malicious links or requesting sensitive information, exploiting the trust associated with SMS communications.



**Impersonation** Impersonation attacks involve masquerading as a trusted entity, such as an authority, figure, or reputable organization, to deceive individuals into disclosing confidential information, transferring funds, or performing actions that compromise security. Attackers often use social engineering tactics to gain the trust of their targets before exploiting it for malicious purposes.

It is worth mentioning that these attack vectors continue to evolve as attackers adapt their strategies to bypass security measures and exploit new human vulnerabilities.

## Step 2: Breakout Groups

Following the introduction and plenum discussion to obtain a common understanding, participants split into two subgroups to discuss the following questions:

### 4.1.3 Guiding Questions

- How do modern hackers target their victims?
- How are you educating yourself/ your team on security awareness?
- What are the limitations of this type of education?
- What problems have you encountered trying to manage threats from Social Engineering?
- What kind of collaborations/technologies would help increase your security?

### 4.1.4 Outcomes

The groups collated a range of techniques of modern attacks, including:

**Deepfake Technology** Attackers may use deepfake technology to create convincing fake audio or video recordings for social engineering attacks, impersonating trusted individuals or manipulating content.

**Machine Learning and Artificial Intelligence** As technologies evolve, attackers may leverage machine learning and artificial intelligence to enhance their attacks. This includes creating more sophisticated malware, evading detection, or automating certain aspects of the attack process, e.g., generating phishing emails.

**Augmented Reality (AR) and Virtual Reality (VR) Threats** Attackers may exploit AR and VR technologies for social engineering attacks, creating immersive scenarios to deceive victims or launching attacks within virtual environments.

When the groups discussed the limitations of training or educating users and the problems of managing social engineering attacks, they came up with several reasons why this can be challenging, including the following:

It is difficult to convince people that damage is possible and real for several reasons:

- **Lack of personal experience:** Users may not have personally encountered an incident, leading to a perception that such events are rare or unlikely to affect them. Without direct experience, it can be challenging to grasp the potential consequences.
- **Trust in technology:** We often trust and rely on technology in our daily lives, which might make us feel that systems are secure and that incidents like data breaches or cyberattacks won't happen to us.
- **Threats are invisible:** Unlike physical attacks resulting in visible damage, the effects of social engineering may be hidden. Data breaches, e.g., might not immediately manifest as tangible harm, making it difficult for individuals to recognize the severity of the situation.
- **Reaction times are generally too long:** Victims may feel overwhelmed when confronted with the potential damage caused by falling for a social engineering attack. This discomfort can lead to denial or avoidance of reporting the issue.

**Training does not work because humans forget / cannot memorize everything:**

One-time or infrequent training sessions may not be sufficient to create lasting awareness.

**Current mitigation strategies are difficult to scale:** Several factors contribute to the difficulty of scaling social engineering mitigation strategies. The discussed challenges covered:

- **Limited resources:** Many organizations have limited resources, both in terms of time and budget, to devote to extensive training programs.
- **Constantly evolving attacks:** Social engineering strategies continuously evolve, and attackers regularly develop new techniques. Staying ahead of these evolving threats and updating training content accordingly is resource-intensive and time-consuming.
- **Measuring effectiveness:** Determining the effectiveness of social engineering training programs is challenging.
- **Privacy concerns:** Balancing the need for effective social engineering training with respect for users' privacy can be a delicate task. Some individuals may hesitate to participate in training that they perceive as invasive.
- **Existence and scalability of Technical Solutions:** Implementing technical solutions to detect and prevent social engineering attacks can be complex and do not exist yet.
- **Lack of ecological validity:** The groups discussed potential collaboration between practitioners and researchers to address the above-mentioned challenges better. The reported concern was the struggle of researchers to move research out of the lab; current approaches are often scenario-oriented (but lack the ecologic validity of real-world settings).

## 4.2 Grand Challenges in Social Engineering

*Matteo Große-Kampmann (Aware7 – Gelsenkirchen, DE, matteo@aware7.de)*

*Angela Sasse (Ruhr-Universität Bochum, DE, martina.sasse@rub.de)*

License © Creative Commons BY 4.0 International license  
© Matteo Große-Kampmann, Angela Sasse

**Objective** This session aimed to identify grand challenges in social engineering from both a practical and academic perspective.

**Methodology** To structure this working group, Chris Hadnagy first introduced the different phases leading up to a successful social engineering attack. Those steps then served as a scaffold for breakout groups in which grand challenges for each phase were identified.

## Phases of a Successful Social Engineering Attack

**Phase 1–OSINT** Open Source Intelligence (OSINT) refers to collecting and analyzing information from publicly available sources. It involves gathering data from various sources such as social media, news articles, online forums, public databases, and websites. OSINT provides social engineers with valuable information about their targets, which can be used to craft convincing narratives and exploit vulnerabilities. By utilizing OSINT, social engineers can gather personal details, interests, affiliations, and even behavioral patterns of their targets. This information enables them to tailor their approaches to appear more trustworthy and increase the chances of successful exploitation.

**Phase 2–Target Selection** Social engineers target individuals who can access valuable information or sensitive systems. This could be employees of a company, individuals in positions of authority, or those with access to financial information. By targeting those with valuable data, social engineers increase the likelihood of a successful attack or fraud.

**Phase 3–Attack Plan** Social engineers craft a highly personalized attack plan using the information gathered during the prior phases. They adopt different personas, using tactics such as impersonation, pretexting, or creating fake online profiles to establish credibility. They exploit emotions and trust by pretending to be someone the target knows or trusts.

**Phase 4–Attack Launch** Social engineers conduct their attacks by manipulating and exploiting human psychology and trust. They use various tactics to manipulate individuals into divulging confidential information or performing actions that could compromise security. They often employ techniques like impersonation, pretexting, and phishing to trick people into believing they are someone they are not or representing a trustworthy entity. Social engineers can access sensitive information and financial data by exploiting human emotions, curiosity, and ignorance or by gaining unauthorized entry into systems. These attacks can occur through various mediums, such as phone calls, emails, social media, or even in-person interactions, to deceive individuals and bypass security systems.

**Phase 5–Evaluation** Throughout the process, social engineers document their actions, record findings, and assess the impact of any successful exploits.

**Phase 6–Reporting** Finally, they provide a detailed report to the organization, outlining the vulnerabilities discovered and recommending remediation measures.

## Practical Challenges and Research Challenges

### Phase 1–OSINT

Participants discussed why educating users to protect themselves from the initial phase of social engineering is complex. The participants identified the following challenges concerning Open Source Intelligence:

**Creating Awareness of Own Vulnerabilities:** A key challenge is making users aware of their vulnerabilities. Users should be aware of what information is publicly shared and could be used by attackers (and which information is not publicly stored but could be accessible by attackers if the platform is breached). Moreover, many users struggle to think that they are targets in the first place.

**Inference of Available Data:** A challenge is keeping an overview of available information about oneself when AI models draw conclusions based on metadata (e.g., relationships).

**Social Media Exploitation:** Social media platforms often contain tons of personal information. Attackers can exploit this information to create phishing messages, impersonate individuals, or conduct other forms of social engineering.

**Dynamic and Evolving Nature of Platforms:** Online platforms and available information constantly change and evolve. Keeping track of these changes and assessing the reliability of information can be challenging, for instance, how social media platforms change their privacy settings and auto-send friend requests.

**Lack of Users' Awareness of the Influence of Their Internal State:** Users' internal states and vulnerabilities play a significant role in their susceptibility to social engineering attacks. As social engineering relies on manipulating individuals' emotions, behaviors, and cognitive processes to deceive them into revealing sensitive information, taking specific actions, or compromising security. Yet, users are not aware of such an influence.

Based on the group discussion and the pointed-out challenges, participants discussed potential solutions to help protect the user and increase their awareness about OSINT. Attackers usually use information aggregation during the OSINT phase, where they gather information from multiple sources to create a comprehensive profile of a target. This profile can be used to craft more convincing and targeted social engineering attacks. Accordingly, participants envisioned a *Cross-Platforms Search Notification System*, where users would be alerted if someone searches them on different platforms, e.g., work/personal website, LinkedIn, Instagram, Facebook, and warn the user if access rates are unusually high. However, this entails technical and privacy challenges, e.g., logging search activities across platforms.

## Phase 2–Target Selection

Identifying potential targets and implementing strategies to mitigate risks can be complex for both defenders and users. Participants openly discussed why this phase might be hard to mitigate. One dominant reason was that many users struggle to think they were targets in the first place. Hence, they exhibit neglectful behavior when dealing with both their own data and institutionally accessible information.

## Phase 3–Attack Plan

The central premise of an attack plan is coming up with a pretext for the attack. During the planning phase, the attacker uses the information gathered from OSINT to devise a strategy. This involves selecting the most appropriate attack vector—whether it be phishing, pretexting, baiting, or another method—based on the target's vulnerabilities and the attacker's objectives. The attacker also crafts the message or scenario they will use to deceive the target, ensuring it is convincing enough to elicit the desired response. This phase requires careful consideration of the psychological and emotional triggers that will be most effective on the target and planning for any contingencies or responses the target might have.

Attack planning also involves the creation of backstories, fake identities, or any necessary props (like counterfeit badges or websites) that will make the attacker's approach more credible. This is where the creativity and insight of the attacker into human psychology are paramount. The success of this phase hinges on how well the attacker can anticipate the target's reactions and prepare for them, ensuring that the attack will not only reach the target but also resonate with them, prompting the desired action or information disclosure.

One of the most effective ways to counteract the planning phase is to limit already access to information that could be gathered during OSINT. Corporations should regularly check what type of information is publicly accessible, including about their employees. Awareness campaigns about public profiles can hone employees' sensitivity to sharing certain information about themselves or their employers.

## Phase 4—Attack Launch

In the attack launching phase, the attacker puts their plan into action and makes direct contact with the target. This could be through email, phone calls, social media, or in-person interactions. The attacker employs the crafted scenario to manipulate the target into performing specific actions, such as divulging sensitive information, granting access to secure systems, or even transferring funds. The success heavily relies on the attacker's ability to adapt to the conversation flow and maintain the deception convincingly.

During this phase, the attacker must remain vigilant and adaptable, as unforeseen variables or responses from the target may require on-the-fly adjustments to the plan. The psychological manipulation skills of the attacker are crucial here, as they must build trust or authority with the target quickly. The ability to read cues from the target and adjust the approach accordingly can make or break the attack's success.

Countermeasures can be taken individually through general awareness training or collectively through peer protection. The latter entails establishing a reporting culture in which employees inform and warn each other about new schemes they encounter. Employers should establish a single point of contact where incidents can be easily reported and that is responsible for disseminating newly identified threats. Another approach that has merit on both individual and collective levels is the introduction of *friction*. Artificially delaying certain actions or procedures, e.g., can create time windows in which reason can kick in, or the attack can be delayed to a point where the risk of exposure becomes too great to continue the attack. Urgency should almost always be a warning sign for an incoming attack.

## Phase 5—Evaluation

Evaluating social engineering attacks, particularly those conducted as part of penetration testing, poses several challenges.

**Ecologic Validity and Generalizability** While pen testers try to act realistically, their actions are still limited by legal and ethical considerations. Also, their customers might exclude certain actions (e.g., accessing sensitive and personal information about employees). Those conditions inevitably influence the ecologic validity of the findings and pose the question of to which degree the findings generalize to settings with real attackers.

**Metrics** Many forms of penetration testing are still strongly limited in terms of the used metrics. For example, click rates are among the most popular metrics for phishing awareness campaigns. At the same time, these allow only very little to be learned and are questionable, as they depend on factors beyond pen testers' control (and assessment). There is a need to rethink metrics currently in use fundamentally.

**Individualization** Measures against social engineering are generally costly from a corporate perspective, as a result of which easy-to-implement solutions are favored (e.g., making users attend talks on awareness once a year). The challenge with such measures is that they might annoy users (as content might be repetitive). Also, employees might have a different level of knowledge and understanding, as a result of which some might struggle with terminology already while others might be bored. A major challenge is the individualization of measures, where users' skills, prior knowledge, and tasks of their everyday job are considered.

**Case Study vs. Large Scale** Due to the required effort and cost, campaigns and research projects often focus only on specific cases rather than large-scale approaches. While (small-scale) case studies might be well suited to identify causes and interesting aspects, more large-scale studies are required to assess effects.

**Priming** For ethical reasons, users or employees might be primed; that is, they are being told upon being employed or signing up for a study that they will / might be subject to a security assessment. This inevitably changes behavior. Research is needed to understand the implications of priming and on approaches that minimize any such priming effects.

**Independent Auditing** Pentesting / auditing is often conducted due to certification or due to being required by insurance companies. As a result of this, companies subsequently hire auditors. The challenge here is that pen tests and audits, in this case, are not independent. It is still an open question about how such independence can be achieved.

**Completeness** While defenders must protect against any vulnerability, attackers only need to find one weakness. Pentesting usually cannot consider any aspect/attack surface.

**Pentester Empathy** A particularly interesting aspect is their actions' implications on pen testers (similar to the Milgram experiment). This is an unexplored area of research.

## Phase 6—Reporting

The group identified several challenges regarding the reporting and, in particular, the way in which recommendations are / should be made.

**Turning lessons learned into positive change** While many pen tests are designed to demonstrate issues/holes that allow attackers to be successful, it is often much less clear how this knowledge can be turned into positive change.

**Targeting Opportune Moments** Closely related to the abovementioned aspect, change must be carefully targeted to opportune moments, i.e., moments in which users are (more) receptive to change and appreciate it. It is well known that in situations of change (e.g., moving to a new house/office, getting a new smartphone/laptop), people are more willing to change habits, but this is hardly explored from a cybersecurity perspective.

**Check-the-Box vs. Organizational Change** Penetration tests/auditing is often seen today as a necessary requirement rather than an opportunity for real (organizational) change. It remains an open challenge to move from just getting things done to a culture in which true organizational change is anticipated.

**Misconceptions about being a target** Many users struggle to understand/accept that they are a target. Reports can surface convincing cases (e.g., kindergartens being targets of cyber attacks).

**Response Costs / Prioritization** Cyber attacks are usually possible through many different approaches, and addressing all of them is costly. There is a need to understand better how countermeasures can be prioritized so as to maximize their impact.


**Change Management / Leadership** Who should drive change is often unclear. Whereas employees often consider employers responsible for cybersecurity, employers want employees to change their habits. A challenge is how to establish a social contract.

**Expressing IT Security as a Business Risk** In particular, companies and individuals struggle to accept IT security as a business risk until they become victims. Research is needed as to how the consequences of successful cyber attacks can be better conveyed.

### 4.3 Towards Solutions: Cognitive Vulnerabilities

Thomas Kosch (HU Berlin, DE, [thomas.kosch@hu-berlin.de](mailto:thomas.kosch@hu-berlin.de))

Yomna Abdelrahman (University of the Bundeswehr – Munich, DE, [yomna.abdelrahman@unibw.de](mailto:yomna.abdelrahman@unibw.de))

License  Creative Commons BY 4.0 International license  
© Thomas Kosch, Yomna Abdelrahman

**Objective** The objectives of this session were to (1) understand how modern sensing technology can be used to assess user states affected by social engineering, to (2) think about how knowledge of those states can be used to design counter measures and (3) how attackers can exploit this knowledge.

**Methodology** Thomas Kosch and Yomna Abdelrahman first introduced the capabilities of modern sensors and machine learning (see talk). Afterward, participants were divided into groups and asked to think about ways of using knowledge obtainable from sensors to defend from social engineering and also which novel attack surfaces this creates.

As previously discussed, users' internal states significantly influence their susceptibility to social engineering attacks. Social engineering relies on manipulating individuals' emotions, behaviors, and cognitive processes to deceive them into revealing sensitive information, taking specific actions, or compromising security. In this session, we leverage design fiction methods to ideate the role of physiological sensors in social engineering from both the attackers' and defenders' perspectives.

#### What can we learn from modern sensors?

Thomas Kosch and Yomna Abdelrahman gave concrete examples of utilizing novel ubiquitous sensors like eye tracking and thermal cameras to detect and leverage cognitive vulnerabilities during social engineering. However, opportunities are not limited to these examples but rather to steer the mindset of the participants towards using modern sensors in different contexts.

**Eye Tracking:** Eye-tracking data in social engineering refers to collecting and analyzing information about a person's eye movements and gaze patterns. While traditional uses of eye-tracking are often associated with research in psychology and HCI usability studies, the application of eye-tracking data in the realm of social engineering introduces additional considerations. Here are some examples:

- **Understanding Attention:** Eye tracking data can provide insights into where a person directs their attention. In a social engineering context, understanding what elements or cues attract a person's gaze can be valuable for attackers. Attackers may use this data to refine deceptive techniques.
- **Phishing and Visual Deception:** Attackers may leverage eye-tracking data to optimize the design of phishing emails. By understanding where users focus their attention, attackers can create more convincing and visually deceptive elements to increase the likelihood of successful social engineering attacks.
- **Defensive Techniques:** While attackers could utilize eye-tracking, they also hold merits for defenders. Researchers and defenders can use eye-tracking data to understand how users visually engage with social engineering attacks and security warnings. This information can inform the design of more effective alerts and communication strategies to raise awareness about potential social engineering threats.



**Thermal Cameras:** Facial temperature can be influenced by emotional states, and temperature changes may reflect variations in emotions and internal states, e.g., cognitive load, stress, anger, etc. These changes could be monitored seamlessly and non-invasively using thermal cameras. While researchers utilized thermal cameras to build cognitive-aware systems in various contexts, it is unexplored in the context of social engineering, yet the potential it holds, for instance:

- **Cognitive Load Detection:** Recent work showed the potential of using thermal cameras to capture facial temperature to quantify cognitive load from low, medium, high, and very high. Research reflected the potential of thermal imaging to further protect the user by knowing the user's state and nudging them not to perform any security critical tasks when they are cognitively vulnerable.
- **Emotions Detection:** Attackers rely on the victims' emotions to conduct social engineering attacks. Namely, they aim to simulate emotions, e.g., fear, guilt, and stress, to make users reveal sensitive information or perform certain actions. Research revealed correlations between changes in facial temperature and these emotions. Emotional-aware systems built using thermal cameras could detect these emotions and either used by the defenders to build protective measures when such emotions are detected or by the attacker to know vulnerable moments.

Following the talk, we had an open discussion on how modern sensors could be deployed in the context of social engineering to serve both attackers and defenders. To this end, participants split into two subgroups to discuss the following questions:

#### 4.3.1 Guiding Questions

- Imagine you are an IT security designer. How would you use psychological real-time data to improve user security?
- Imagine you are a hacker who has access to psychological real-time data. How would you utilize the data for a social engineering attack?
- How can users be made aware of their individual cognitive vulnerabilities?

#### How can we use the knowledge of user states to build better protection mechanisms?

The groups came up with the following ideas for novel protection mechanisms.

**Mechanism 1** During face-to-face situations, use a video-based assessment of physiological data to give insights into the current level of fatigue, stress/arousal, and identify health status. Based on the signals, the system would recognize if the user is sleepy/exhausted, provide recommendations, and flag potential threats.

**Mechanism 2** During reading emails, use data such as heart rate, EDA, pupil dilation, blink rate (fatigue), reading speed, eye tracking for speed of reading emails, and nonverbal body posture to determine when the user is vulnerable, and the email client color changes to indicate cognitive vulnerability.

**Mechanism 3** One group proposed using physiological data not to detect victims' vulnerabilities but the attacker's intent. Once an attacker is detected, it is flagged.

## How can attackers exploit knowledge of user states?

The groups came up with the following ideas for novel possible attacks.

**Attack 1** Attackers could exploit the knowledge about users' states to tailor the attack and abuse the user during their vulnerable state.

**Attack 2** Having access to heart rate, EDA, pupil dilation, blink rate (fatigue), and reading speed data, attackers could use the data to find a point when the target is most stressed and fatigued, being overly busy. During that time, a phishing email is sent from someone in authority over the target requesting immediate action.

Interestingly, the participants discussed how they can benefit from the potential of using physiological real-time data without the risk of falling into the attackers' hands. For instance, participants proposed randomly introducing noise to the data or using differential privacy. Adding enough noise makes the attack infeasible but still allows data to be used legitimately.

## 4.4 Social Engineering for Good (A Walk to the Castle Ruins)

*Tilman Dingler (TU Delft, NL, t.dingler@tudelft.nl)*

*Claude Kirchner (INRIA Institut National de Recherche en Informatique et en Automatique – Rocquencourt, FR, claude.kirchner@inria.fr)*

License © Creative Commons BY 4.0 International license  
© Tilman Dingler

So far, social engineering has been mainly discussed in association with deceptive practices aimed at exploiting human psychology for malicious ends. The goal of this session, however, was also to consider the use of its principles and mechanics for positive individual and social outcomes, *i.e.* for good. Social engineering (SE) can be defined as “any engineered act that influences a person to take an action that may or may not be in their best interest”. This allows us to insist on the specially designed intention (engineered) as well as to emphasize that it could be for malicious but also positive “for good” reasons. In this context, we should consider the involved ethics, understood as the thinking process about human conduct and the values on which they are based. Indeed, either for good or bad, social engineering may not respect human autonomy, transparency, or explainability, and of course, the non-maleficence principle will be strongly questioned.

During a joint walk up to the old castle ruins, seminar participants were thus presented with two leading questions and invited to discuss in present company and eventually report back. The two leading questions posed were:

1. How would you use insights, techniques, and methods of Social Engineering to do good?
2. What are ethical boundaries and obligations when “manipulating” people for good?


Examples discussed by participants included personal health and environmental conservation. In the realm of public health, social engineering and, specifically, nudging can play pivotal roles. For instance, designing environments that subtly encourage physical activity, such as strategically placed stairs over escalators, can significantly impact public health outcomes. Similarly, nudges in cafeterias or grocery stores, like placing fruits and vegetables at eye level, can make healthy food choices more appealing and accessible. These interventions use our natural tendencies and decision-making shortcuts for positive ends, making the healthier choice, the easier or more attractive option.

On another note, environmental sustainability can benefit from social engineering techniques aimed at encouraging eco-friendly behaviors. For example, utility companies have successfully used social norms to influence behavior by showing customers how their energy consumption compares to their neighbors, nudging them to reduce energy use. Similarly, simple prompts or reminders to recycle or making recycling bins more visible and accessible can significantly increase recycling rates. These strategies rely on our innate desire to conform to social norms and our responsiveness to environmental cues, guiding us toward more environmentally sustainable actions.

A commonly mentioned critique of nudges was the user's agency, which might potentially be violated. Even nudges "for good" are construed by a choice architect, *i.e.* an individual or group of people who deem one choice *better* than another. Conflicting moral and value systems can, therefore, give precedence to choices that go against what the individual might have selected in a more conscious choice scenario. In the end, any technique deemed as social engineering entails some kind of manipulation. The question of which manipulation is deemed "good" or "bad" needs to be discussed in light of differing moral and ethical frameworks. People's agency should, at best, be preserved, while transparency should always be provided about how certain choices are presented.

## 4.5 Identifying Research Areas and Research Questions

Mohamed Khamis (*University of Glasgow, UK, me@mkhamis.com*)

License  Creative Commons BY 4.0 International license  
© Mohamed Khamis

**Objective** This session aimed to identify specific research questions that could be addressed through joint research projects and initiatives.

**Methodology** Based on the grand challenges and discussions from the first day of the seminar, Mohamed Khamis synthesized different areas of research participants could vote on. Afterward, breakout groups were built where people identified specific research questions based on their interests.

We compiled a list of *main research areas* based on the outcomes of previous sessions. Participants then voted on which areas they would like to explore. The main areas were:

1. Social engineering vulnerability: self-assessment and misconceptions
2. Organizational changes to defend against social engineering
3. Frictions and warnings: How? When? What?
4. Evaluation of solutions: threats to validity
5. Evaluation metrics
6. Datasets

All areas that received four or more votes proceeded to the next stage, in which the participants chose the area they were interested in exploring further to produce research questions that can be addressed by (a) a Ph.D. thesis, (b) a research grant, or (c) a dedicated research center.

## Research Areas

The participants voted for the following research areas:

**Evaluation of Solutions** The breakout group discussed methodological challenges of evaluating solutions against social engineering attacks.

**Finding and Supporting Routines Against Attacks** This breakout group discussed research questions related to the development of routines, aiming at minimizing risks from social engineering attacks.

**Datasets** This breakout group aimed to identify research questions that could be answered by having access to different practitioners' datasets.

## Research Questions

### 4.5.1 Evaluation of Solutions

This breakout group focused on the methodological challenges of evaluating solutions against social engineering attacks, namely phishing and vishing. The group discussed and identified the following research questions.

**How can we collect data on successful and unsuccessful attacks?** One of the major challenges is the limited access to realistic, ecologically valid data. Access to such data is restricted due to legal and privacy constraints. Additionally, users may exhibit reporting bias, i.e., they may be hesitant to report social engineering attacks, whether successful or unsuccessful. Accordingly, the data collected may be skewed, and the true impact of social engineering attacks may be underestimated. While researchers try to overcome this challenge by relying on simulated or controlled environments, this usually does not fully capture the actual attacks.

**How can we identify and support protection strategies among real users?** Another interesting research question is how to crowd-source protection strategies from users' common practices. While reporting bias is one challenge, another entailed challenge is how to develop protection strategies that are adaptable and customizable to the diverse needs and characteristics of different user groups. Furthermore, establishing effective feedback loops for users to report social engineering attacks or provide input on protection measures and strategies might be methodologically challenging.

**How can we run evaluations in different attack phases?** Participants categorized the solution space into three phases: pre-, during-, and post-attack. This categorization introduces a set of research questions: How can attacks be reliably detected? What are effective intervention designs for the different phases? What are the appropriate research methods for each phase? When are solutions most effective?

### 4.5.2 Finding and Supporting Routines Against Attacks

While much knowledge and tools exist that can help protect users from social engineering attacks, a prevalent challenge is establishing secure routines. An example is using a password manager whenever logging into a website, as, in this way, links to fake phishing websites would be easily identified. The group identified the following research questions.

**What is the role of routines?** As a first step, researchers could explore the role routines could play in users' everyday lives.

**Which routines work across contexts?** A challenge is that routines (think about using VPNs) might work in one context, for example, during a business trip – but not in other contexts, such as being on vacation. The development of routines would benefit from an in-depth understanding of which routines work across contexts and which do not.

**How to design cues/reminders of risks/security behavior? How to communicate them?**

To support the development and habituation of routines, an interesting question is how users can be reminded of them, particularly as they are about to behave insecurely/riskily.

**How can the community support routines?** Another interesting question is the role of the community, particularly the question of how the fact that a community agreed on certain routines would affect the individual.

**What are easy routines?** Some routines are easier to habituate than others. Identifying easy routines would be valuable information to support self-efficacy. Routines that are easy for one user might not be easy for another user.

**How can routines be supported through AI?** Participants of the breakout groups also discussed the question of whether routines could be supported through AI, for example, models that predict opportune moments.

**How can developing a security mindset be supported?** An interesting question is how the gradual establishment of routines might ultimately lead to a “security mindset” among users and whether this makes adopting routines for other security contexts more likely.

### 4.5.3 Data Sets

The breakout group on data sets identified questions that could be answered as researchers have access to (historical and real-time) information on threat actor communication, traceable transactions on the dark web, knowledge of the organizations of these actors, and observable actions (e.g., increasing network traffic towards potential victims).

**What are observable attacker movements?** First, a comprehensive understanding of attacker actions, their characteristics, and how they could be tracked would be interesting.

**How can we associate movements with attackers?** A current challenge is linking observable movements with particular threat actors, as these are often difficult to identify (due to using TOR, VPNs, etc.). At the same time, close temporal or spatial proximity might hint at movements being associated with particular threat actors, allowing a more comprehensive picture to be drawn.

**How can predictive models for attackers from “movement sequences” be built?**

Researchers were particularly excited about the ability to not only understand what common sequences of movements are but to, based on this knowledge, predict the next steps of attackers. This would give potential victims time to prepare their defense and expect attacks.

**What are opportune moments to intervene?** From a practical point of view, an interesting question is when to intervene; that is, when to approach and warn potential victims. A predictive model might hint at a particular time window in which an attack is likely.

**What should interventions look like?** As an attack is likely, an interesting question is how to intervene. Should potential victims be sensitized? Or should they be trained through fake campaigns to (at least temporarily) improve their detection skills?

**When should a pen test be run?** Along the same lines, an interesting question could be when to launch a penetration test so as to test defenses.

**How can we identify opportunities for attackers (victims are unaware of)?**

Participants found the idea of learning more about the attackers and what opportune moments they exploit. In that way, a better understanding of threats can be obtained that might help victims develop better routines that minimize opportunities for attackers.

## 4.6 Towards Collaborations in Social Engineering Research

Florian Alt (University of the Bundeswehr – Munich, DE, [florian.alt@unibw.de](mailto:florian.alt@unibw.de))

License  Creative Commons BY 4.0 International license  
© Florian Alt

**Objective** This session aimed to identify specific topics and areas of collaboration.

**Methodology** For this session, a poster was created for each research question participants identified in the previous session. Then, participants were asked to indicate their interest in each research question by writing their names on the poster. Afterwards, in several rounds, participants met at the poster to discuss the following questions: (1) Who would fund this research? What would be the scope of a project? (3) How would you pitch the topic (i.e., write an abstract)?

The following list describes the different research projects and initiatives.

### A European Research Center for Awareness, Detection, and Mitigation of Social Engineering

**Funding** NL: NCSC-NL (National Cyber Security Center), NCTV (National Coordinator For Counterterrorism and Security, Ministry of Justice and Security), Cyberveilig Nederland, HighTechCrime Police/Europol (i.e., Law Enforcement), AIVD (General Intelligence and Security Services, Ministry of Interior)

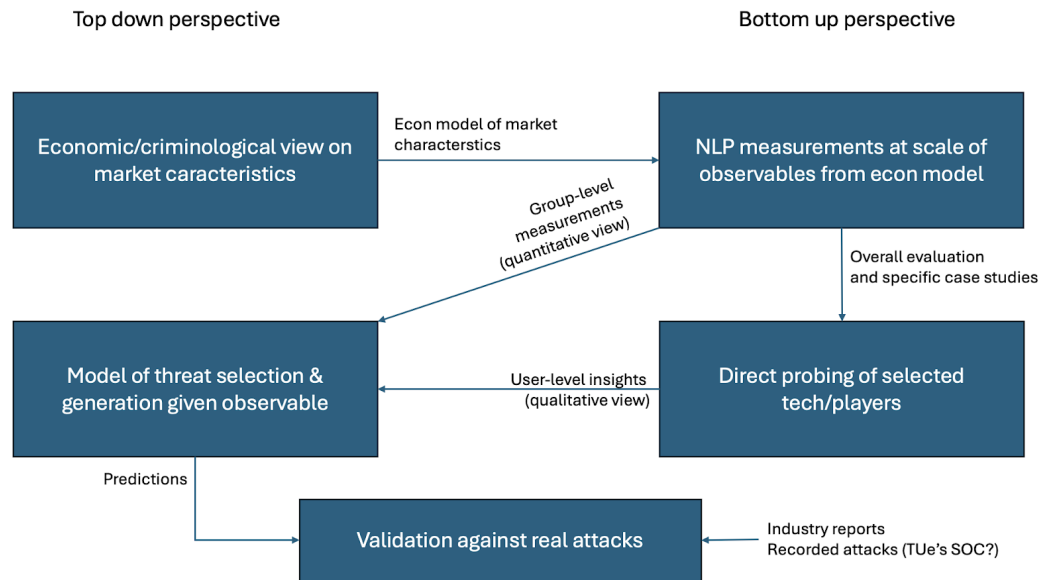
FR: ANSSI (National Cyber Security Center), Viginum (Service of vigilance and protection against foreign digital influence), Inria (National Research Institute on Informatics and applied mathematics, CNRS (National Center for Scientific Research).

DE: BSI (Bundesamt für Sicherheit in der Informationstechnik), BMI (Bundesministerium des Inneren), Cyberagentur; BKA (Bundeskriminalamt)

**Pitch** Social engineering attacks have emerged as a predominant threat, exploiting human psychology to manipulate individuals into revealing confidential information, compromising their financial security, and influencing their decision-making, including voting behavior. These tactics bypass traditional cybersecurity measures by directly targeting the most vulnerable link in the security chain: the human. Recognizing the gravity and complexity of this issue, the proposed *enter title here* aims to serve as a pioneering institution dedicated to combating these threats across Europe.

The center will be a collaborative hub, uniting industry practitioners and academic experts to develop comprehensive strategies against social engineering attacks. Its primary objectives will include raising awareness about the nature and tactics of these attacks, enhancing the ability to detect and identify such threats promptly, and devising effective mitigation strategies to reduce their impact. By fostering a multidisciplinary approach, the center will address current challenges and anticipate emerging trends in social engineering, such as the threat of generative AI, ensuring a proactive stance against these evolving threats.

This research center will formulate goals to fortify individual privacy, financial integrity, and democratic processes against the influence of social engineering. The establishment of this research center represents a significant step forward in strengthening Europe's resilience against these sophisticated psychological attacks, thereby protecting its citizens and institutions in an increasingly interconnected world.



■ **Figure 1** The envisioned approach to identify trends of emerging threats and attacker patterns.

## Observing / Modeling Attacker Movements

**Funding** DoD (Department of Defense), GCHQ (Government Communications Headquarters), EU, DFG (German National Research Foundation), Cyber Insurances

**Scope** Creating threat actor ecosystem by 1) identifying trends of emerging (credible) threats and patterns of attacker interest in different attack capabilities made available in the darknet. 2) Characterize specific and emergent criminal convergence spaces from internet forums/anonymous markets to Telegram/Discord channels. 3) Quantify/qualify the effect of the appearance of an offensive capability in one of those venues for realizing an attack.

**Pitch** Social engineering attacks follow a pattern and defined phases (see 4.2). The pre-attack phases involve OSINT, target selection, and attack planning. This project aims to model the attacker patterns to predict social engineering attacks even before happening (i.e., prevention instead of mitigation). The overarching goal is developing an investigation infrastructure that triggers based on detecting attack patterns.

As shown in Figure 1 the approach would cover different aspects. Define economic and criminological theoretical underpinnings to identify measurables within forum and telegram/discord channels to characterize communities regarding the type of support (e.g., moral hazard/adverse selection mitigation mechanisms) they provide to criminal activities. NLP topic analysis will be used to identify both discussion topics within communities and user perception/feedback related to attack technology (sentiment analysis) to characterize user interactions at scale. Language and slang challenges must be addressed, especially on less verbose channels like instant messaging platforms. Active probing with direct interaction (either as potential customers/providers) or face-to-face interviews (remote setting) with offenders/perspective offenders to investigate underlying decision mechanisms/factors (e.g., why joining community x/choosing product y to do offense x rather than product y'). Develop a model of threat selection. Relate model predictions to high-level trends in emerging attack tech with known incidents and see if there is a credible link between what these markets enable and what attackers do.



Based on this *Threat Actor Ecosystem*, attacker movement would be modeled by:

1. Identifying patterns of attacker actions within network monitoring data.
2. Coding techniques on network packets/sequencing and MITRE ATT&CK mapping to evaluate qualitatively attack processes.
3. Building the conceptual model from that understanding.
4. Developing a data model for detecting those patterns from network data for scalability.
5. Evaluating the correlation of historical trends in those attack patterns with trends from our approach.

### Live Threat Detection and Intervention (e.g., vishing)

**Funding** NSF (National Research Foundation), DFG (German National Research Foundation), Security companies

**Scope** Social Engineer's vishing dataset or self-data collection; project might include other modalities (keystroke dynamics, etc.)

**Pitch** An increasingly popular form of social engineering attacks is vishing (voice phishing). Little effective means for protection exist today against such attacks beyond raising awareness in cyber education. At the same time, the human voice holds rich information about (1) the current user state (i.e., whether they are stressed and what their current level of awareness is) as well as (2) techniques in use to social engineer somebody (firm voice to sound authoritative, pleading tone to beg for help, etc.). This project proposes to design, implement, and evaluate in-situ interventions protecting users from falling for vishing, that is, mechanisms that are capable of detecting in real-time if a caller is trying to social engineer somebody or if the callee is being socially engineered and provide active guidance as to how the legitimacy of a call can be verified.

The project will address the following objectives:

- (1) building predictive models based on real vishing data, allowing common manipulation strategies and callee reactions to be detected;
- (2) designing, implementing, and evaluating interventions to assist end users during vishing calls;
- (3) assessing social and ethical implications of vishing mitigation technologies and strategies.

### Using AI to support sharing content

**Funding** Social Media Platforms, Security Companies, and Research Foundations.

**Pitch** Social network users share personal information online that might be misused in several ways incl. social engineering. In this project, we want to investigate the usage AI-based support to inspect the information users want to share for (a) identifying content that could be misused or does not match the user's privacy needs, (b) educating the user on sharing consequences and (c) helping the users avoiding to share such information in the future, and (d) modify the content to mitigate the probability of misuse.

## Using AI to design personalized support to mitigate vulnerabilities toward social engineering attack

**Funding** Research Foundations and Cyber Security Companies.

**Pitch** This funding proposal seeks support for an innovative project to leverage artificial intelligence (AI) to design personalized support systems that effectively mitigate vulnerabilities towards social engineering attacks. The proposed initiative recognizes the multifaceted nature of susceptibility, focusing on personability, age, cultural differences, and accessibility needs, among others, as critical dimensions to tailor interventions. The advent of sophisticated social engineering attacks demands a nuanced approach that adapts to individual characteristics. Our project will employ advanced AI algorithms to analyze and understand diverse user profiles to generate tailored responses and interventions resonating with users personally and build resilience to manipulation attempts.

The following ideas were identified but not further discussed:

- Utilizing AI for routine building and providing support
- Plugins and feature highlighting
- Designing social engineering interventions
- AI-based validation
- Self-assessment and self-reflection to mitigate vulnerabilities
- Design targeted support for vulnerable groups
- Understanding routines to mitigate vulnerabilities
- Characteristics of targets and social engineers

## 5 Report Summary

This report documents the outcomes of a three-day seminar that brought together experts from academia, industry, and authorities to address the escalating threats posed by social engineering in the digital age. The seminar aimed to develop a common understanding of social engineering, identify grand challenges, work on a research agenda, and foster collaboration in addressing social engineering vulnerabilities. Key themes included the professionalization of cyber attacks, the proliferation of sensors in everyday life as an opportunity for developing protection solutions, and the need for interdisciplinary collaboration to address evolving social engineering threats.

The seminar featured various sessions, including keynote speeches, group activities, and breakout discussions. Key discussions revolved around the fundamentals of social engineering, attack vectors, and the application of modern sensors to assess user states affected by social engineering. The participants also explored the ethical boundaries and obligations when using social engineering techniques for positive individual and social outcomes.

Furthermore, the seminar identified specific research areas and questions to be addressed through joint research projects and initiatives. Research areas included the evaluation of solutions, finding and supporting routines against attacks, and using datasets to understand threat actor communication and traceable transactions on the dark web.

The seminar outcomes underscore the importance of collaborative efforts between researchers and practitioners in fortifying against evolving social engineering threats. The insights gained from the seminar lay the foundation for future research and initiatives in addressing psychological manipulation in the digital age, including using modern sensors, ethical considerations in social engineering, and the development of protective measures against social engineering attacks.

Overall, the seminar provided a platform for interdisciplinary collaboration, fostering a deeper understanding of social engineering and its cognitive vulnerabilities. The identified grand challenges and proposed research projects highlight the significance of collaborative efforts in addressing the emerging threats posed by social engineering in the digital realm. The seminar outcomes provide valuable insights and potential research directions for fortifying against psychological manipulation and cyber threats.

## 6 Reading List

### References

- 1 Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Yang Wang, and Shomir Wilson. “Nudges for Privacy and Security: Understanding and Assisting Users’ Choices Online.” *ACM Comput. Surv.* 50, no. 3 (May 2018): 44. <https://doi.org/10.1145/3054926>
- 2 Florian Alt, Mariam Hassib, and Verena Distler. “Human-centered Behavioral and Physiological Security.” In *Proceedings of the 2023 New Security Paradigms Workshop (NSPW ’23)*, 2023: 48–61. <https://doi.org/10.1145/3633500.3633504>
- 3 Nattapat Boonprakong, Xiuge Chen, Catherine Davey, Benjamin Tag, and Tilman Dingler. 2023. Bias-Aware Systems: Exploring Indicators for the Occurrences of Cognitive Biases when Facing Different Opinions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI ’23)*. Association for Computing Machinery, New York, NY, USA, Article 27, 1–19. <https://doi.org/10.1145/3544548.3580917>
- 4 Lina Brunken, Annalina Buckmann, Jonas Hielscher, and M. Angela Sasse. (2023). “To Do This Properly, You Need More Resources”: The Hidden Costs of Introducing Simulated Phishing Campaigns.
- 5 J-W. Bullée, and M. Junger. “How effective are social engineering interventions? A meta-analysis.” *Information and Computer Security* 28, no. 5 (2020): 801–830. <https://doi.org/10.1108/ICS-07-2019-0078>
- 6 J.-W. Bullée, L. Montoya, W. Pieters, M. Junger, P. Hartel. “On the anatomy of social engineering attacks – A literature-based dissection of successful attacks”. *J Investig Psychol Offender Profil.* 2018; 15: 20–45. <https://doi.org/10.1002/jip.1482>
- 7 Pavlo Burda, Luca Allodi, and Nicola Zannone. “Cognition in Social Engineering Empirical Research: a Systematic Literature Review.” 2023. <https://doi.org/10.1145/3635149>
- 8 Felix Dietrich, Pascal Knierim, Yomna Abdelrahman, Ahmed Shams, Ken Pfeuffer, Mariam Hassib, and Florian Alt. “The Influence of Context on Response to Spear-Phishing Attacks: an In-Situ Deception Study.” In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI ’23)*, 2023: 619. <https://doi.org/10.1145/3544548.3581170>
- 9 Nina Gerber and Karola Marky. 2022. The nerd factor: the potential of S&P adepts to serve as a social resource in the user’s quest for more secure and privacy-preserving behavior. In *Proceedings of the Eighteenth USENIX Conference on Usable Privacy and Security (SOUPS’22)*. USENIX Association, USA, Article 4, 57–76.
- 10 Martina Angela Sasse and Iacovos Kirlappos, “Security Education against Phishing: A Modest Proposal for a Major Rethink” in *IEEE Security and Privacy*, vol. 10, no. 02, pp. 24–32, 2012. <https://doi.ieeecomputersociety.org/10.1109/MSP.2011.179>
- 11 Karola Marky, Martin Schmitz, Verena Zimmermann, Martin Herbers, Kai Kunze, and Max Mühlhäuser. 2020. 3D-Auth: Two-Factor Authentication with Personalized 3D-Printed Items. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing*

- Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3313831.3376189>
- 12 Karola Marky, Shaun Macdonald, Yasmeen Abdrabou and Mohamed Khamis. 2023. In the Quest to Protect Users from Side-Channel Attacks – A User-Centred Design Space to Mitigate Thermal Attacks on Public Payment Terminals. 32nd USENIX Security Symposium (USENIX Security 23). <https://www.usenix.org/conference/usenixsecurity23/presentation/marky>
  - 13 Florian Mathis, John H. Williamson, Kami Vaniea, and Mohamed Khamis. 2021. Fast and Secure Authentication in Virtual Reality Using Coordinated 3D Manipulation and Pointing. *ACM Trans. Comput.-Hum. Interact.* 28, 1, Article 6 (February 2021), 44 pages. <https://doi.org/10.1145/3428121>
  - 14 Peter Mayer, Yixin Zou, Byron M. Lowens, Hunter A. Dyer, Khue Le, Florian Schaub, and Adam J. Aviv. 2023. Awareness, Intention, (In)Action: Individuals' Reactions to Data Breaches. *ACM Trans. Comput.-Hum. Interact.* 30, 5, Article 77 (October 2023), 53 pages. <https://doi.org/10.1145/3589958>
  - 15 Simone Ooms, Minha Lee, Pablo Cesar, and Abdallah El Ali. 2023. FeelTheNews: Augmenting Affective Perceptions of News Videos with Thermal and Vibrotactile Stimulation. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 137, 1–8. <https://doi.org/10.1145/3544549.3585638>
  - 16 Justin Petelka, Yixin Zou, and Florian Schaub. 2019. Put Your Warning Where Your Link Is: Improving and Evaluating Email Phishing Warnings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, Paper 518, 1–15. <https://doi.org/10.1145/3290605.3300748>
  - 17 Christina Schneegass, Thomas Kosch, Andrea Baumann, Marius Rusu, Mariam Hassib, and Heinrich Hussmann. 2020. BrainCoDe: Electroencephalography-based Comprehension Detection during Reading and Listening. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376707>
  - 18 Steeven Villa, Thomas Kosch, Felix Grelka, Albrecht Schmidt, and Robin Welsch. 2023. The placebo effect of human augmentation: Anticipating cognitive augmentation increases risk-taking behavior. *Comput. Hum. Behav.* 146, C (Sep 2023). <https://doi.org/10.1016/j.chb.2023.107787>
  - 19 Melanie Volkamer, Martina Angela Sasse, and Franziska Boehm. 2020. Analysing Simulated Phishing Campaigns for Staff. In *Computer Security: ESORICS 2020 International Workshops, DETIPS, DeSECSys, MPS, and SPOSE, Guildford, UK, September 17–18, 2020, Revised Selected Papers*. Springer-Verlag, Berlin, Heidelberg, 312–328. [https://doi.org/10.1007/978-3-030-66504-3\\_19](https://doi.org/10.1007/978-3-030-66504-3_19)
  - 20 Yixin Zou, Kevin Roundy, Acar Tamersoy, Saurabh Shintre, Johann Roturier, and Florian Schaub. 2020. Examining the Adoption and Abandonment of Security, Privacy, and Identity Theft Protection Practices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. <https://doi.org/10.1145/3313831.3376570>

## Participants

- Yomna Abdelrahman  
European Universities in Egypt –  
Cairo, EG
- Luca Allodi  
TU Eindhoven, NL
- Florian Alt  
University of the Bundeswehr  
Munich, DE
- Nathan Berry  
Nexus – Leeds, GB
- Jan-Willem Bullee  
University of Twente, NL
- Mary D'Angelo  
Searchlight Cyber –  
Washington, DC, US
- Felix Dietz  
University of the Bundeswehr  
Munich, DE
- Tilman Dingler  
The University of Melbourne, AU
- Verena Distler  
University of the Bundeswehr  
Munich, DE
- Abdallah El Ali  
CWI – Amsterdam, NL
- Jerry Färdigs  
Swedish Armed Forces –  
Uppsala, SE
- Ann Fernström  
Swedish Armed Forces –  
Uppsala, SE
- Matteo Große-Kampmann  
AWARE7 GmbH –  
Gelsenkirchen, DE
- Christopher Hadnagy  
Social-Engineer – Orlando, US
- Mohamed Khamis  
University of Glasgow, GB
- Claude Kirchner  
CCNE – Paris, FR & INRIA –  
Rocquencourt, FR
- Thomas Kosch  
HU Berlin, DE
- Karola Marky  
Ruhr-Universität Bochum, DE
- Abbie Maroño  
Social-Engineer – Orlando, US
- Alexander Nussbaum  
University of the Bundeswehr  
Munich, DE
- Alia Saad  
Universität Duisburg-Essen, DE
- Martina Angela Sasse  
Ruhr-Universität Bochum, DE
- Florian Schaub  
University of Michigan – Ann  
Arbor, US
- Christina Schneegass  
TU Delft, NL

