# When Awareness Fades and There Is No Support, the Phisher Has an Easy Game

Zur Erlangung des akademischen Grades eines

**Doktors der Ingenieurwissenschaften (Dr.-Ing.)**

von der KIT-Fakultät für
Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

**DISSERTATION**

von

**M.Sc. Benjamin Maximilian Berens (geboren Reinheimer)**
geb. in Seeheim-Jugenheim

| | |
|---|---|
| Tag der mündlichen Prüfung: | 01.03.2024 |
| Hauptreferent: | Prof. Dr. Melanie Volkamer |
| Korreferent: | Prof. Dr. Florian Schaub |

# Abstract

Background: Phishing has plagued our digital communication for many years now. Every day more phishing attacks are perpetrated and more reports of phishing infiltration in businesses or private lives make headlines. Phishing presents a particularly insidious threat to communication because the difference between phishing and legitimate emails can be all but imperceptible, even to users who have received training in phishing prevention. Therefore, phishing awareness measures and tools need to improve in design and thus enable users to decide more effectively between legitimate and phishing massages. However, one major challenge to the effective prevention of phishing is the long-term impact of awareness interventions. Previously, there has been a gap in knowledge about how long phishing awareness measures last and what refresher awareness measures are potentially effective. Furthermore, more information is needed, whether phishing awareness measures or tool support is more effective and if a combination of both could further improve the user's decision process. This thesis tries to close these research gaps and is therefore separated in two parts: Part I is systematically analyzing the effectiveness of three phishing awareness measures over different periods of time, including refreshment and Part II is a systematic evaluation of a phishing awareness measure against and in combination with a tool support. Additionally, the strengths and weaknesses with regard to specific phishing tricks are analyzed for all interventions of Part I and II.

Methods: Overall, four online user studies have been conducted to evaluate three different phishing awareness measures (video, e-learning, and workshop), four refreshers (video, short text, long text, interactive e-mail example), and one tool support (link-centric warning). In Part I, three retention studies have been conducted, whereby in study 1 a short awareness measure (video) was implemented and tested over a period of eight weeks (N = 89) and in study 2, a more extensive e-learning was conducted and tested over a period of five months (N = 46), both as repeated measure design studies. Additionally, in study 3, a between-subject design was chosen to systematically evaluate a workshop over a period of twelve months with five retention time points (after 4, 6, 8, and 12 months) in a sample of N = 439 employees. Additionally, four refreshers (video, short text/poster, long text/leaflet, interactive e-mail example) were evaluated at the point in time, where the awareness gained through the workshop had no significant improvement anymore. In Part II, a phishing awareness measure (NoPhish video) and a tool support (TORPEDO link-centric warning) was systematically evaluated in a sample of N = 420 clickworker participants as individual and combined interventions over eight groups: two control groups (status quo: status bar and tooltip), two intervention groups with awareness measure NoPhish video (video + status bar, video + tooltip), two interventions groups with tool support (TORPEDO with and without tutorial), and two combined interventions (NoPhish video + TORPEDO with and NoPhish video + TORPEDO without tutorial). Across all studies, signal detection theory was applied to evaluate the effectiveness of the interventions according to the values sensitivity $d'$ and criterion $C$. Statistical analyses were dependent on the study design as repeated or single measures ANOVAs conducted with statistical program R.

Results: Overall, n = 856 participants were analyzed (n = 22 in study 1, n = 16 in study 2, n = 409 in study 3, and n = 409 in study 4). As a result of Part I, it was found in study 1 that the trained awareness of the video was still significant after eight weeks and that in study 2, the gained awareness through the e-learning was still significantly improved after five months. The evaluation in study 3 revealed that the trained awareness by the workshops drops back to the starting level between the fourth and sixth month. However, it could be found that three out of four refreshers (video, long text/leaflet, and interactive e-mail example, but not the short text/poster) had been successful to refresh the lost awareness after six months. With the help of the three refreshers, the increased awareness lasted another 6 months until the evaluation after 12 months. In addition, different phishing tricks were evaluated separately across the three studies. In particular, it was found that the phishing trick "Small Deviations in the Domain“ was difficult to detect across different phishing awareness measures. The results of Part II showed that a video as a phishing awareness measure and also a link-centric warning as a tool support both significantly improved sensitivity when compared to status-quo measures and tools alone. The link-centric warning and the accompanying tutorial outperformed the video, while the tutorial played an essential role for the link-centric warning to be most effective. The combination of tool support with tutorial and video achieved the best sensitivity values overall. Additionally, it is only this combination that achieves near-optimal effectiveness with regard to the dangerous phishing trick "Small Deviations in the Domain“ category.

Conclusion: This work provides important insights into the effectiveness of phishing awareness measures over time and how the effect of phishing awareness measures could be maintained by refreshers or further improved by the combination with phishing tool support. This evidence-based recommendations will help users to make informed decisions in the face of phishing threats and help to decide when to plan refresher training and in which form to give that training: Firstly, the workshop and the e-learning as interactive measurements performed very well and the awareness lasted for a period of four to six months till a refresher was needed. Secondly, a combination of a short video with a phishing tool support provided excellent results. The success of the video must be particularly emphasized, as it is a very short five-minute measure that is therefore easy to implement. Additionally, the phishing trick “Small Deviations in the Domain” seemed to be the most difficult to detect and should therefore be given greater consideration in phishing awareness measures. Overall, these findings can make a difference for both the advancement of the research about phishing prevention as well as for organizations and everyday users coping with the ever-evolving threat of phishing.

# Acknowledgements

Als Erstes möchte ich mich an dieser Stelle bei meiner Erstbetreuerin Prof. Melanie Volkamer bedanken. Du hast mich immer besonders gefördert und unterstützt und ich konnte durch deine enge Betreuung enorm profitieren. Vor allem die vielen Diskussionen zu den verschiedensten Themen haben zu zahlreichen spannenden Ideen in der Vergangenheit, aber auch für die Zukunft, geführt. Zudem möchte ich mich besonders für die vielen eindrücklichen Erlebnisse (Konferenzbesuche, Auslandsaufenthalte oder Summerschool) auf dem Weg der Dissertation bedanken, die den Weg bereichert und die Motivation immer wieder gesteigert haben.

Außerdem danke ich meine Eltern (Gaby und Rudolf), die mich schon mein ganzes Leben unterstützen und auch wenn der Weg nicht immer geradlinig verlaufen ist - doch immer hinter mir und meinen Entscheidungen gestanden haben. Vor allem habt ihr auch einen entscheidenden Einfluss, dass ich überhaupt meine erste Stelle damals noch an der Technischen Universität in Darmstadt angetreten habe.

Ebenfalls möchte ich auch allen meinen Kollegen danken, die mich in der Zeit in der Forschungsgruppe begleitet haben. Im Besonderen nenne ich hier meine Kollegen Peter und Nina, mit denen ich zusammen gestartet bin. Wir haben vieles in der Zeit erlebt und gelernt und vor allem konnte ich auch viel von euch lernen.

Auch meine (Schul)-Freunde sollen an der Stelle nicht unerwähnt bleiben. Ihr habt meine gesamte Zeit miterlebt und musstet euch die ein oder andere Klage, aber auch schöne Geschichten anhören. Besonders möchte ich hier auch meinen ehemaligen Kollegen und mittlerweile Freund Marco erwähnen, mit dem ich die ein oder andere Pause dafür genutzt habe, Bouldern zu gehen.

Abschließend möchte ich mich bei meiner Frau Sabrina bedanken. Du begleitest mich jetzt schon meine gesamte akademische Karriere, angefangen vom ersten Studium bis hin zum Ende meiner Doktorarbeit. Auch wenn die Zeit mal wieder knapp geworden ist für eine Abgabe oder Arbeit, haben wir immer zusammengehalten und ich konnte jedes Problem auch mit deiner Stärke bewältigen.

# Contents

**Abstract** . . . . . . . . . . **i**

**Acknowledgements** . . . . . . . . . . **v**

**Acronyms and symbols** . . . . . . . . . . **xi**

**1 Introduction** . . . . . . . . . . **1**

**2 Related Work** . . . . . . . . . . **3**

2.1 Phishing Definition . . . . . . . . . . 3

2.2 Awareness Retention . . . . . . . . . . 4

2.3 Security Interventions . . . . . . . . . . 6

**3 Background** . . . . . . . . . . **11**

3.1 Phishing Tricks . . . . . . . . . . 11

3.2 NoPhish Awareness Measure . . . . . . . . . . 15

3.2.1 General Idea & Concept . . . . . . . . . . 16

3.3 Implementation of the Awareness Measures for NoPhish . . . . . . . . . . 17

3.3.1 Video . . . . . . . . . . 18

3.3.2 E-Learning . . . . . . . . . . 20

3.3.3 Workshop . . . . . . . . . . 22

3.3.4 Leaflet . . . . . . . . . . 24

3.3.5 Poster . . . . . . . . . . 27

3.3.6 Interactive Example . . . . . . . . . . 28

3.4 TORPEDO . . . . . . . . . . 29

3.4.1 General Idea . . . . . . . . . . 29

3.4.2 Risk Levels . . . . . . . . . . 30

3.4.3 Tutorial . . . . . . . . . . 32

3.5 Dimensions of Study Examples . . . . . . . . . . 33

3.6 Signal Detection Theory . . . . . . . . . . 34

**4 Part I: Long-term Effects of Phishing Awareness Measures** . . . . . . . . . . **39**

4.1 Video . . . . . . . . . . 41

4.1.1 Research Questions and Hypotheses . . . . . . . . . . 42

4.1.2 Methodology . . . . . 43
4.1.3 Results . . . . . 47
4.2 E-Learning . . . . . 54
4.2.1 Research Questions and Hypotheses . . . . . 55
4.2.2 Methodology . . . . . 56
4.2.3 Results . . . . . 63
4.2.4 Limitations . . . . . 68
4.3 Workshop . . . . . 69
4.3.1 Research Questions and Hypotheses . . . . . 71
4.3.2 Methodology . . . . . 72
4.3.3 Results . . . . . 79
4.3.4 Limitations . . . . . 88
4.4 Discussion Part I . . . . . 92
4.4.1 Research Questions . . . . . 92
4.4.2 Future Work . . . . . 98
4.4.3 Overall Limitations . . . . . 100

**5 Part II: Phishing Awareness and Tool Support: A Synergistic Approach . . 105**
5.1 Research Questions and Hypotheses . . . . . 106
5.2 Methodology . . . . . 108
5.2.1 Study Design . . . . . 110
5.2.2 Study Material . . . . . 114
5.2.3 Recruitment & Ethical / Data Protection Considerations . . . . . 117
5.3 Results . . . . . 120
5.3.1 RQ1 - Video Effect . . . . . 121
5.3.2 RQ2 - Link-Centric Warning Effect . . . . . 121
5.3.3 RQ3 - Video vs. Link-Centric Warning . . . . . 121
5.3.4 RQ4 - Combination of Video and Link-Centric Warning . . . . . 122
5.3.5 RQ5 - Phishing Tricks . . . . . 122
5.4 Discussion Part II . . . . . 125
5.4.1 Research Questions . . . . . 125
5.4.2 Future Work . . . . . 127
5.4.3 Limitations . . . . . 128

**6 Overall Discussion and Outlook . . . . . 131**
6.1 Thesis Discussion . . . . . 131
6.2 Final Remarks . . . . . 132

**A Appendix . . . . . 135**
A.1 All pages of the leaflet . . . . . 135
A.2 All pages of the tutorial . . . . . 138

**List of Figures** . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . **141**

**List of Tables** . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . **143**

**Bibliography** . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . **147**

# Acronyms and symbols

## Acronyms

| | |
|---|---|
| **KIT** | Karlsruher Institute for Technology |
| **OS** | Operating System |
| **SDT** | Signal Detection Theory |
| **C** | Criterion (from Signal Detection Theory) |
| **d′** | Sensitivity (from Signal Detection Theory) |
| **HTTP** | Hypertext Transfer Protocol |
| **HTTPS** | Hypertext Transfer Protocol Secure |
| **CISM** | Chief Information Security Manager |
| **PoC** | Persons of contact |
| **InfoSecs** | Information security concerns |
| **SOGGS** | State Office for Geoinformation and State Survey |
| **GDPR** | General Data Protection Regulation |

# 1 Introduction

Preventing phishing messages from reaching people's inbox by technical measures is important and constantly improved. However, filtering suspicious messages does not catch all phishing attempts [1, 2], leaving users somewhat vulnerable if they cannot effectively detect such messages. In real world, two approaches exist to support the user: 1) creating awareness and 2) tool support while deciding. While several phishing awareness measures have been proposed and shown in evaluations to be effective (e.g., games [3] or text material [4]), less is known about how long this effect lasts and whether awareness can be guaranteed for longer with short refreshers. The most promising tool support approaches are "just in place and time" security interventions with tooltips (link centric warnings) right at the point of focus of the user [5]. To sum up, according to the first approach, little is known about how long phishing awareness measures last and what effective refreshing awareness measures are. Furthermore, according to the second approach, more information is needed, whether phishing awareness measures or tool support is more effective and if a combination of both could further improve the user's decision process. This thesis tries to narrow these gaps by: 1) systematically analyzing three measures in three different contexts, including refreshment (Part I) and 2) a systematic evaluation of a short phishing awareness measure (video) against and in combination with a tool support (link-centric warning) (Part II). Additionally, the strengths and weaknesses for the different measurements of Part I as well as for the individual interventions and the combination in Part II with regard of specific phishing tricks are analyzed. The overarching goal of this thesis is to evaluate different phishing awareness measures and tool support in terms of their effectiveness, to raise users' phishing awareness and to support them in their decision-making, to gain insights into the sustainability of the raised awareness and to identify suitable times and measures for refreshers. To achieve this, a total of four studies were conducted, based on various awareness measures implementing the NoPhish concept (more in Section 3.2) and the TORPEDO based tool support (more in Section 3.4). My contribution from Part I "Long-term Effects of Phishing Awareness Measures" are the following:

- Even a very brief measure (e.g, an anti-phishing awareness video of five minutes) significantly improved users' ability to distinguish between phishing and legitimate messages (called user sensitivity) for a two-month period [6].

- A more extensive and interactive measure (e.g., an anti-phishing awareness e-learning), which included exercises and lasted around two hours, significantly improved user sensitivity over a period of at least five months [7].
- Even a very brief measure (e.g, an anti-phishing awareness video of five minutes) significantly improved users' ability to distinguish between phishing and legitimate messages (called user sensitivity) for a two-month period [6].
- An anti-phishing workshop with the train-the-trainer approach also including exercises with a duration of around three to four hours significantly improved user sensitivity for a period of around four to five months [8].
- When user sensitivity is no longer significantly improved three of four evaluated refresher measures again achieved significant improvements. Furthermore, these three refresher measures achieved significant improve awareness for a period of at least twelve months [8].
- In two out of three of the studies, the lowest number of correct answers are measured for the phishing trick “Small Deviations in the Domain” (e.g., `lufthansa.com`). Therefore, the most important focus for future anti-phishing interventions should be on this difficult-to-identify phishing trick.

My contribution from Part II "Phishing Awareness and Tool Support: A Synergistic Approach" are the following:

- Confirmation of previous research on a phishing awareness measure (video) and a phishing tool support (link-centric warning) to be effective in increasing the sensitivity in a single, between-subjects study design [9].
- Comparison between a phishing awareness measure (video) and a phishing tool support (link-centric warning) showed the tool support to be a better single intervention [9].
- Combination of a phishing awareness measure (video) and a tool support (link-centric warning) has proven to be more effective than a single intervention [9].
- Evaluation of specific phishing tricks revealed that only the combination achieved near-optimal effectiveness with regard to the phishing trick „Small Deviations in the Domain“ category [9].

# 2 Related Work

In this chapter, the previous research in the area is presented. The focus is on the two core elements of this dissertation, i.e., awareness retention and support through security intervention. First, it is located where in the previous phishing literature this work ties in. Afterwards, the focus is on the current status quo of awareness retention. The chapter closes with an overview of the security intervention research that has previously been conducted.

## 2.1 Phishing Definition

Phishing has been the subject of many studies in the past. Nevertheless, there are slight differences in the different definitions of what (still) belongs to phishing. Some consider phishing more in the field of website investigation, i.e. a person has already clicked on a link and landed on a website [10, 11]. The website then tries to obtain sensitive data, e.g. passwords or entire account data, and has created a particularly genuine-looking website for this purpose. Others consider phishing one step ahead, namely what methods attackers use to get a person to click on a link, open an attachment or perform an action that is actually unwanted [12, 13]. The focus of the thesis includes fraudulent links and, at least to a small extent, the topics of unwanted actions and executable files. Accordingly, the choice falls on the broader definition of phishing and thus on the assessment of messages (e.g. emails) and not websites.

In the previous literature, both phishing [14] or fraudulent messages [15] are used as synonyms. Other terms, such as scam or spam, are sometimes used in the context of unwanted messages, e.g., emails in the inbox. When the definition of phishing messages becomes very broad, one could also use the term fraudulent message as an equivalent. In such cases where it can be used as equivalent, there are reasons to use both phishing as well as fraudulent to describe those messages. Phishing, which is more prominent in media reports, can be priming people into a particular mindset where "fraudulent" is more abstract in describing a characteristic of the message. In this thesis, some studies use the term phishing, and some use the term fraudulent when participants are asked to categorize a message. Both are used as synonyms and, therefore, defined as mentioned above.

## 2.2 Awareness Retention

In the context of cyber security, there is no standard definition of awareness [16, 17]. However, current overviews conceptualize awareness as a multidimensional construct, that combines perception, understanding, as well as behavioral parts [16, 18] and is composed of elements of threat and coping appraisals [17, 19]. Therefore, in this thesis, phishing awareness combines 1) the knowledge of common phishing threats and their consequences 2) the attention on potential phishing threats 3) the understanding, how the threat potential can be evaluated, and 4) the applicability of coping behaviors for responding to these threats. In this sense, the awareness measures and tools should not only impart knowledge about phishing, but also draw attention to the relevant cues (e.g., URL), help with evaluation and provide instructions for action.

Creating awareness is an important part of protecting people from danger and its consequences. However, it is also known from many sources that knowledge "fades" again over time [20]. The rate at which this occurs depends on many factors. For example, whether knowledge is refreshed over time (see Table 2.1 for an overview of the different studies), whether knowledge is applied in everyday life or used only sporadically, and also what kind of knowledge it is [21]. Accordingly, it was examined in different areas, how long the knowledge lasts and in which forms the knowledge decreases, e.g., in the area of economics [22], basic education [23] or also management [24]. In the field of cyber security, there are also already many study to create knowledge. But how long this knowledge remains, that is not yet in detail examined. There are single studies, which showed that it still exists at different times. From these studies, however, it is not yet possible to deduce a point in time at which the knowledge should be renewed. In principle, knowledge can also be renewed in fixed periods of time without the knowledge. Especially in the organizational context, however, the cost/benefit question is raised at this point. Organizations are in a field of tension here that on the one hand they are obliged to a certain interval, e.g. by legal regulations, and on the other hand they want to find an optimal interval for themselves. They want to maintain a certain level and yet not invest excessive resources in constant training.

There have been already some initial indications of how much refresher or retraining is needed or what this should look like in order to reach the same level again. Evidence suggests that less than half the original training time is sufficient to return to the same level [25]. Regarding the form of the refresher, it seems that the simplicity of delivery has a moderating influence on how well the refresher works [26]. So possibly a simpler measure helps better as a refresher than a very complicated one. So it could be helpful if a complicated topic is first trained with an extensive and thus more complex measure in order to retrain the knowledge again with a simple measure afterwards. Repetitions for the sake of repetitions, i.e. in a too strong accumulation have also turned out not to be beneficial [27]. The interval between repetition or training has an important influence on effectiveness [27]. There must be enough time in between to be able to process

**Table 2.1:** Overview of awareness retention studies with first author, retention period, result (either * for significant or n.s. for non significant) and special characteristic of the study. “Significance” means that after the measure, the tested period still shows a significant result compared to before.

| Literature | Retention Period | Significance | Special |
|---|---|---|---|
| Zhang et al. [21] | A week | n.s. | Students, control group plus multiple training over six weeks (between subject) |
| Jensen et al. [29] | Ten days | * | Students, control group plus different forms of training (between subject) |
| Lastdrager et al. [30] | 14, 28 and 64 days | n.s. (28 days) | Pupils, control group plus one training (between subject) |
| Kumaraguru et al. [31] | 28 days | * | University member, control group + two training (between subject) |
| Nguyen et al. [32] | Eight weeks | * | Students, control group plus different forms of training (between subject) |
| Canova et al. [33] | Five months | * | Students, one training (within subject) |
| Berens et al. [34] | Five months | (*) | Panel participants, two control groups and two with measure (between subject) |

the knowledge. As an extension, it has been found that repetition with the same thing over a short period of time is also extremely ineffective [28]. Therefore, the distance turns out to be an important criterion, but also the form of repetition. A new form can possibly bind the attention better, set new accents and thus lead to a more effective refreshment.

Especially in the area of phishing awareness, there have been a few studies that have looked at the time period over which awareness is retained. However, many studies have rather dealt with short periods of time. Most of the studies deal with a period up to about two months. There have been studies on the effectiveness of mindfulness on the long-term effect [29], in which a period of ten days was considered. Over the ten-day period here, mindfulness training was better than rule-based training and both were better than no training [29]. A short five minute SETA program (security education, training, and awareness) was found to no significantly improve phishing detection of a single email sent to the participants after a week [21]. In this study for one of the groups the program was repeated six times over the span of six weeks and for the second group twelve times over the span of six weeks. Improvements through an instructed-based training of 40 minutes with trainers being specialized in cyber-security were also found with a sample of pupils. But these improvements, especially for phishing, returned to baseline levels over the four-week period [30]. Another study looked at long-term outcomes again with a focus on mindfulness vs. rule-based training where the training took around 1 hour. In there, identification tests were administered immediately after the intervention and ten weeks later [32]. In addition, a phishing message was sent after one week and after eight weeks. The mindfulness training proved to be significantly better after eight weeks [32]. Both for organization and for normal users a awareness retention of four weeks or even two months seems unrealistic. Especially as these were mostly still significant results, it should be checked which longer periods of time are still producing significant results. A study with an extensive android game where participants played for 30 minutes found that after five months after the gaming session the participants decreased in their

**Table 2.2:** Overview of security intervention studies with first author, kind of tool support, result (either * for significant or n.s. for non significant) and special characteristic of the study. “Significance” means that the tool support tested showed a significant result compared to either a control group or to a measurement before.

| Literature | Kind of tool support | Significance | Special |
|---|---|---|---|
| Schechter et al.[10] | Warning page | * | Lab study, https indicators, site-authentication image, warning page |
| Akhawe et al.[37] | Warning page | * | Field study, existing warnings firefox malware, google malware firefox ssl |
| Egelman et al.[38] | Warning page | * | Lab study, active vs. Passive warning, warning perception |
| Felt et al.[39] | Warning page | n.s. | Field study, new warning vs. existing warnings, warning comprehension |
| Sunshine et al.[40] | Warning page | * | Lab study, new warnings vs. existing warnings |
| Brustoloni et al.[41] | Tooltip | * | Lab study, active vs. Passive warning, warning perception |
| Stembert et al.[42] | Tooltip | * | Lab study, active vs. Passive warning, three different studies |
| Anderson et al.[43] | Warning page | * | Lab study, fMRI, polymorphic warning, habituation |
| Wu et al.[44] | Toolbar | * | Lab study, toolbars, different phishing tricks |
| Marchal et al.[45] | Banner | * | Lab and field study, add-on implementation, user preference comparison with existing warnings |
| Silic et al.[46] | Warning page | * | Online study, color appeal, perceived risk and behavioral intention |

awareness, but still were significantly better than before [33]. The game had multiple levels and changed between learning sections and exercises to train themselves.

Although refresher training is important, there are also findings that refresher training is not needed indefinitely. At a certain level of repetition, knowledge can persist for a very long time and the forgetting curve can be greatly flattened [20, 35, 36]. For this reason, it is first necessary to determine whether there is a loss of knowledge before investigating which time and which content offers the best refresher.

## 2.3 Security Interventions

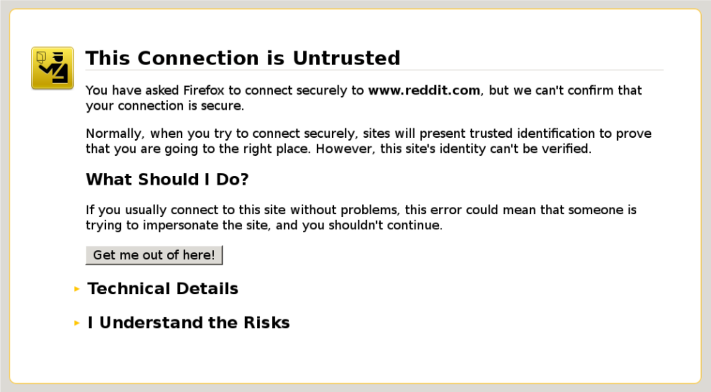
The creation of awareness to detect phishing can also be supported by tools for people in their everyday lives. In the past, many different forms and types of tools have been developed.

Basically, active security interventions have proven to be more effective than their passive counterparts [15]. Passive means that the user’s current activity is not interrupted. Rather, he is passively accompanied by an indicator and does not necessarily have to interact with it (see Table 2.2 for an overview of the security intervention studies).

One problem that can arise with active interventions is that it can lead to some fatigue towards the intervention or its messages [37]. Fatigue in this case describes that repeatedly showing a message leads to the reduced effectiveness of that message [37]. It is not possible to prevent this fatigue from occurring, but it is possible to reduce or delay the onset of fatigue and stretch it out over a longer period of time. One way to ensure this is that the intervention is only active in a certain situation or that the intervention is divided into different cases [47]. This subdivision with different cases can then lead to the reduction of fatigue. It is also important to communicate the case distinction clearly, so that users understand what its meaning is and how they come to a decision [48]. In particular, communicating this distinction and its meaning is often an area of security interventions that still needs development [49]. Interventions need sufficient information about this classification or assessment to achieve their effectiveness [37, 48]. Another issue for improvement is what possible consequences and risks should be part of the communication [50].

Furthermore, at the time when such a security intervention appears, different characteristics have initially emerged. In the context of phishing, for example, there are warning messages that appear after clicking on a link if the link has been classified as dangerous [37] (see Figure 2.1). The difficulty with these interventions is that they train the user to behave in a way that can be even more dangerous in certain situations than without this learned behavior. If the user becomes accustomed to it, then in the event that the intervention no longer works or has been circumvented, an action could lead to danger. Moreover, this behavior could transfer to other contexts and lead to dangerous situations as well. Accordingly, users should rather be directly taught to behave safely, even in the context of an intervention. So that even without the intervention, the behavior still provides some protection in the worst case. Another danger of this downstream warning is the psychological phenomena of anchoring bias and loss aversion [51]. Anchoring bias here means that by clicking on the link, a decision has already been formed to some extent. Even if a warning comes now, the basic tendency is to stay with the first decision (the anchor) especially if the credibility of the warning is not clear. In contrast, in the same case, if the warning comes first, the pendulum might swing in the direction of the warning. Loss aversion here describes the “loss” of access to the link that initially occurred by clicking and now is to be taken away by the warning. Tversky and Kahnemann described the strength of the loss compared to the gain as almost double [52]. Accordingly, it must be noted here that such an effect can also have a strong influence on “following” a warning.

But warnings can also be presented in different forms before clicking. Here, very different forms have been used in the past. For example, so-called banners have been used to provide further guidance about the assessment in the margin of an email [45] (see Figure 2.2). In contrast, other forms display the information directly at the cursor. However, these banner warning show certain weaknesses in the past, which are difficult to eliminate in this form. Typically, they are very limited in size and accordingly can only carry a small amount of information. This, in combination with the fact that they do not directly address the problematic aspect of the message, e.g. the link, can lead to ambiguity and thus reduce the effectiveness of the warning. This can lead from misunderstanding to ignoring the message [5]. In contrast, those interventions that are



This Connection is Untrusted

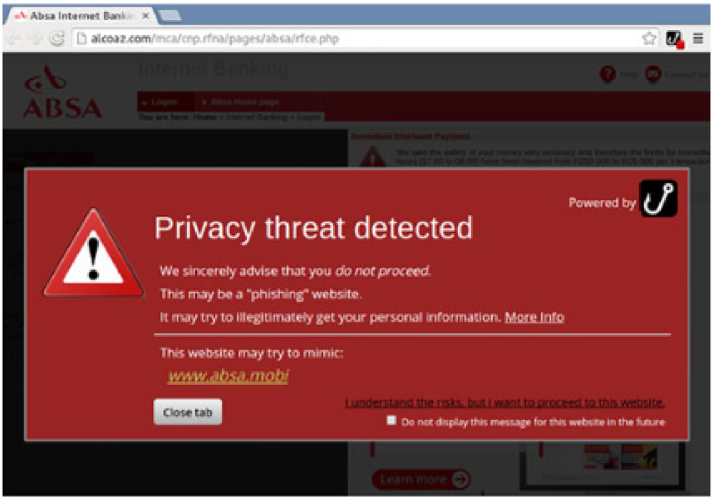
You have asked Firefox to connect securely to **www.reddit.com**, but we can't confirm that your connection is secure.

Normally, when you try to connect securely, sites will present trusted identification to prove that you are going to the right place. However, this site's identity can't be verified.

What Should I Do?

If you usually connect to this site without problems, this error could mean that someone is trying to impersonate the site, and you shouldn't continue.

**Figure 2.1:** Example for a warning page from Akhawe et al. [37].



Absa Internet Bankin
alcoaz.com/mca/cnp.rfna/pages/absa/rfce.php
ABSA
Internet Banking
Powered by

Privacy threat detected

We sincerely advise that you *do not proceed.*

This may be a "phishing" website.

It may try to illegitimately get your personal information. More Info

This website may try to mimic:

www.absa.mobi

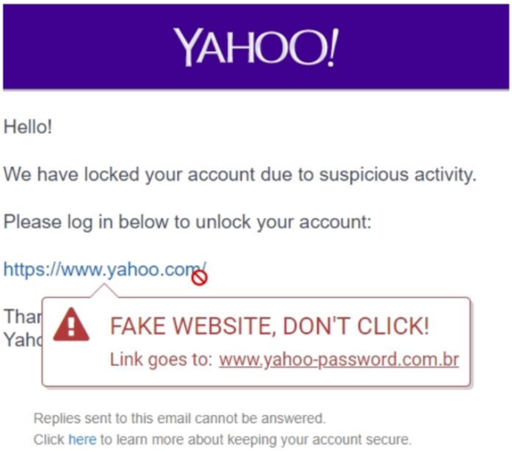
Close tab
I understand the risks, but I want to proceed to this website.

□ Do not display this message for this website in the future

**Figure 2.2:** Example for a banner warning from Marchal et al. [45].

in the user's focus perform better than those that are out of focus [5]. An example of the user's focus in the phishing context would be directly on the mouse cursor that is currently over a link.

Other forms of decision support such as chat bots have also shown utility in a previous study [53]. The drawback of this form, though, compared to those that directly address the user's focus, is that the user must first make a context switch to the domain of the chat bot. This context switch is a hurdle that can possibly be overcome in studies. But in reality, in certain situations and with strategies such as time pressure, this switch can be blocked and the intervention can thus not take effect. Most importantly, studies have already shown that interruptions can limit productivity [54]. People need time to recover between each change again [54]. To be sure, short interruptions



YAHOO!

Hello!

We have locked your account due to suspicious activity.

Please log in below to unlock your account:

https://www.yahoo.com/

Thar
Yahc

FAKE WEBSITE, DON'T CLICK!
Link goes to: www.yahoo-password.com.br

Replies sent to this email cannot be answered.
Click here to learn more about keeping your account secure.

**Figure 2.3:** Example for a link-centric warning with a tooltip from Petelka et al. [5].

are less disruptive than longer ones, and the same is true for such interruptions that are part of the task [54, 55]. Nevertheless, more research would be needed to clearly establish that assessing whether a URL is phishing is actually part of the task of processing emails.

Although such changes may not lead to a change in productivity at first, because the missing time is compensated by working faster [55]. However, in the long run, this leads to increased workload, increased stress perception and thus can strongly influence productivity [55]. Based on the previous studies, it can also be emphasized that the design of the warnings is crucial and that just-in-place support and domain highlighting can be helpful [56]. The highlighting is intended to help focus attention directly on the important part of the URL. It is, in a sense, an extension of the just-in-place principle that the tooltip has already applied to the banner.

To sum it up, security interventions in a form of a just in place and time warning can help people to distinguish between phishing and legitimate messages. But they should be implemented before the actual action of a user happen so that users are not getting used of clicking or acting before making a decision about the trustworthyness.

# 3 Background

In this thesis, the research questions described are answered on the basis of existing interventions (both phishing awareness measures and phishing tool support). The existing measures and the selection criteria for those are described in Section 3.2. Their necessary adaptations are described in the corresponding section of the study. Additionally in Part II a security intervention is tested against a phishing awareness measure. This intervention is described in Section 3.4. In order to be able to make as broad a statement as possible about the ability to detect phishing, various phishing tricks from literature are combined and examined (more on this in Section 3.1). Furthermore, the focus is not only on phishing detection but also on the ability to differentiate between legitimacy and phishing. The Signal Detection Theory with its measures of sensitivity and criterion is used throughout the thesis (see Section 3.6).

## 3.1 Phishing Tricks

In the past, many investigations have used very different forms of phishing tricks. The exact list of phishing tricks or strategies used in this thesis and from which literature they are extracted can be found in Table 3.1. Advice often involves looking at the URL to make a final assessment of where a link goes [57, 58]. Accordingly, various phishing tricks targeting the URL, in particular, must form an essential part of analyzing phishing awareness measures and tool support. Often, the main focus was on the sender's investigation or the email's plausibility [59, 60]. Nevertheless, the URL or attachments were already the focus of research [13, 29]. Especially in the URL area, there are many possibilities based on different "vulnerabilities" in the evaluation to manipulate the URL or domain. There are many different attack possibilities with the help of phishing links. In addition, the manipulation of the URL makes very different demands on the awareness of the user and also a very different focus down to very small subtleties. In the past, many measures have indeed addressed URLs or links. But without the appropriate depth and with only very abstract assistance, these measures do not allow sufficient protection to emerge. Accordingly, the focus should be on looking at just such phishing tricks for manipulating URLs. Tricks that cannot be detected directly when looking at the URL, even by means of a measure, are to be left out. An example of this would be so-called short URLs or redirects. Here, the obfuscation of the target URL can only be resolved with technical measures and then a decision can be made. Consequently, an evaluation of such a URL could only lead to the result that no final decision can be made. The decisions should be able to be made however clearly, in order to be able to make later a classification into correct and wrong decision.

**Table 3.1:** Overview of the different phishing tricks or manipulation strategies across different papers. An "x" signals that the mentioned literature included the phishing trick.

| Phishing Trick | Strategy | [61] | [3] | [33] | [4] | [11] | [62] |
|---|---|---|---|---|---|---|---|
| Content | Implausible Sender | | | | x | | |
| | Implausible Content | | x | | x | | |
| Non-Brand related Domain | IP | x | x | x | | | |
| e.g. https://host745.com/ | Random | x | x | | | | |
| Non-Brand related Domain + Brand Outside | Subdomain | | x | x | x | | |
| e.g. www.google.com.megahoust.ru/ | Path | | | x | x | | x |
| Small Deviations in the Domain | Typo | | | | | | x |
| e.g. https://www.netfllx.com/ | Letter Swap | x | | x | x | | |
| | Similar Character | x | | x | x | | |
| | Extension | x | | | | | |
| Special Link Manipulation | Mismatch | | | | | x | x |
| | Faketooltip | | x | | | | |
| Attachment | .exe | | | | x | | |

The tricks can be separated into three different groups: (1) Content, (2) URL and (3) Attachment. Whereas the focus is mainly on (2) URL. (1) Those tricks do not have any link/URL or attachment, therefore they can only be detected by checking aspects like: sender, design or the text content (e.g. pressuring people into actions). (2) For those tricks, emails are 100% copied from legitimate examples and only the URL is changed to being fraudulent. Therefore checking aspects like for (1) does not help in detecting them. (3) For those tricks, emails are similar to (2) copied from legitimate examples and only the attachment type is changed.

Based on these categories phishing tricks from various sources were collected and put into one scheme. The scheme does not claim to be complete. The used phishing tricks and special strategies represent a mix of different literature, which are currently dealing with the topic of phishing. In the literature, strategies for manipulation were not always grouped into categories of phishing tricks. However, some strategies start from very similar points e.g. using the pretend organization in a section of the URL or modifying the pretend organization name with different strategies. To account for this, the phishing tricks have been introduced.

**Content** E-Mails that fall in this category do not include a link. Therefore the decision needs to be based on other indicators. The user should instead focus on either the sender or the content to decide whether it is a phish or not. An example could look like that the sender address is amazon.de@phish.me. The trick focuses on the user's lack of awareness that the part behind the link is similar to the domain the relevant part to check. For the e-mails that fall in the content category the phisher asks for either actions that should not be taken by the user (e.g. transfer money) or for sensitive information that organization normally do not ask directly (e.g. birthdaydate or bankdata).

**Non-Brand related Domain** E-Mails from this category have nothing in the URL that does help to identify the organization one is communicating with. Therefore they target on users that lack the awareness about the importance of the URL as indicator. As this trick can easily detected when directly looking at the URL. The URL either includes an IP address or some random characters. Even though every website has an IP address, only a small amount of people know what IP is behind a specific website. Therefore normally users should be very wary when confronted with an IP. Also as mentioned before, the domain should give information about with whom one is communicating and if there are random characters, this is a strong indicator for phishing.

**Non-Brand related Domain + Brand Outside** The trick with the brand outside the domain is then focusing on people that know that they should check the URL, but that lack more specific awareness about where to check. In such cases people could be tricked by the brand name they think they are communicating with somewhere outside the domain. This trick can rely on different context variables. One could be that people (specially from the western world) read an URL as a normal text and therefor start from the left. At the point when they find the name of the brand they stop reading and classify the URL as trustworthy. Another possibility is that the brand is even after the domain. Looking at status bars that is also pretty obvious, because people normally shift there focus from the link and with the resolution increasing the status bar is moved further in the left bottom corner. So even though people would read the URL from left, they might first touch the URL with their gaze somewhere in the path. Using the brand somewhere in the path and in the subdomain also helps phishers create a multitude of phishing URLs with just a single domain. As every domain has to be registered, but the subdomain and path can be freely chosen for one single domain.

**Small Deviations in the Domain** Now it becomes even more complicated. Every domain has to be registered, but even small deviations in the domain lead to it being a new domain and a new registration is needed. A domain can look 99.9% similar to a trustworthy domain from a known organization, but can still be registered by someone else. So even when users are aware of the

importance of the domain, they still can miss such small deviations in the domain. The phisher uses the phenomenon “word recognition” or “lexical processing”. The phenomenon states that, for reasons of efficiency, not every letter is read letter by letter, but rather recognizes known words based on their shape and letter order [63, 64]. In the case of slight deviations, the reader may still recognize the known word. Accordingly, there are different possibilities for phishers, e.g. they can swap two letters in the order (mircosoft.com), they can replace one or more letters with a very similar letter (arnazon.de) or they simply insert slight spelling mistakes (tagessschau.de). Depending on e.g. the size or the font, such mismatches can be very difficult to notice.
Another possibility is to extend the domain’s range. This does not rely on the change not being detected by the user. Rather, it relies on the fact that this also happens in legitimate cases, and the user needs to know in the specific case whether this extension is authentic or just introduced by the phisher. Such an extension could be like amazon-sicher.de. Most of the time, either a word that is supposed to generate trust or a word that fits the context, e.g. “login”, is used. In this case, the user must either have the appropriate awareness or gather more information through other channels to make a decision.

**Special Link Manipulation** A link usually consists of a link text, button or image and the URL behind it. Phishers can also create so-called mismatches between the link text and the link URL or display a fake tooltip. In both cases, the link text or false tooltip displays a URL that is legitimate. However, the URL behind it is different from this URL and then leads to the phisher. This trick also relies on a lack of awareness on the part of the user. On the one hand, that the user is not aware that the link text can take any form from “Click here” to a URL and this says nothing about where the link actually leads. On the other hand, that the user does not know that the own program displays the URL only in the status bar. In both cases the user must have the awareness where in his environment the actual URL is displayed. In a way, this trick represents an additional dimension of the tricks, because it can also be combined with the tricks mentioned before. It is also a bit more technical in a way, because with the help of additional code in the email the mismatch or the fake tooltip has to be created first. Accordingly, these errors should be considered particularly critically.

**Attachment** In addition to grabbing data by typing it on a fake website, scammers sometimes want to trick the user into opening a file. This file can also be hidden behind a link and downloaded and launched when clicked. Phishers can also attach the file directly to the email and then use the content to try to get the user to open it. Here, too, there are various starting points for the phishers. They can rely on the fact that the user does not have the awareness to recognize which file formats are particularly dangerous. Also depending on the environment, the file formats are not always displayed directly and the user must first have the awareness of how to get to the display. In the case of file formats, a distinction must also be made between those that should be considered particularly dangerous per se (.exe) and those that at least harbor the danger (.doc). In the case of those that pose a risk, the point is that they are not directly dangerous in themselves, but they contain the possibility for macros. These macros can then cause similar functions as an .exe and should be handled with care.

Especially the strategies where the link text is changed so that a mismatch between link text and link URL is created have rarely been evaluated in the past. However, this form of phishing trick is a very easy way, which does not even require registering new domains, and can be reused very often. If the phisher assumes that the attacked person does not even have the basic awareness about this form of attack, then he can even choose the URL completely randomly and thus bypass many technical measures such as filters.

## 3.2 NoPhish Awareness Measure

The NoPhish concept contains different measures that were developed over time. Those measured were developed in an collected effort by various people from the SECUSO research group [1].

Even though there was contribution to various of these measures, their development is not a contribution of the thesis. The measures are therefore presented in this background chapter.

The NoPhish concept as such and the different phishing awareness measures are now going to presented in more detail. Starting with an overview of the NoPhish concept. Finally, presenting all forms or specific measures that were used in the evaluations. The NoPhish concept stands as the basis for the evaluation done in Chapter 4. All the materials used for Part 1 are part of the NoPhish concept and represent a different format, but with the same overall goal and core content. It was originally developed starting with the NoPhish App and leaflets at the Technical University of Darmstadt and transferred to the Karlsruhe Institute of Technology. So part of the measures presented are first developed in Darmstadt and others at the Karlsruhe Institute of Technology. Also, existing misconception are addressed, which have been elaborated from various previous researches. Therefore, the misconceptions collected from previous literature and how they are addressed is listed below:

- Only emails can be phishing messages. Instead, phishing links can also appear on websites, in social posts or in short messages.[65, 66]
- Phishing only aims to make you provide your banking information. Instead, phishers want to get all sorts of information from email account credentials or even just from other websites.[67, 65]

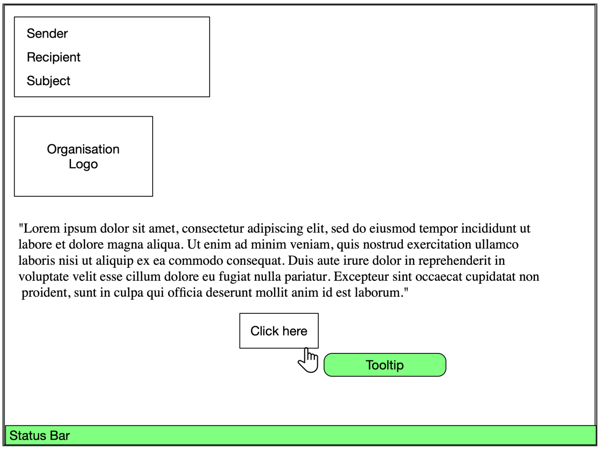
[1] `https://secuso.aifb.kit.edu/Team.php`

- The sender name is a trustworthy criterion that can always be relied upon. Instead, phishers can easily spoof the sender without having much technical knowledge. The sender can be looked at, but is not a reliable indicator for detection.[68, 66]
- Only people who have a high position or are particularly rich or famous are targeted by phishing attacks. Instead, anyone can become a victim of such an attack. After all, everyone has interesting data in some form, whether it is access data or contacts.[69]
- Phishing attacks can be completely prevented by technical measures. Instead, technical measures can help, but do not provide 100% security and accordingly users have to learn to deal with it themselves.[70]
- HTTPS compared to HTTP already means that a URL is secure. Instead, attackers can also use an HTTPS URL very easily nowadays and it does not guarantee security against phishing attacks. [66]
- Already trustworthy words like secure in the URL mean that it is trustworthy. Instead, such words are rather used as a trick to increase the trustworthiness. Thus, special attention should be paid to such words especially in the domain. [71]

### 3.2.1 General Idea & Concept

NoPhish represents a general concept around getting people to understand what phishing is and how detect phishing respectively differentiate between phishing and legitimate messages. It is supposed to create a general awareness about the topic of phishing ,e.g., that everyone could be a potential target, what potential damage is possible and should know at least the basics to defend against it. After this general introduction the first indicators to detect a phishing message are explained such as checking the sender or the general content of the message for plausibility. These are not the final indicators to base a decision on, but can help to detect the most obvious phishing messages. They are asked to examine the link instead as the sender can be spoofed and the email content can be copied from existing messages easily, too.

Then the focus moves to the URL as the most important and reliable factor of a messages. It is presented, how to find the URL in different context as this might differ from context to context e.g. from a mobile app to a desktop client (but even differs between desktop e-mail clients). There are currently two status quo for the representation of the URL for the desktop context as shown in Figure 3.1. First, there can be a tooltip directly at the cursor or a status bar appears at the lower left edge. This is then followed by instructions to look at the URL in either the tooltip or status bar and examine it for suspicious content, as an example `megahost.ru` being shown instead of mein-paketservice.de.



"Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum."

**Figure 3.1:** Abstract representation of an email highlighting the status bar and the tooltip.

Various URL related phishing tricks and advice how to detect those tricks are presented. The tricks are explained with examples so that it is more clear how such a URL would look like in the wild. In the end email attachment as a special case are explained. There a list of most probably "safe" file extensions and such that are more dangerous are explained.

Awareness in the context of the NoPish concept means making aware of the threat and providing knowledge of how to differentiate between legitimate and phishing messages.

## 3.3 Implementation of the Awareness Measures for NoPhish

Currently, there exist a variety of different implementations of the NoPhish concept into actual measures[2]. Some are developed as main measures, others to refresh awareness some time after, and other for both. The measures range from very small and short info cards over videos of around five minutes to workshop material of multiple hours. In the following, the three measures and four refreshment measures (video counts for both) are explained in detail, which are later on used in Chapter 4. Note, the video is also used in Chapter 5. The measures are presented in the chronological order of their usage in Chapter 4, starting with the video implementation. In this chapter, the original measures are presented, for those changes that were needed for the study, the adaptions made to the measures are presented in the corresponding study material sections.

[2] `https://secuso.aifb.kit.edu/betruegerische_nachrichten_erkennen.php` last accessed 15th January 2024

Those measure share the same foundation with the NoPhish concept, but also because of the variance in length, only parts of the whole concept are integrated in some of the measures. The included topics to the measure can be seen in Table 3.2 and afterwards the integration of the concept for the various measures is described in more detail.

**Table 3.2:** Overview of six phishing tricks from Section 3.1 included per measure and refresher measures.

| Topics | Video | E-Learning | Workshop | Long text | Short text | Interactive email example |
|---|---|---|---|---|---|---|
| Content | | x | x | x | | x |
| Non-Brand related Domain | x | x | x | x | x | x |
| Non-Brand related Domain + Brand Outside | x | x | x | x | x | x |
| Small Deviations in the Domain | x | x | x | x | x | x |
| Special Link Manipulation | | x | x | | | |
| Attachment | | x | x | x | | |

The target audience of the NoPhish concept is laypeople or the general population. Therefore, no special knowledge, technical knowledge or even known terminologies are required to understand and follow the video. The aim is to raise awareness of phishing and to show how attackers create phishing messages and what psychological tricks they use. Concrete consequences of one's own behavior are also pointed out in order to make it clearer what dangers can arise.

### 3.3.1 Video

In total, there are three videos, each with a different focus. Each video explains an aspect of the NoPhish concept in about five minutes. The first and original video has the focus on the links, the second video has the focus on the attachment and the third video has the focus on the content. As only one of those videos is used for this thesis, the content of the video with focus on links is presented in detail.

The video is made in a comic style (see Figure 3.2). There is both a graphic story that runs through the entire video and there is also a narrator voice that guides the viewer through the story and explains the animations. The whole story follows one person and tells several little stories around that person in the context of phishing messages. The video starts with a basic introduction that there have always been scammers and depending on the time they just had a different form (e.g. pirate) or different goal (e.g. steal piggy bank). This analogy is meant to build a bridge

**Figure 3.2:** Start sequence of the video showing the comic style and main character on the right side.

even for non-technical users. Then the connection to today is made that fraudsters pretend to be known person or grandchildren and ring the doorbell of the protagonist. Again, using an example from the analog world, the idea is to create a connection for people with little prior knowledge so that they have an idea. The door is opened for these people in the video, but they are not let in. Now it is explained how attackers make it so that you don't have to gain their trust first, but enjoy it right from the start. It is explained that such scammers create genuine-looking messages in order to lead the reader to an equally genuine-looking website and trick him into entering his data there. Afterwards, the data can be used to order something or withdraw money from the account. The possibility of installing malware on one's own computer by clicking on the link is also presented. Through such malware, data can then also flow out and end up with the attacker. The data can also be encrypted by the attack and thus not be available anymore. Then still another important aspect is explained that this method is very simple and from everywhere in the world with thousands of humans at the same time to be accomplished. It is explained to the viewer that thus this type of attack can hit anyone in the world. Then, different contexts are presented and illustrated in which such a link can occur (email, social media post or website). The viewer is then taught that it is important to recognize whether a link is trustworthy or not. The link can be hidden behind a number of different things, e.g. a button or an image (see Figure 3.3). It is explained that you first have to move the mouse over the link and wait a short time.

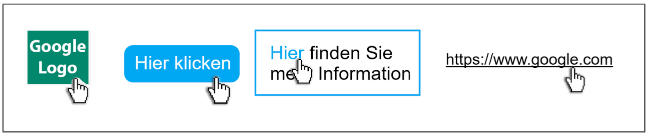


**Figure 3.3:** Different forms of a link. Google logo censored due to missing copyrights.

Tip 1 is: "Watch out for the "who-area". The tip starts with two representations of URLs and the information that attackers adopt the behavior of users. In fact, most read URLs from left to right like normal text. This is followed by an explanation of the structure of the URL with the different sections. Also the explanation, the meaning of dots and slash in the address or the importance of the third slash and the so-called “who-area” (combination of domain and top-level domain) in front of it. The focus here is on a German sample and does not address topics like a double top-level domain from countries like UK (.co.uk) or New Zealand (.co.nz). This is also made graphically clear again with a marker (see Figure 3.4). Some examples follow to illustrate how different URLs can look like and where the “who area” would be. Then with an animation, that you can put unlimited words and dots in front of the “who-area”, the URL is extended further and further in the subdomain.

Tip 2 is: “Watch out for the fake who-area”. In this type of phishing trick, the attackers rely on the reader not looking closely. Accordingly, they use a domain that looks very similar to the fake domain and can be easily missed if the focus is not completely on the URL. For example, letters are rotated or replaced with similar looking letters. Again, various examples are shown that run across the screen one after the other from bottom to top, e.g. paypa1.com or paketsrevice.de. The examples are also explicitly shown in detail and the domain is zoomed in again. This is to make clear that at first it looked very authentic, because the font was rather small. However, when you look closer, the amazon.de actually becomes arnazon.de. Therefore, one should always look at the domain character by character. The video ends with some teaser on more technical examples like the use of ShortURL services with three examples for such services and the reference for further information to click on the following link.

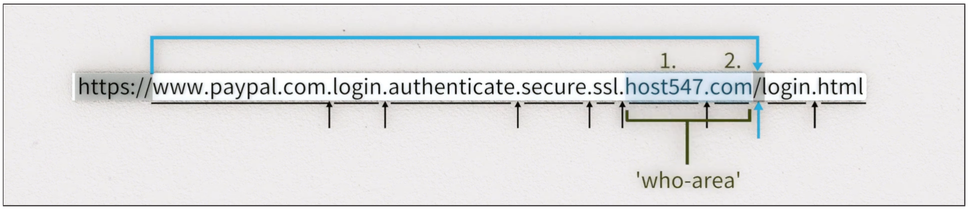

1.
2.
https://www.paypal.com.login.authenticate.secure.ssl.host547.com/login.html
'who-area'

**Figure 3.4:** URL with highlighted ’who-area’.

## 3.3.2 E-Learning

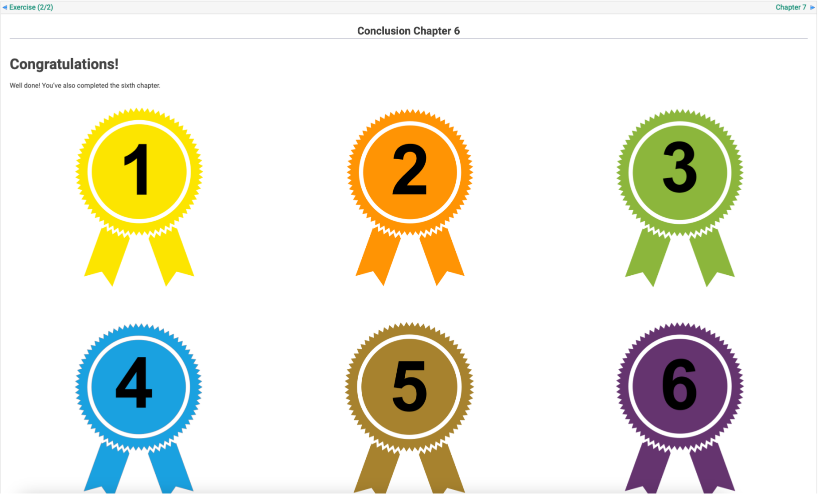
In addition to very short measures such as the video, the NoPhish concept also includes more extensive measures such as E-Learning. This training is based on the findings of several evaluations of a presence or self-learning training. These forms have also been evaluated in the past in different contexts with different samples to create the most mature concept for a training. One study compared different measures such as the instructor-based training, a computer-based measure and a text-based measure in a school context, finding the instructor-based training performing better than the other two [61]. Another study used the training also in an instructor-based variant with lay people in an adult education centre, showing significant improvements [72]. Finally, a third study used the training in a text-based version for self-learning in a organization context [4]. This resulted in a concept that was transformed into the interactive e-learning environment with the professional support of a service of the Karlsruhe Institute of Technology. The e-learning course is based on an ILIAS system [73], which is classified as an open source learning management system. Compared to the previous iterations e.g. as PDF [4] the ILIAS system offers the possibility to include exercises and other interactive elements like videos. Compared to other measures the E-Learning also has the benefits that participants can take a break at any time they want. This includes even closing the browser or shutting down the computer. Starting again, they are put directly at the spot they stopped and can resume from there on. The correctness of the exercises can also be checked directly, thus automatically regulating the learning progress without manual intervention. The E-Learning consists of twelve chapters. Every chapter has a different topic (see Section 3.2.1).

Each chapter has its own focus, but the chapters can still be divided into topic areas. The E-Learning starts with the section that first explains the context for phishing and phishing as such (chapters 1 to 3). Then the foundations for the subsequent behavioral instructions are laid by explaining the structure of the URL (chapters 4 and 5). As a third topic area, there are different tricks that can be detected with different behaviors (chapters 6 to 9). Finally, there is the subject area of attachments (chapters 10 to 12). Each chapter had different number of pages both in content and in the area of exercises (see Table 3.3). The number of pages is related to the amount of information that was needed to explain the subject matter or the nature of the exercises.

**Table 3.3:** Number of pages for each chapter both for content and exercise part.

| Chapter | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pages Content | 7 | 8 | 4 | 6 | 3 | 5 | 4 | 5 | 4 | 4 | 4 | 2 |
| Pages Exercises | 2 | 3 | - | 8 | 12 | 2 | 2 | 2 | 1 | 6 | 1 | 3 |

With a starting page, it was always made clear to the readers whether a content section or an exercise section would follow next. In addition, there was always an additional page after the exercise tasks where a badge was awarded for each successful exercise for a chapter and then contained the entire overview of all previous badges (for an example see Figure 3.5). The goal was to reward the readers through the badges and the progress shown and thus further motivate them. To give readers further orientation, each page was always labeled with the number of the current page and the maximum pages of the chapter in addition to the topic e.g. “(1/7) Introduction "Fraudulent Messages".


◂ Exercise (2/2)

Chapter 7 ▸

Conclusion Chapter 6

Congratulations!

Well done! You've also completed the sixth chapter.

**Figure 3.5:** Example of a page that concludes a chapter with badges of previously completed chapters.

### 3.3.3 Workshop

Similar to the E-Learning, the workshop is based on the findings of the previous training. Workshop in contrast to the E-Learning means that the measure consists of at least a couple of people being taught by an instructor the content. The workshop has also only been evaluated in a pre-post measurement. The training was evaluated in presence in the school context [61] and as self-study material in the organizational context [4]. In this regard, the current workshop is based on the findings from these evaluations. Similar to these studies, the workshop was in the German language. Most of the slides consisted of either an example, which was used to explain facts with markers and short texts, or larger text blocks which were supported with graphics, e.g. the structure of the URL.

The workshop consisted of four parts:

1. How to handle emails
2. How to handle links
3. How to read URL's correctly
4. How to handle attachments

Similar to the measures before, the workshop starts with the basic context on phishing and this includes first awareness of potential dangers when "carelessly clicking" on links and attachments. Here concrete examples are given for possible dangers and consequences for the individual or the organization. These include spying on data, carrying out actions on the device, and blocking certain activities. Then, three specific possibilities are mentioned that phishers use, e.g. performing unwanted actions such as bank transfers, clicking on links and opening or executing files.

Next, the awareness of the participants is tested for the first time by showing two examples of an email at the same time. The two e-mails look almost exactly the same and differ only in that the sender ends once in ".ru". This should catch the attention of the viewers and directly address a misconception that phishing e-mails can often be easily recognized by their grammar or spelling. Now comes the division into three aspects that are important for the later course when judging an e-mail: Checking the plausibility (topic, sender and content), dangerous links and dangerous attachments.

Afterwards, the last section of the first part on dealing with e-mails starts directly with three examples from the organizational context for checking plausibility. Here, the participants should first get a feeling for which e-mails have already been on their way in the context of the organization and which criteria can be used to recognize them. For this purpose, the important areas of the e-mail are also highlighted to direct the focus of the participants to these areas. Afterwards, the participants are presented with questions that are intended to help them check the plausibility as a behavioral cue. This is followed by another four examples to directly apply the learned questions. Further aggravating factors such as the use of pointer pressure are also discussed. Then the page with the two examples from the beginning is shown again to make a bow to it. This part closes again with an example for the plausibility examination, which is not from the organization context. This is intended to sensitize employees to the fact that phishing does not only apply to their own organization, but also in a private context, e.g. in the case of a fictitious Telekom invoice.

The second part on dealing with dangerous links starts with an example and already shows a rather complicated trick, a mismatch between link text and link URL. This should make clear at this point that it is especially important to know how to get to the URL behind a link and not to focus on the link text. At this point it is also discussed that it is important to know which display form status bar or tooltip your own software offers. This is the transition to the next three pages where a fake tooltip is shown, which is similar to the link text a mismatch to the actual URL in the status bar at the bottom left.

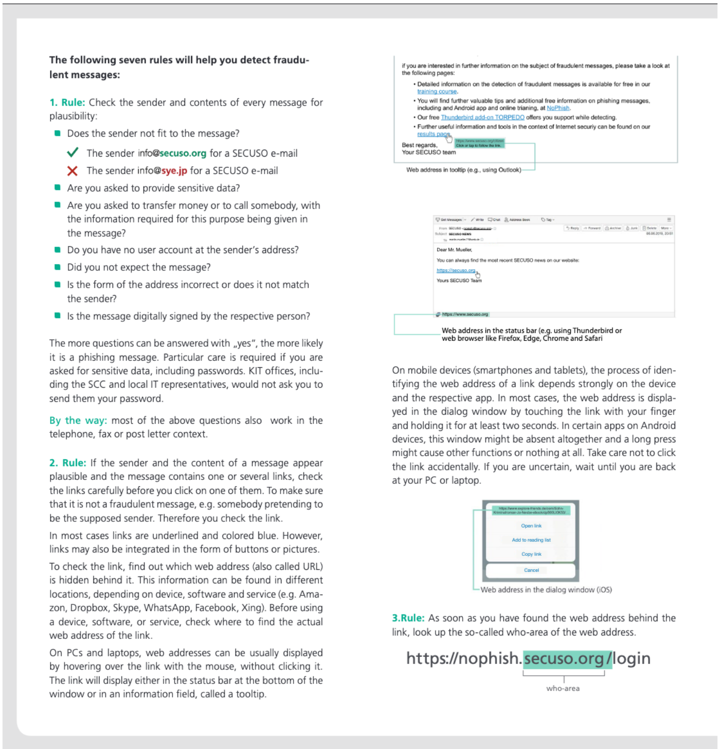
After establishing how to get the URL behind a link, the third part explains the structure of the URL with the important components. Similar to the video and the E-Learning it is then explained that the "interlocutor" is in the area of the domain and top-level domain and that this is in the area directly before the third slash by two blocks. On several pages, each section of the URL is now explained again with markers. This goes over into the representation of different URLs on a side with marking of the Domain and Top level Domain. The URLs represent a mixture of different factors e.g. legitimate and phishing or short and long. Here the participants should practice recognizing the correct target of the URL with a subsequent resolution. On the next page there are examples with different phishing tricks e.g. replacing letters with similar looking letters, slight spelling mistakes or extending the domain with matching words like “ebay-kleinanzeigen.de”. To increase the importance of the awareness of especially the slight adjustments of the domain, there is a short example text that illustrates how easy it is for the human brain to read a text, even if the letters are twisted. The section ends with a summary of everything that has been learned up to this point, in order to check the plausibility of the email and the links.

The fourth section on dangerous attachments begins with the first behavioral tips to help avoid clicking on dangerous content. This includes contacting the sender via another channel (not the one in the e-mail), obtaining further information from colleagues or asking for a different file format. Now it goes into the concrete examination of the file format and for it the consciousness must exist that depending upon used Client not always directly the complete file format is to be seen. This is followed by a brief classification of different file formats into potentially dangerous ones (e.g. .exe) and those that have at least the danger of macros (e.g. .doc). These lists are clearly presented as not complete and only a brief overview of the most common ones. It also explicitly addresses special cases, e.g. archive file formats, use of duplicate extensions and unknown file formats. There is also more information about sites that collect potentially dangerous formats and settings on your own client, e.g. preventing the direct display of images. As an alternative if still unsure, participants were presented with the option to search for the sender or content of the email using search engines to find known phishing cases involved. The section concludes with a summary of all important behaviors regarding plausibility, links and attachments.

### 3.3.4 Leaflet

The long text represents the first refreshment measure and has been part of the NoPhish concept for a long time. It represents a much used format that has already been used in other contexts to create awareness. The advantages of a long text are that many people are already familiar with the format. In addition, it is also very suitable for distribution because it requires few prerequisites and can be easily distributed in large quantities. The long text also offers the possibility of providing a lot of text despite its relatively small size (see a part of the long text in Figure 3.6). For the full content of the long text see Appendix A.1.

The long text starts with an introduction very similar to the other measures already described on the context of phishing. This includes that popular attack strategies focus on either spreading malware to access the device or to obtain sensitive information such as login credentials. Phishing is a widespread possibility for this. Phishing does not only occur in the email context, but also in short messages or messages in social networks. Then it is explained again that phishing can have different forms, e.g. asking for sensitive data, requesting unwanted behavior (e.g. bank transfers), enticing people to click on links or attachments. Next, eight rules are presented, each highlighting a different aspect to protect yourself from phishing.



**The following seven rules will help you detect fraudulent messages:**

**1. Rule:** Check the sender and contents of every message for plausibility:

- Does the sender not fit to the message?
  - ✓ The sender info@**secuso.org** for a SECUSO e-mail
  - ✗ The sender info@**sye.jp** for a SECUSO e-mail
- Are you asked to provide sensitive data?
- Are you asked to transfer money or to call somebody, with the information required for this purpose being given in the message?
- Do you have no user account at the sender's address?
- Did you not expect the message?
- Is the form of the address incorrect or does it not match the sender?
- Is the message digitally signed by the respective person?

The more questions can be answered with „yes", the more likely it is a phishing message. Particular care is required if you are asked for sensitive data, including passwords. KIT offices, including the SCC and local IT representatives, would not ask you to send them your password.

By the way: most of the above questions also work in the telephone, fax or post letter context.

**2. Rule:** If the sender and the content of a message appear plausible and the message contains one or several links, check the links carefully before you click on one of them. To make sure that it is not a fraudulent message, e.g. somebody pretending to be the supposed sender. Therefore you check the link.

In most cases links are underlined and colored blue. However, links may also be integrated in the form of buttons or pictures.

To check the link, find out which web address (also called URL) is hidden behind it. This information can be found in different locations, depending on device, software and service (e.g. Amazon, Dropbox, Skype, WhatsApp, Facebook, Xing). Before using a device, software, or service, check where to find the actual web address of the link.

On PCs and laptops, web addresses can be usually displayed by hovering over the link with the mouse, without clicking it. The link will display either in the status bar at the bottom of the window or in an information field, called a tooltip.

if you are interested in further information on the subject of fraudulent messages, please take a look at the following pages:

- Detailed information on the detection of fraudulent messages is available for free in our training course.
- You will find further valuable tips and additional free information on phishing messages, including and Android app and online trianing, at NoPhish.
- Our free Thunderbird add-on TORPEDO offers you support while detecting.
- Further useful information and tools in the context of Internet securiy can be found on our results page.

Best regards,
Your SECUSO team

https://www.secuso.org/citizen
Click or tap to follow the link.

Web address in tooltip (e.g., using Outlook)

From SECUSO
Subject SECUSO NEWS
06.06.2019, 20:51

Dear Mr. Mueller,

You can always find the most recent SECUSO news on our website:

https://secuso.org

Yours SECUSO Team

Web address in the status bar (e.g. using Thunderbird or web browser like Firefox, Edge, Chrome and Safari

On mobile devices (smartphones and tablets), the process of identifying the web address of a link depends strongly on the device and the respective app. In most cases, the web address is displayed in the dialog window by touching the link with your finger and holding it for at least two seconds. In certain apps on Android devices, this window might be absent altogether and a long press might cause other functions or nothing at all. Take care not to click the link accidentally. If you are uncertain, wait until you are back at your PC or laptop.

Web address in the dialog window (iOS)

**3.Rule:** As soon as you have found the web address behind the link, look up the so-called who-area of the web address.

https://nophish.secuso.org/login

**Figure 3.6:** First part of the back pages of the leaflet from the original measure.

Rule 1 starts with the information to check the sender and the content for plausibility. So the reader expects just this message and does the sender match the content of the message. Here are two examples of a fraudulent looking sender and a legitimate sender on an Amazon email.

Rule 2 then explains that just because the content or sender are plausible, you should not click directly on the link just yet. First, if it exists, it should be investigated further. In the text you can recognize the link for example by blue writing and underlining the words. It can also appear in the form of buttons and images. Once you have found the link, you still need to know how the link URL is displayed in the device or software used. Here are examples of a tooltip with reference to the Outlook client and a status bar with reference to Thunderbird or web browsers. Also the possibility of a mismatch between link text and actual URL is shown with a short explanation as well as with a graphic. After that, the mobile context is addressed, since it depends even more clearly on the device which behavior must be shown to get the URL, e.g. hold your finger on the link for 2 seconds (as of 2018).

Rule 3 is based on now seeing the URL and having to decide if it is a phish or a legitimate URL. To do this, the term “who area” is introduced for the combination of domain and top level. This area is described as essential and highlighted with a graphic (see Figure 3.6). As in the other measures, it is described as a range separated by 2 dots before the third slash. First, the example of an IP is shown with the indication that it is not trusted (writing highlighted in red).

Rule 4 builds on the awareness of the who area and shows two more tricks. Sometimes the organization that should be in the who section is in other sections of the URL, e.g. the path or the subdomain. Three examples are given which show the legitimate form and then the other two tricks.

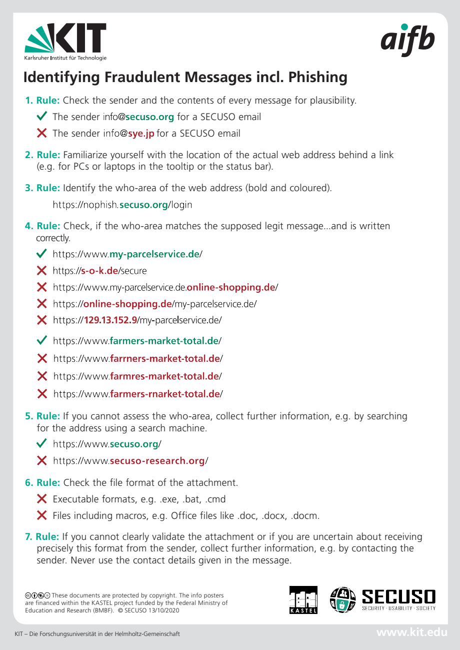
In rule 5 it gets more complicated, because now changes are made in the domain. The idea is to use slight adjustments in the domain, e.g. twisting the letters, replacing letters with similar letters and small spelling mistakes. Again, the example is shown for the legitimate URL and then for one of the three tricks. The goal is always to give a clue for changing the URL, because the example URLs are unknown and otherwise the small mistakes would not be detected directly.

Rule 6 is again about an adaptation of the domain and this time a “trusted” word is added, in the example “research” to the context of SECUSO. This is particularly difficult to detect, as it can also be the case with legitimate messages. Accordingly, the tip in case of uncertainty is to ask for help, e.g. via a search engine search. Most of the time, in cases where you see URLs with such extensions, search engines either display the correct URL of the organization without the extension (if it is a phish) and with the extension if it was a legitimate message. The last two tricks then deal with attachments.

Rule 7 mainly divides three different file formats that can be dangerous: 1) Directly executable file formats; 2) File formats that can contain macros and 3) File formats that you don’t know. The first two are also given direct examples e.g. .exe or .bat for 1) and .doc or .xls for 2).
Rule 8 then gives behavioral hints on how to deal with it should you find such a file format and be unsure. For example, you can ask for more information, but not with the stored information from the message. Also one should think briefly, if one should be asked by the Office program whether macros are to be executed.

### 3.3.5 Poster

The short text as the second refreshment measure also represents a longer existing measure of the NoPhish concept. It has not yet been evaluated for its effectiveness. Similar to a long text, a poster is a way of communicating an issue to a larger number of people. In contrast to the long text, other prerequisites are needed here for distribution, since not everyone can simply get hold of the poster. Rather, a poster is more suitable to be placed in a highly frequented place in an organization or in a square. Due to the high frequency and the rather short dwell time, posters have to focus on the core elements in a very short form. Since most people just do not spend much time on it.



KIT
Karlsruher Institut für Technologie
aifb

**Identifying Fraudulent Messages incl. Phishing**

**1. Rule:** Check the sender and the contents of every message for plausibility.
- ✓ The sender info@**secuso.org** for a SECUSO email
- ✗ The sender info@**sye.jp** for a SECUSO email

**2. Rule:** Familiarize yourself with the location of the actual web address behind a link (e.g. for PCs or laptops in the tooltip or the status bar).

**3. Rule:** Identify the who-area of the web address (bold and coloured).

https://nophish.**secuso.org**/login

**4. Rule:** Check, if the who-area matches the supposed legit message...and is written correctly.
- ✓ https://www.**my-parcelservice.de**/
- ✗ https://**s-o-k.de**/secure
- ✗ https://www.my-parcelservice.de.**online-shopping.de**/
- ✗ https://**online-shopping.de**/my-parcelservice.de/
- ✗ https://**129.13.152.9**/my-parcelservice.de/
- ✓ https://www.**farmers-market-total.de**/
- ✗ https://www.**farrners-market-total.de**/
- ✗ https://www.**farmres-market-total.de**/
- ✗ https://www.**farmers-rnarket-total.de**/

**5. Rule:** If you cannot assess the who-area, collect further information, e.g. by searching for the address using a search machine.
- ✓ https://www.**secuso.org**/
- ✗ https://www.**secuso-research.org**/

**6. Rule:** Check the file format of the attachment.
- ✗ Executable formats, e.g. .exe, .bat, .cmd
- ✗ Files including macros, e.g. Office files like .doc, .docx, .docm.

**7. Rule:** If you cannot clearly validate the attachment or if you are uncertain about receiving precisely this format from the sender, collect further information, e.g. by contacting the sender. Never use the contact details given in the message.

These documents are protected by copyright. The info posters are financed within the KASTEL project funded by the Federal Ministry of Education and Research (BMBF). © SECUSO 13/10/2020

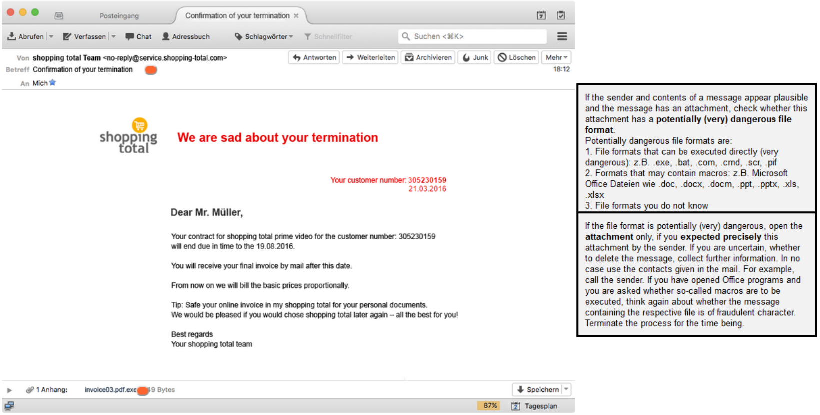
KASTEL
SECUSO
SECURITY · USABILITY · SOCIETY
KIT – Die Forschungsuniversität in der Helmholtz-Gemeinschaft
www.kit.edu

**Figure 3.7:** Poster from the original measure.

Accordingly, the short text offers the same eight tips as the long text (see Section 3.3.4). But with the difference that there are only one or two sentences about the basic rule. This sentence is underpinned in each case with the same examples as in the long text. So the short text (see Figure 3.7) is a strongly shortened and compressed form of the long text. The idea behind it is that the examples with the short explanation are sufficient to understand the phishing tricks and to take appropriate countermeasures.

### 3.3.6 Interactive Example

The interactive example is also based on the long text, similar to the short text. However, a completely different form was chosen. Similar to the video, this measure has more requirements that must be met. For example, at least one device is needed that enables the display of the interactive example. Compared to the video, even a mouse is needed to ensure interactivity. The interactive example represents a new measure, which was also not accessible to the public so far in comparison to all measures described so far. It is also the only measure where the user interacts with the measure. Similar to the long text or short text, and unlike the video, users can choose their own pace here. But unlike the other measures, the users themselves are active here and receive the information directly at the point where they later have to carry out the respective check. The interactive example consists of three pages. Two pages (page 1 and page 3) are screenshots of e-mails. These e-mails are based on actually sent e-mails but have been modified to represent a fictitious "explore-friends" service (see Figure 3.8). These two pages each have so-called hot spots, i.e. orange dots that trigger an action when touched with the cursor. These hot spots are located exactly where one of the eight tips starts, e.g. the sender or the link. When the user touches these points with the mouse, the corresponding information is displayed both in a tooltip directly at this hot spot and to the right of the e-mail, which is important for checking the respective aspect. In the case of the sender, for example, that the sender and the content should be checked for plausibility. When leaving the hotspot, the tooltip is hidden, but the tip on the right



Von shopping total Team <no-reply@service.shopping-total.com>
Betreff Confirmation of your termination
An Mich
18:12

**We are sad about your termination**

Your customer number: 305230159
21.03.2016

**Dear Mr. Müller,**

Your contract for shopping total prime video for the customer number: 305230159
will end due in time to the 19.08.2016.

You will receive your final invoice by mail after this date.

From now on we will bill the basic prices proportionally.

Tip: Safe your online invoice in my shopping total for your personal documents.
We would be pleased if you would chose shopping total later again – all the best for you!

Best regards
Your shopping total team

1 Anhang: invoice03.pdf.exe 9 Bytes

If the sender and contents of a message appear plausible and the message has an attachment, check whether this attachment has a **potentially (very) dangerous file format**.
Potentially dangerous file formats are:
1. File formats that can be executed directly (very dangerous): z.B. .exe, .bat, .com, .cmd, .scr, .pif
2. Formats that may contain macros: z.B. Microsoft Office Dateien wie .doc, .docx, .docm, .ppt, .pptx, .xls, .xlsx
3. File formats you do not know

If the file format is potentially (very) dangerous, open the **attachment** only, if you **expected precisely** this attachment by the sender. If you are uncertain, whether to delete the message, collect further information. In no case use the contacts given in the mail. For example, call the sender. If you have opened Office programs and you are asked whether so-called macros are to be executed, think again about whether the message containing the respective file is of fraudulent character. Terminate the process for the time being.

**Figure 3.8:** First page of the interactive example.

edge remains. So that by going through all the hotspots on the page, the respective tips slowly build up similar to the order in the long text. Page 2 only shows the exact same text for rules four to six as the leaflet on one page. Page 3 is then again a new screenshot with further hot spots on the topic of dangerous attachments.

## 3.4 TORPEDO

In addition to the phishing awareness measures already described, there is a phishing tool support called TORPEDO that has also been was developed in 2015, published in 2016 [74] and has already been evaluated in two user studies. The first publication describes the development of a relatively simple tooltip with extended information to an even more complex tooltip with different representations [6]. Furthermore, the second study describes a small field study on the usability of the phishing tool support [75]. More about the basic structure of TORPEDO in the next section.

### 3.4.1 General Idea

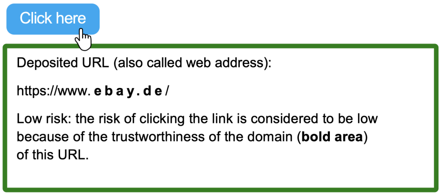
There are several basic assumptions that are made and on which the use of TORPEDO is based. One of them is that the obvious phishing emails e.g. larger campaigns on several people are automatically blocked by a system and do not even show up in the user's inbox. Accordingly, TORPEDO does not need another representation for such obvious phishing emails or links and does not include any form of block list. The focus of TORPEDO is therefore on links that cannot be automatically classified and must be decided by the user. Since its release, TORPEDO has not only existed as an artifact or proof-of-concept, but is now freely accessible to all users. TORPEDO can be used in its current form both in the web browser (1.8.1) for web email services such as Gmail or GMX or as an extension for Thunderbird (3.2). For the study in the later section, version 1.7.1 for the Google Chrome browser was used.

TORPEDO represents an extended tooltip that is intended to help the user make a faster and better decision regarding a URL, both in terms of content and design elements. As soon as the user moves the mouse over a link, the tooltip is displayed. TORPEDO makes a risk assessment based on the URL. Based on different rules, this risk is determined and the text and the display form are adapted accordingly. For example, a green frame appears with the indication of a low risk. More about the different risk levels later in Section 3.4.2. To better detect various phishing tricks, further modifications are made to the domain or top-level domain. Firstly, the domain is highlighted with bold font. This is intended to focus the user's attention even more on this area and to emphasize its importance. This is also intended to ensure that the overview is not lost in the case of particularly long subdomains. Secondly, a space is inserted between each character in this area. For example, “arnazon.com” becomes “a r n a z o n . c o m”. This should also help to detect such phishing tricks more easily. Since humans normally do not read words character by character, such phishing tricks can be easily overlooked in normal cases [63, 64]. With the inserted spaces this should be prevented or at least made more difficult.

### 3.4.2 Risk Levels

The current version of TORPEDO distinguishes between four different risk levels: Low Risk (Green, see Figure 3.9), Low Risk (Blue), Unknown Risk (Gray, see Figure 3.10), Unknown Risk with Indicator (Gray, see Figure 3.11).

The first risk level describes the case that the domain (in the future always a combination of domain and top-level domain) appears in a list of the most visited websites. At the moment this list is based on the Alexa list for the most visited websites with slight adjustments. The Alexa list is now no longer updated. However, at the time of installation it provided a good basis to have at least a certain base of websites that can be considered “safe”. Safe here not in the sense that there can be no problems on the website, but that it is initially safe to click on this domain. This list is static and cannot be changed by the user. In this case the tooltip gets a green frame and informs the user in the text that the domain is trustworthy and you have a low risk to click on it.



Deposited URL (also called web address):

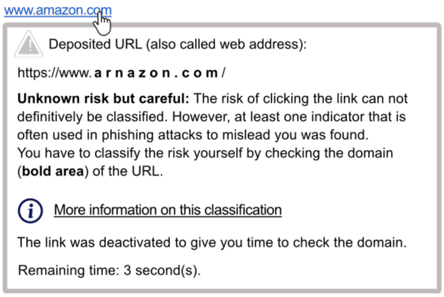
https://www. **e b a y . d e** /

Low risk: the risk of clicking the link is considered to be low because of the trustworthiness of the domain (**bold area**) of this URL.

**Figure 3.9:** TORPEDO’s low risk level warning.

A second low-risk level was introduced so that over time the user has to make fewer and fewer decisions based on a very high level of uncertainty. This describes domains that were either added manually by the user or clicked at least twice by the user. This list therefore becomes larger and larger as the tool is used, in contrast to the first case. The users have the option to add a domain to the list via the tooltip with a right click, as well as to expand the list for their own use via the settings directly at the start. In this case the tooltip gets a blue frame. In addition, the user is informed that based on previous assessments, the domain is considered low risk.

If none of the rules for the first two risk levels apply, the decision is made between two levels with unknown risk. Which of the two levels is displayed depends on whether additional indicators are found. For example, TORPEDO checks if there is a mismatch between link text and link URL, i.e. if a link text tries to suggest a URL and the link URL behind it is different. It also checks if the domain has the form of an IP. IPs are not readable for normal users and it is usually not known which target is hidden behind them. Both are no 100% certainty for a phishing attack, but at least an indicator. Therefore, this level gets a gray frame, a gray warning triangle in the upper left corner and a text that indicates to be particularly attentive and to check the URL, because at least one indicator was found. The gray frame is intended to distinguish itself from the two low-risk levels by its color. On the other hand it should symbolize the unfamiliarity. In the past, stronger signal colors were used here, but they turned out to be misleading in studies [76, 77]. A link can also be legitimate in the case of an assessment as Unknown Risk with Indicator and a too strong signal color has led to a high false positive rate.



www.amazon.com

⚠ Deposited URL (also called web address):

https://www.**arnazon.com**/

**Unknown risk but careful:** The risk of clicking the link can not definitively be classified. However, at least one indicator that is often used in phishing attacks to mislead you was found. You have to classify the risk yourself by checking the domain (**bold area**) of the URL.

ⓘ More information on this classification

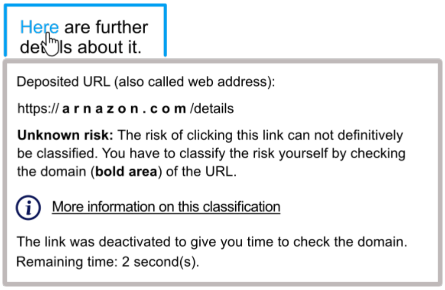
The link was deactivated to give you time to check the domain.

Remaining time: 3 second(s).

**Figure 3.10:** TORPEDO's unknown risk warning with indicator. Note, the "unknown risk with indicator" level does not necessarily means that it is a phishing link as shown here, a mismatch may also occur in benign emails.

If none of these indicators is found and also no entry in the lists for Low Risk, the Unknown Risk Level is displayed. This case thus represents the most open assessment. Neither a positive indicator was found, which speaks for a trustworthy domain. Nor was a negative indicator found, such as an IP or a mismatch, which speaks more for a phishing domain. The assessment therefore remains solely up to the user to decide. Accordingly, only a gray frame with an explanation is displayed, which makes clear to the user that the risk is unknown and must be decided by him.

For both levels, "unknown risk" and "unknown risk with indicator", clicking the link is blocked for three seconds. This is to prevent unintentional clicking while examining the link. It is also intended to result in no hasty gut decisions being made as a result of the deliberate interruption. Rather, it is intended to encourage conscious decisions. These three seconds are the default state and can also be switched on for the two low risk levels via the settings or reduced (or switched off) or increased in terms of duration.



Here are further details about it.

Deposited URL (also called web address):

https://**arnazon.com**/details

**Unknown risk:** The risk of clicking this link can not definitively be classified. You have to classify the risk yourself by checking the domain (**bold area**) of the URL.

ⓘ More information on this classification

The link was deactivated to give you time to check the domain. Remaining time: 2 second(s).

**Figure 3.11:** TORPEDO's unknown risk warning.

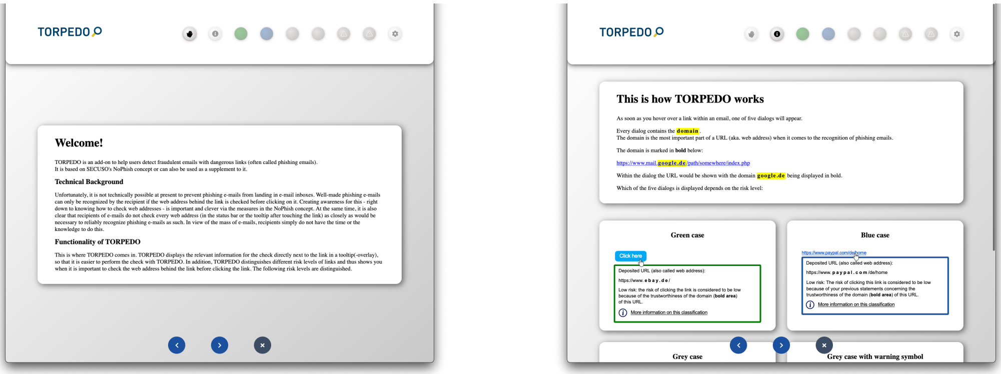
### 3.4.3 Tutorial

The TORPEDO intervention also includes a corresponding tutorial. This tutorial opens automatically directly after the installation and explains the most important functions of TORPEDO (see Figure 3.12 for the exemplary first two pages and Appendix A.2 for the complete tutorial.). It explains when the TORPEDO tooltips appear and which risk levels are available. It also gives the different risk levels and explains their meaning. At the end the settings with the possibilities for individualization are shown. Each step is accompanied by both text and pictures.

The tutorial starts with a short introduction to the technical background of why TORPEDO is needed. It explains that it is currently not technically possible to filter every phishing email before it arrives in the inbox. Therefore, it is the user's responsibility to make a decision. Accordingly, awareness can help with the decision, e.g. the NoPhish concept. But with the ever-increasing number of emails, there is no time for this. This is where TORPEDO helps by displaying important information directly in a tooltip. It is explained that TORPEDO distinguishes between different risk levels, which are described below.

The second page starts with a basic explanation of the functions of TORPEDO by focusing on the domain with a highlight. An example of where the domain is located within a URL is shown. The different risk levels are then presented, each with an example, before each risk level is discussed in more detail in the following pages.

The risk levels are then described on the next six pages. On each page, an example of the TORPEDO tooltip is shown on the left and explained in more detail on the right. The reasons for the classification and the specifics of the risk level, e.g. link disabled for three seconds, are explained. Two pages are used for each unknown risk level. The first page shows a phishing example and the second page shows a legitimate example. This is to make it clear that the unknown risk levels are not necessarily phishing. The tutorial concludes on the last page with the special case of short URL and redirect, which are resolved by TORPEDO. In addition, the options available to the user are mentioned.



**Welcome!**

TORPEDO is an add-on to help users detect fraudulent emails with dangerous links (often called phishing emails).
It is based on SECUSO's NoPhish concept or can also be used as a supplement to it.

**Technical Background**

Unfortunately, it is not technically possible at present to prevent phishing e-mails from landing in e-mail inboxes. Well-made phishing e-mails can only be recognized by the recipient if the web address behind the link is checked before clicking on it. Creating awareness for this - right down to knowing how to check web addresses - is important and clever via the measures in the NoPhish concept. At the same time, it is also clear that recipients of e-mails do not check every web address (in the status bar or the tooltip after touching the link) as closely as would be necessary to reliably recognize phishing e-mails as such. In view of the mass of e-mails, recipients simply do not have the time or the knowledge to do this.

**Functionality of TORPEDO**

This is where TORPEDO comes in. TORPEDO displays the relevant information for the check directly next to the link in a tooltip(-overlay), so that it is easier to perform the check with TORPEDO. In addition, TORPEDO distinguishes different risk levels of links and thus shows you when it is important to check the web address behind the link before clicking the link. The following risk levels are distinguished.

**This is how TORPEDO works**

As soon as you hover over a link within an email, one of five dialogs will appear.

Every dialog contains the **domain**.
The domain is the most important part of a URL (aka. web address) when it comes to the recognition of phishing emails.

The domain is marked in **bold** below:

https://www.mail.**google.de**/path/somewhere/index.php

Within the dialog the URL would be shown with the domain **google.de** being displayed in bold.

Which of the five dialogs is displayed depends on the risk level:

**Green case**

Deposited URL (also called web address):

https://www.**ebay.de**/

Low risk: the risk of clicking the link is considered to be low because of the trustworthiness of the domain (**bold area**) of this URL.

More information on this classification

**Blue case**

https://www.paypal.com/de/home

Deposited URL (also called web address):

https://www.**paypal.com**/de/home

Low risk: The risk of clicking this link is considered to be low because of your previous statements concerning the trustworthiness of the domain (**bold area**) of this URL.

More information on this classification

**Grey case**

**Grey case with warning symbol**

**(a)** TORPEDO's tutorial page 1.

**(b)** TORPEDO's tutorial page 2.

**Figure 3.12:** Exemplary the first two pages of the original tutorial.

## 3.5 Dimensions of Study Examples

Next to the different phishing tricks that are the most important aspect of the examples used to evaluate the effectiveness, there exist also other dimensions which lead to a complete description of the example: 1) phishing or legitimate; 2) phishing type; 3) phishing tricks; 4) email sender; 5) interaction type; 5) environment; 6) context.

- Phishing or legitimate: The overall example being either legitimate or phishing
- Phishing type: Content (self-made or cloned), Sender, URL, attachment
- Phishing trick: see Section 3.1
- Email sender: Organization or person (known or unknown)
- Interaction type: static screenshot, interactive screenshot, actual inbox, simulated phishing to own inbox
- Environment: OS (Windows, macOS, Linux, iOS, Android etc.), Client (Thunderbird, Mail, Outlook, etc.), URL display (Statusbar and/or Tooltip)

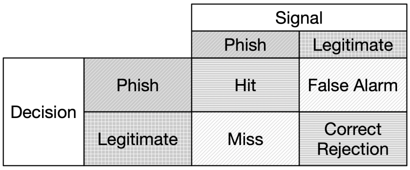
The first dimension is the categorization of the example either being legitimate or phishing. Some dimensions recreate the way phishers create their messages e.g. what aspects they can manipulate. The phisher can create the content of the message from scratch e.g. when impersonating an message from a known person rather than from an organization, or he can also clone an existing message from an organization and change the original link to his own URL or change the attachment. The phisher can also spoof the sender so that for the normal user the message looks like it was send from the impersonate person (known or unknown to the user) or organization. There are also some more general dimensions to describe an example e.g. the design of the environment the email seems to be from or the way the participant can interact with the email during the study. The environment can have various different characteristics such as different OS and for those OSs different clients. Also within a study there is the option to have a variety of characteristics in mixing different OSs, Clients and URL display or stay with a simple combination of such characteristics, e.g., windows examples in the Thunderbird client with a status bar. Lastly studies can have different possibilities for the interaction with such examples during the study, e.g., static screenshots that display a specific status and no interaction is possible. An other alternative is the possibility to use static screenshot and add interactive elements such as hover events for showing to URL similar to an email client. Also one could give participants access to an actual email inbox or send emails as part of a phishing campaign to their own inbox.

In the later parts of this thesis every study methodology describe with which characteristics every dimension was used.

## 3.6 Signal Detection Theory

In the past, studies have already evaluated the effectiveness of measures and for this purpose, the rate of correct answers for phishing examples was often used [53, 61, 6]. However, this only considers a part of the messages to be evaluated that come into consideration in the phishing context. In reality, there is the confrontation with both phishing messages and legitimate messages and the right decision must be made for each.

Depending on the definition of the target, i.e. whether phishing messages were recognized as phish or legitimate messages as legitimate, a different four-field matrix results for the decision. Assuming that the target is to correctly identify phishing as such , there are four possible assessments of the decision (see also Figure 3.13): 1) The decision is phishing and the example is a phish equals a “Hit”; 2) The decision is phishing and the example is legitimate equals a “False Alarm”; 3) The decision is made in favor of legitimate and the example is legitimate equals a “Correct Rejection”; 4) The decision is legitimate and the example is phishing equals a “Miss”.



| | | Signal | |
|---|---|---|---|
| | | Phish | Legitimate |
| Decision | Phish | Hit | False Alarm |
| | Legitimate | Miss | Correct Rejection |

**Figure 3.13:** Overview of the combination of the dimensions decision and signal in a 4x4 fields matrix.

As soon as only phishing examples are used to evaluate a measure and a decision is asked for, only the hit or miss rate can be considered. There is a risk that a measure that generates excessive fear leads to an increased hit rate, but the false alarm rate, which is likely to increase at the same time, is not taken into account. In reality, an increased hit rate is desirable. Nevertheless, the question arises as to whether an increased false alarm rate undermines the measure for future use and may even have more negative effects in the long term. After an increased number of false alarms, users could lose confidence in the measure and no longer follow also meaningful behavioral cues in the future. Such a loss of trust would then require further measures.

The Signal Detection Theory [78] involves all four types of decision. It has also been shown to be helpful in the past in individual studies in the context of phishing: Some used it to evaluate a phishing game with pre-test and post-test evaluation [3], differences between simple comic and more complex games [79] or visual discrimination training [80]. Additionally, it has also often been used in the context of modeling human behavior such as finding the most vulnerable users [81], modeling security behavior [82, 83] or comparing different forms of it [84]. Besides, the theory has also been applied to check for other factors influencing the decisions like individual or cultural differences [85], confidence [86], task-level variables [87] or conscientiousness [88]. Furthermore, aspects of the email such as authority, scarcity and social proof have been examined for their influence on the decisions [89].

In particular, it reduces the four types of decision to two separate values: the sensitivity ($d'$) and the criterion ($C$). The sensitivity describes the ability to distinguish between the signal (in this context = phishing) and the noise (in this context = legitimate). The greater the sensitivity, the greater this discriminative ability. It is, so to speak, a combination of the hit and the correct rejection rate. The criterion describes the general tendency of all decisions either in the direction of signal or noise. A “perfect” decision would be according to the criterion = 0. It would have neither the tendency to make a decision in the direction of phishing or legitimate. Therefore, there is neither a tendentially increased miss rate nor an increased false alarm rate.

The following section describes how the two values relevant to this thesis, sensitivity and criterion, are calculated. First, the theoretical basis and formula for the calculation is described, and afterwards, it is exemplified how the calculation is carried out.

First, the literature describes the following formula for calculating the sensitivity.

$$d' = z(\text{Hit}) - z(\text{FA})$$

"Hit" describes the ratio of correctly recognized phishing examples to the total number of phishing examples. Accordingly, the number ranges between 0 and 1.0, where 0 means that no phishing example was recognized and 1.0 means that all phishing examples were recognized as such. "FA" stands for false alarm and describes the number of falsely categorized legitimate examples as phish by the total number of legitimate examples. This value also ranges from 0 to 1.0.

Due to the $z$-transformation of the values the exact boundaries depend on the amount of signals and noise used for the study. In this thesis, the boundaries of the respective study are reported in the corresponding study design part. In general, a sensitivity of 0 describes a neutral sensitivity and would correspond to the guessing probability of 50% correct answers. At the threshold values in the range of 100% of correctly recognized phishing examples and 100% of correctly recognized legitimate examples, it then becomes very steep so that the number of examples strongly influences the final value for $d'$. An example for 90% correct answers for both, hit rate and correct rejection rate, is a value of $d'$ = 2.563, and for 95% correct answers, a value of $d'$ = 3.29.

Both the "Hit" value and the "FA" value are also relevant for calculating the criterion. The formula for the criterion is as follows:

$$C = -\frac{z(\text{Hit}) + z(\text{FA})}{2}$$

The calculations in this thesis are based on the r package "psycho" [90], which contains all necessary values such as sensitivity ($d'$) and the criterion ($C$).

Signal detection theory encompasses two key values: sensitivity and criterion. This thesis primarily focuses on the ability to distinguish between phishing and legitimate messages (sensitivity). Consequently, all hypotheses in the subsequent chapters related to the signal detection theory are based on the sensitivity. The objective of interventions is to enhance sensitivity. However, it is also important to examine how the decision-making process evolves: if an intervention improves the phishing detection rate but at the same time increases the false alarm rate, then only the overall alertness has been increased. Therefore, in addition to the general ability to discriminate (sensitivity), the criterion should provide important information about the decision-making tendency. With regard to the criterion, usually, no hypotheses are formulated due to the lack of clarity in prior experiences. Nevertheless, statistical significance testing is used to enable detecting a significant increase of alertness. Depending on the direction and extensiveness of the significant development of the criterion, the significant improvement of the sensitivity needs further discussion.

# 4 Part I: Long-term Effects of Phishing Awareness Measures

Knowledge transfer is not a one-off event [25]. It is based on a continuous retrieving process which is subject to forgetting. Consequently, the retention of any acquired knowledge depends on appropriate refresher measures. Furthermore, the particular effectiveness and efficiency of a refresher itself depends on timing. In this sense, the measures should ideally be implemented at a point in time before significant forgetting sets in. In the field of phishing prevention, successful initial transfer of awareness is therefore only one step in the learning process, which must be followed by a second step in order to retain the awareness in the longer term. Thus, the fundamental question is, when have the participants in a phishing awareness measure lost so much awareness that they are no longer significantly better than before the measure?

To date, there is very limited research on the long-term effect of phishing awareness measures and there is no research on refreshment measures. Jensen et al. [29] evaluated two different phishing awareness measures (rule-based and mindfulness training) with significant results for phishing detection ten days later. Kumaraguru et al. [31] analyzed the effect of a phishing awareness measure (two short messages) and found significant results for a period of under a month. Similar research was conducted by Nguyen et al. [32] for two phishing awareness measures (rule-based and mindfulness training) crossed with overlearning with significant results for mindfulness training after two months. Canova et al. [33] studied another phishing awareness measure (game) and found significant more correct answers five months after playing the game.

Part I of this thesis contributes to the field of research about long-term effects of phishing awareness measures by evaluating different type of measures: video, e-learning, workshop, long text (see Section 3.3.4), short text (see Section 3.3.5), and interactive email example, in different settings (private context, university, and organization), with different types of study designs (between- and within-subject), while always using the signal detection theory (sensitivity and criterion) to measure participants awareness. While both are important and described in the research question and hypotheses sections, the hypotheses are focused on sensitivity. Sensitivity should be explicitly mentioned here, as it enables the precise measurement of improvements in capability. While the criterion from Signal Detection Theory also represents an important measure and should thus be described descriptively in addition. As previously mentioned (see Section 3.6), in the area of the criterion, several aspects are crucial for describing positive development, e.g., the criterion of the groups being compared, as well as the absolute position of the assessed criterion in comparison to neutrality, which is at zero. Besides the difficulty of categorizing the development of the criterion, both measures are interrelated, and a significantly improved sensitivity represents a notably more important development.

The chapter is structured into three sections each on a specific study. Every study section includes the research questions of the study, the study methodology (study design and study material), the study results and study specific limitations. In the first study, a video is evaluated, the second study focuses on an e-learning and the third study tests a workshop together with the video, a long text, a short text and an interactive email example as refreshment measures. At the end of the chapter, there is a joint discussion of all results. At this point, the results of the overall research questions are summarized and discussed in context. In addition to the subsections of the studies with their limitations, there is a final section with limitations that apply to all three studies.

**Contributions described in this chapter:**

- Even a very brief measure (e.g, an anti-phishing awareness video of five minutes) significantly improved users' ability to distinguish between fraudulent and legitimate messages (called user sensitivity) for a two-month period [6].
- A more extensive and interactive measure (e.g., an anti-phishing awareness e-learning), which included exercises and lasted around two hours, significantly improved user sensitivity over a period of at least five months [7].
- An anti-phishing workshop with the train-the-trainer approach also including exercises with a duration of around three to four hours significantly improved user sensitivity for a period of around four months [8].
- When user sensitivity is no longer significantly improved (i.e., from six months onward and sometimes before), the refresher measures of video, long text and interactive email example again achieved significant improvements. Furthermore, these three refresher measures achieved significant long-term effects for a further six months, helping to improve awareness over twelve months [8].
- In two out of three of the studies, the lowest correct answer values are measured for the phishing trick "Small Deviations in the Domain" (e.g., `luftthansa.com`). Therefore, the most important focus for future anti-phishing interventions should be on this difficult-to-identify phishing trick.

## 4.1 Video

To address problem of phishing, various forms of measures have been developed to train users in awareness and to support them in building effective protection together with technical tools. For example, apps or even longer texts or entire training courses were developed to provide effective support in varying degrees of detail. These have also proven to be effective by significantly improving phishing detection. However, these measures often took between half an hour and a whole hour. This carries the risk that both companies and organizations will refrain from using them due to the time required, and also that normal users will be discouraged. In addition, the measures were often visually unappealing forms of awareness transfer or at least very static forms without moving elements that can attract attention.

The main advantage of videos in the context of phishing is that they can be comparatively short, as the information is presented not only as text, but also as a visual cue. In practice, this means that organizations save a lot of time. The question is whether a short video is just as effective as longer or more extensive measures and how long the effect can be expected to last. Previously, a study that evaluated a game found significant improvements for the overall detection rate after five months. The game took the participants 30 minutes in the study to complete and therefore around six times the duration of the video from Section 3.3.1. Additionally, as other studies with shorter measures [31] found significant results after around a month . As the video is more comparable in length to the shorter measures, the decision was made to go for the conservative approach with a less extensive long-term measurement of eight weeks.

For this reason, the video (see Section 3.3.1) was examined for its effectiveness with a sample of 89 participants. Particular attention was paid not only to short-term effectiveness, but also to whether the video could contribute to significant improvement over a longer period of time. As part of this, participants were assessed on their ability to distinguish between phishing and legitimate messages before watching the video, after watching the video, and eight weeks later. First, the general effectiveness and a long-term but not to distant retention period should be tested before moving on to more extended retention periods. The results showed that watching the video both immediately afterwards and eight weeks later led to significantly higher sensitivity. Suggestions were also made as to which particular phishing tricks should be repeated or refreshed after the eight-weeks.

**Parts of the results described in this chapter have been published in:**

- Volkamer, M., Renaud, K., Reinheimer, B., Rack, P., Ghiglieri, M., Mayer, P., Kunz, A. and Gerber, N. (2018). Developing and evaluating a five minute phishing awareness video. In Trust, Privacy and Security in Digital Business: 15th International Conference, TrustBus 2018, Regensburg, Germany, September 5–6, 2018, Proceedings 15 (pp. 119-134). Springer International Publishing.

### 4.1.1 Research Questions and Hypotheses

In the following, the concrete research questions and the related hypotheses are presented. The general idea is to learn more about the effectiveness of the sensitivity ($d'$) for the video directly after the measure and in the long-term, as well as to get differentiated findings with regard to the effectiveness of the specific phishing tricks. The question now is whether such a short and passive measure as the video over eight weeks can also significantly improve sensitivity. Also, due to the time limit of the video, all phishing tricks were explained, but in a relatively short scope. This condensed presentation could lead to differences in the phishing trick results over a longer period of time. Based on these open questions, the first research question is formulated:

**RQ**$_1$: *Is there a long-term effect of the video as a phishing awareness measure?*

Before a long-term effect can be measured it should be made clear that the measure itself provides an improvement in the pre-post comparison. Otherwise, a long-term measurement would not calculate any effects that should be associated with the measure. This leads to the following two hypotheses, whereby the fulfillment of the first hypothesis is a prerequisite for testing the second main hypothesis:

**H**$_1$: Participants show significantly higher sensitivity immediately after video compared to the pre-test.

**H**$_2$: Participants show significantly higher sensitivity eight weeks after the video compared to the pre-test.

In addition to this significant effectiveness in increasing sensitivity, there is also the question of whether the measure has a different effect on different phishing tricks over time. Previous research on a game [33] found that some phishing tricks like "Non-brand related Domain + Brand Outside" remained at the same level, while others such as "Small Deviations in the Domain" decreased over time. Now, it would be interesting to know whether these trends are similar or different for the shorter time period and the video as a more passive and compressed measure.

**RQ**$_2$: *Does the long-term effect of the video as a phishing awareness measure differ for specific phishing tricks?*

### 4.1.2 Methodology

This section describes the procedure for evaluating the above research questions and hypotheses. First, the study design and the procedure of the study with the individual phases (see Section 4.1.2.1) are described in detail. Particular attention is paid to the material used to evaluate the sensitivity (see Section 4.1.2.2). For this purpose, a kind of quiz was used, in which the participants had to answer the question whether an example is phishing and legitimate. It is explained how the structure of each example looks like and which phishing tricks respectively manipulation strategies were used. Afterwards, the sample of the study and the ethical considerations are explained (see Section 4.1.2.3).

#### 4.1.2.1 Study Design

The study was conducted using a repeated measures design. This involved repeated surveys of all participants at three fixed measurement points in time. The survey took the form of an online experiment on the Sosci Survey platform. All participants were surveyed before watching the video, immediately after the video, and a third time after eight weeks. Accordingly, there were two survey dates, with the pre-measurement, the intervention and the post-measurement taking place at one point in time and the retention measurement taking place at a second point in time after eight weeks (see Figure 4.1). Thus, the structure of the first phase was as follows for the study design):

First, participants were educated about the study procedure. Subsequently, the participants received an introduction to the scenario. The goal here was to ensure that the participants did not directly reject examples, e.g., from providers they did not use themselves, as phishing based on their own experience. Therefore, participants were given the following text:

> “To determine how well you can recognize fraudulent messages, the following are 16 examples with a messages in different contexts. Knowing that you may not know or use all senders, service providers, operating systems and programs, we present received messages for different senders, service providers, operating systems and programs. In order to better evaluate the senders, service providers, operating systems and programs to which you have no connection in reality, please assume in the following that...
>
> You are Martin Müller. Your email address is martin.mueller.77@web.de
> Jonas Schmidt is your colleague. His email address is jonas.schmidt.77@web.de
> You are using all services used in this questionnaire.
>
> Readability and enlargement of the images:
> You can use the keyboard shortcuts on your keyboard to resize the images. You can use “Ctrl/cmd” + “+” to enlarge any web page and “Ctrl/cmd“ + “-” to reduce it. You can also use the “Ctrl/Ctrl/cmd” key and the mouse wheel to change the display size.
>
> To start now with the quiz, please click on next” [6]

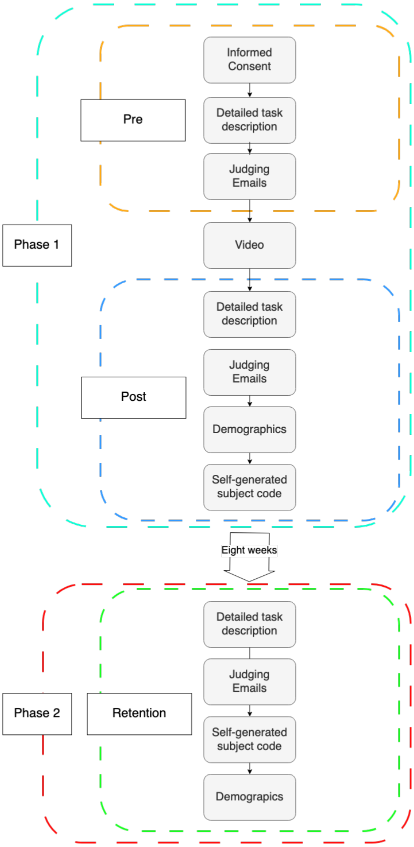

Informed Consent
Detailed task description
Judging Emails
Pre
Phase 1
Video
Detailed task description
Judging Emails
Post
Demographics
Self-generated subject code
Eight weeks
Detailed task description
Judging Emails
Phase 2
Retention
Self-generated subject code
Demograpics

**Figure 4.1:** Study design for the three time points (pre, post, and retention).

The scenario was followed by 16 example messages (see Section 4.1.2.2 for more details about the examples). One example was displayed per page. The examples were presented in a separate random order for each participant. On the page the question was asked: Is this message fraudulent? This was accompanied by the choices “Fraudulent” and “Not Fraudulent”. Fraudulent would mean a classification as phishing.

Afterwards, participants had to create a subject code. This is called personal coding and had the purpose to ensure a pseudo-anonym assignment of the data between the measurement times. The code was generated from data that only the participant knows and that should not change during the study, e.g., the first letters of the parents or parts of their birth data. Finally, participants were asked to provide an email address if they wished to be invited for the second phase. This email address was stored separately from the study data using a special feature of Sosci Survey, so there was no possibility of linkage.

For the retention phase participants received again the scenario with a detailed task description to refresh their knowledge about what they are supposed to do. Then the same 16 examples as in both the other phases were displayed again. As before every example was displayed on a single page with no option to move back and forth between the examples. To connect all the data between the two phases again the same method to create a subject code was presented and participants had to follow the guidelines.

#### 4.1.2.2 Study Material

This subsection deals with the study material used to measure the sensitivity in the pre-test, post-test and the two month retention-test. For testing the sensitivity, screenshots of emails were used. These examples are based on messages that were sent in exactly this form by the used organization (like Lufthansa) or real existing persons (changed to Jonas Schmidt). The organizations used represent well-known organizations that are familiar to most people in Germany like Google, Vodafone, 1und1 or Bahn. Even if people are not familiar with the specific example they should be familiar with the topic or organization and not reject the example due to the unfamiliarity. The language of the emails was German.

For all messages, the content, the sender and the form were plausible, so that these factors do not act as confounding variables that make a message appear to be phishing. Accordingly, the assessment of whether a message is phishing or legitimate should only be made on the basis of the manipulated factor (URL). Half of the examples (n = 8) were legitimate examples and the other half were phishing. No information was provided about this distribution to the participants. Examples were collected through various channels from both organizations and private individuals.

All images of the emails were static i.e. there was no way to interact with the example screenshots. The mouse was already placed over the respective link in the images, so the URL of the link was already visible. Depending on the software displayed, the URL was either visible in the tooltip (for the Outlook examples) or in the status bar (for the web browser or Thunderbird examples). The display as status bar or tooltip was also evenly distributed across the phishing or legitimate examples. The phishing examples were intended to cover all of the tips from the video. For more information about the actual URLs used or how the URL was presented, the subject and the sender, see Table 4.1 for the phishing examples and Table 4.2 for the legitimate ones. One email is representing the phishing trick “Non-Brand related Domain” with an IP in the domain. Two emails fall in the category of “Non-brand related Domain + Brand Outside” with one being a subdomain trick (`google.com.best-photos.com`) and one with a path trick (`zehrukol.com/ebay.com`). A total of two emails represent the “Small Deviations in the Domain” with one using a similar letter combination and one a single typo in the domain. Three emails represent “Special Link

Manipulation" with all of them using a mismatch between the link text (looking like a legit URL) and the actual link URL. One example is simpler being a random URL for the actual URL, but the other two even used additionally the "Small Deviations in the Domain" trick so that both had a letter swapped compared to the actual legitimate domain.

#### 4.1.2.3 Recruitment & Ethical / Data Protection Considerations

Participants in the study were recruited through a snowball system using various channels such as social networks, online platforms, leaflets, and personal invitations. There was no monetary compensation for the participants, rather they were told that they would receive idle compensation by gaining awareness in dealing with phishing.

The ethical requirements followed the guidelines of the Technical University of Darmstadt and its ethics committee. This includes that all data is stored independent of specific data of the person. The stored email addresses were used exclusively for the purpose of contacting participants for retention and were subsequently deleted. In addition, the data were stored in a separate database independent of the actual study data. This was done via an integrated function of the Sosci Survey study platform [91]. Participants' data could be linked solely by the subject code they generated and entered for each study time point. This was done at the risk of data being unassigned and therefore discarded in the event of errors in generation. In the end, however, it represented the best balance for preserving pseudonymity in combination with the study's interest in linking the data.

### 4.1.3 Results

Overall, the sample was n=89, but those that took part in the phase 2 eight weeks later consisted of 22 participants. Of these, 12 were female and 10 male, and the mean age was 38.09 years with a standard deviation of 12.72. The minimum age was 20 years, the maximum 61 years. Additionally, 10 had a university degree, 5 had an a-level qualification, and 7 had some other form of education.

The following section reports the results in terms of the detection rate of phishing and legitimate for pre and post as well as in the reduced retention sample. Before moving to the sections with the hypothesis tests, the descriptive statistics of the average detection rates for phishing, legitimate and all examples as well as sensitivity and criterion for all three tests are reported (see Table 4.3)

Table 4.1: Overview of presented phishing messages. "TT" stands for tooltip and "ST" stands for status bar.

| Name | Phishing Trick | More specific Phish | Sender | Subject | TT or ST | Indicator |
|---|---|---|---|---|---|---|
| P1 | Non-Brand related Domain | IP | bestellstatus@eventim.de | eventim.de - Shipping notification - Your order no. 342491238 | Tooltip | 162.179.34.56/login |
| P2 | Non-brand related Domain + Brand Outside | Subdomain | None | Your Google Photos storage space is almost used up | Status | www.google.com.best-photos.com/... |
| P3 | Non-brand related Domain + Brand Outside | Path | Jonas Schmidt | Software package | Tooltip | www.zehrukol.com/ebay.com/... |
| P4 | Small Deviations in the Domain | Similar Letter | None | Your My Packages service - change of tariff ordered | Status | control-center.1uncl1.de/... |
| P5 | Small Deviations in the Domain | Typo | gewinnbestaetigung@hujilorik.cn | Travel three times, save once - quickly secure a free ride! | Tooltip | www.bahncard.bahm.de/... |
| P6 | Special Link Manipulation | Fake Tooltip | congstar <kundenservice@congstar.de> | Your congstar invoice | Status | www.cognstar.de/... |
| P7 | Special Link Manipulation | Fake Tooltip | Volksbank Südhessen eG | Confirmation of your newsletter subscription | Tooltip | www.volksbanknig.de/... |
| P8 | Special Link Manipulation | Mismatch | Jonas Schmidt <jonas.schmidt.77@web.de> | Notes | Status | www.secure-documents-online.com/... |

**Table 4.2:** Overview of presented legitimate messages. In the sixth column TT stands for tooltip and ST stands for status bar.

| Name | Sender | Subject | TT or ST | Indicator |
|---|---|---|---|---|
| L1 | Apple Market Research <MarketResearch_Europe@InsideApple.apple.com> | Tell us about your iPhone | Status | https://marketresearch.apple.com/SE?Q_DL=231h534j1234h\T1\ss0823rhjas/=3221akdsjoiqhdas |
| L2 | Dropbox <no-reply@dropbox.com> | You have an Android device connected to Dropbox. | Tooltip | https://www.dropbox.com/l/0Gn8vLUksmfUQ02knHbHkH |
| L3 | Jonas Schmidt | None | Dialog | https://photos.google.com/share/AFQ1ii... |
| L4 | Jonas Schmidt <jonas.schmidt.78@web.de> | Research assistant at the TU Darmstadt | Tooltip | https://www.tu-darmstadt.de/forschen/id=324438213 |
| L5 | Lufthansa | Travel information for your Lufthansa flight to London on 08.09 | Status | https://buchung.lufthansa.com/servlet/cc?soDBYCTTDVYTEz0.26wa7uDU.261f7uuF.3df4D.2e.26EaEXEPNRTOOL_LINKEhttp:DVMDVredirect.2elufthansa.2ecomDVMdefault.2easpxDWMd.3http:DVM |
| L6 | Vodafone Team <nicht.antworten@kundenservice.vodafone.com> | Link to change the internet password for MyVodafone | Tooltip | https://www.vodafone.de/ussa/resetPasswordLink/rese.. |
| L7 | None | None | Dialog | https://accout.wire.com/reset/?key=Z9zUm2Q9MGR_unXEJi_OLOx-t5jknu3xceDWNnOTcU=&code=VGWoCWPgNtLoEwMN9Oylbaf aVbywlskKR |
| L8 | register@gutefrage.net | Please confirm registration | Tooltip | https://www.gutefrage.net/registrierungsbestaetigung/martinmueller77/754c043dbfb63c8dd1080fb4d141109c |

**Table 4.3:** Overview of the average rates for all three points in time in % for the correct answers for phishing and legitimate examples as well as overall. Also adding the values for the signal detection theory.

| Example Type | Pre | Post | Retention |
|---|---|---|---|
| Phishing | 42.6 | 86.9 | 81.3 |
| Legitimate | 75.0 | 88.1 | 83.0 |
| Overall | 58.8 | 87.5 | 82.15 |
| Values Signal Detection Theory | | | |
| Sensitivity | 0.56 | 2.23 | 1.83 |
| Criterion | 0.43 | 0.03 | 0.05 |

To test the hypotheses H1 and H2 repeated measures ANOVAs were calculated that analyzed the 3 time points or pre-, post-, and retention-test. Two different repeated measures ANOVAs were conducted to evaluate sensitivity and criterion.

#### 4.1.3.1 RQ1 - Long-term Effect

To ensure the comparability of the studies, a value had to be used that was consistent across all studies. Here, the values of the Signal Detection Theory should be taken since they best describe the ability to distinguish between legitimate and phishing examples, as described in Section 3.6. As the original study did not include Signal Detection Theory values, these values were recalculated as part of the thesis. The original data was taken for this purpose. In addition, the same data-cleaning procedure was used in the study.

A repeated measures ANOVA was used to compare the sample related to the sensitivity and criterion. Sensitivity was statistically significantly different at the three measurement times, F(2,54) = 23.385, $p < .0001$, $eta_g^2$ = .464. Following the specifications of Cohen [92] (.01 > small effect, .06 > medium effect and .14 >= large effect), $\eta^2$ = .464 is a large effect.

The post-hoc tests with Bonferroni correction showed that both the post-test groups (p < .0001) and the retention test (p < .0001) differed significantly in sensitivity from the pre-test group.

The post-hoc tests with Bonferroni correction showed that both the post-test ($p < .0001$) and the retention-test ($p < .0001$) differed significantly in sensitivity from the pre-test (see Figure 4.2 for a boxplot of the post-hoc tests). In contrast, post-test and retention-test did not significantly different from each other ($p = .109$). The descriptive statistics show that the sensitivity was initially low ($d'$ = .559). The sensitivity increased for the post-test ($d'$ = 2.23) and decreased slightly over the two months until the retention time ($d'$ = 1.83).

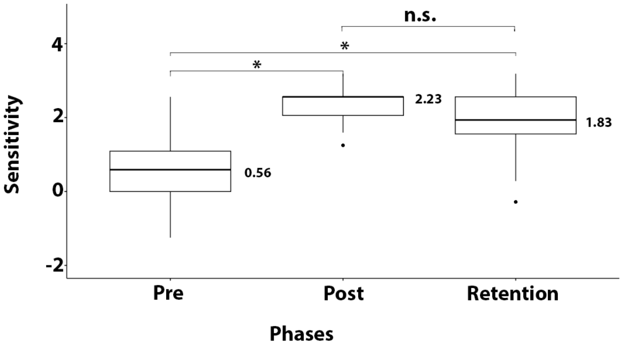


**Figure 4.2:** Boxplot for sensitivity for all three points in time. The mean sensitivity for the sample is on the right side of the boxplots. "*" marking those post-hoc comparison that were significant and "n.s." those that were not significant.

The repeated measures ANOVA for the criterion was statistically significantly different across the three measurement time points, F(2,54) = 8.22, $p = .0008$, $eta_g^2$ = .233. Following the specifications of Cohen [92] (.01 > small effect, .06 > medium effect and .14 >= large effect), $\eta^2$ = .233 is a large effect.

The post-hoc tests with Bonferroni correction showed that both the post-test ($p = .007$) and the retention-test ($p = .016$) differed significantly for the criterion from the pre-test. Furthermore, post-test and retention-test did not significantly differ ($p = 1$). Based on the descriptive statistics, it can be seen that there was initially a high criterion ($C$ = .43) that decreased to the post-test ($C$ = .029) and then only slightly increased two months later at the the retention-test ($C$ = 0.05).

#### 4.1.3.2 RQ2 - Phishing Tricks

This section starts looking at the individual examples, both phishing and legitimate. For this purpose, the average of the correct answers of all participants is considered. The detailed listing with mean and standard deviation for phishing URLs can be seen in Table 4.5 and the similar listing for legitimate URLs in Table 4.6.

**Phishing Tricks** When looking at the phishing tricks (see Table 4.4), it becomes clear that none of the phishing tricks were easy to recognize at pre-test before the intervention. Initially, all tricks were even below 50% correct answers with “Special Link Manipulation” (36.36%) the lowest, followed by “Non-brand related Domain” (40.91%). The two tricks “Non-brand related Domain + Brand Outside” and “Small Deviations in the Domain” performed the best, but still at guess

**Table 4.4:** Detection rate in % for phishing tricks for the two phases.

| Phishing Trick | Pre | Post | Retention (eight weeks) |
|---|---|---|---|
| Non-brand related Domain | 40.91 | 95.45 | 81.82 |
| Non-brand related Domain + Brand Outside | 47.73 | 81.82 | 68.18 |
| Small Deviations in the Domain | 47.73 | 95.45 | 90.91 |
| Special Link Manipulation | 36.36 | 81.82 | 83.33 |

probability with 47.73% each. Immediately after the video, the order shifts. All phishing tricks were now recognized much better with values above 80%. Non-brand related Domain + Brand Outside and Special Link Manipulation each achieve 81.82%. Non-brand related domain" and "Small Deviations in the Domain" performed even better with 95.45% each. Over the next eight weeks, the values for three of the four phishing tricks had fallen again. "Non-brand related Domain + Brand Outside" then only achieved a value of 68.18% and thus brought up the rear. "Special Link Manipulation" remained relatively constant and still reached 83.33%, while "Non-Brand related Domain" lost almost 15% and only reaches 81.82%. 'Small Deviations in the Domain" then achieved the best value with slight losses at 90.91%.

**Phishing Examples** Looking at the Table 4.5, it is noticeable that before watching the video, P6 performed the worst with a value of 22.73% way below the guessing rate. P2 and P8 did not fare much better with 36.36% and P1 and P4 with 40.91%. The example P7 was right on the spot of the guessing rate with 50%, while two examples were slightly above (P5 with 54.55% and P3 with 59.09%).

Basically, it can be stated that for all examples the number of correct answers increased from the pre-test to the post-test. But the examples benefited differently. Furthermore, the order of the best recognized examples changed slightly. P3 moved from first to last place with only 68% correct answers, while P8 remained on the second to last place with 72.73% (even though improving by 36.36%). Except for P6 with 81.82% all other examples achieved a score of 90% (P7) or even above 95% (P2, P1, P4 and P5). So except for P3, all other examples improved around 36.36% (P8) to even 59.09% from pre-test to post-test.

Comparing the post-test and retention-test, some anomalies can also be seen. While some examples showed only a very small change, e.g. P4, P7 and P5 with -4.55%, the number of correct answers for other examples decreased much more, e.g. P2, P1 and P3 with -13.64%. Two examples even slightly improved their score with 4.55% more correct answers (P6 and P8). P3 moved back to it's previous level and even slightly below with only 54.55% and way below all

**Table 4.5:** Detection rate in % for individual phishing URLs for the two phases.

| Name | Phishing Trick | Pre | Post | Retention (eight weeks) |
|---|---|---|---|---|
| P1 | Non-Brand related Domain | 40.91 | 95.45 | 81.82 |
| P2 | Non-brand related Domain + Brand Outside | 36.36 | 95.45 | 81.82 |
| P3 | Non-brand related Domain + Brand Outside | 59.09 | 68.18 | 54.55 |
| P4 | Small Deviations in the Domain | 40.91 | 95.45 | 90.91 |
| P5 | Small Deviations in the Domain | 54.55 | 95.45 | 90.91 |
| P6 | Special Link Manipulation | 22.73 | 81.82 | 86.36 |
| P7 | Special Link Manipulation | 50.00 | 90.91 | 86.36 |
| P8 | Special Link Manipulation | 36.36 | 72.73 | 77.27 |

the other examples. Two examples still achieved a score of 90.91% (P4 and P5), two a score of 86.36% (P6 and P7) and two a score of 81.82% (P2 and P1).

**Legitimate Examples** For the legitimate examples (see Table 4.6) there are also differences between the examples for the number of correct answers. In the pre-test, the example L6 achieved the worst value with 50%, followed by the examples L1 with 54.55% and L7 with 54.55%. In contrast, the example L4 already reached a value of 100%. Two other examples, L3 and L5, achieved already scores of 90.91%. While the other two examples achieved mediocre scores: L2 with 77.27% and L8 with 81.81%.

In the comparison between pre-test and post-test, it is noticeable that the top spot continued to be taken by L4 with 100% - closely followed by L8 and L3 with 95.45%. Now the worst example was L2 with still 77.27% correct answers. Three examples improved massively with over 30% more correct answers: L1 with 86.36%, L6 with 81.82%, and L7 with 86.36%. The L5 example even deteriorated slightly with -9.09%. While the legitimate examples performed way better in the pre-test compared to the phishing exampels (42,6% to 75%), this gap has now almost evened out with phishing = 86.93% and legitimate = 88.07%.

Comparing post-test to retention-test, a mixed picture for the changes emerged similar to the phishing examples. Examples like L7 showed no change. Other examples even increased slightly like L2 with +4.55%. However, most examples decreased. Two examples in particular stand out here, losing over 10%. L6 lost 22.73% in correct answers and L8 lost even 13.64%. Thus, L6 lost over 66% of the number of correct answers previously gained by watching the video and again represented the worst example with only 59.09%. It was also interesting that L5 achieved a worse

**Table 4.6:** Detection rate in % for individual legitimate URLs for the two phases. In contrast to the phishing examples, there is no phishing trick included in the legitimate examples, so the column is not present.

| Name | Pre | Post | Retention (eight weeks) |
|---|---|---|---|
| L1 | 54.55 | 86.36 | 81.82 |
| L2 | 77.27 | 77.27 | 81.82 |
| L3 | 90.91 | 95.45 | 95.45 |
| L4 | 100 | 100 | 100 |
| L5 | 90.91 | 81.82 | 77.27 |
| L6 | 50.00 | 81.82 | 59.09 |
| L7 | 54.55 | 86.36 | 86.36 |
| L8 | 81.82 | 95.45 | 81.82 |

value than in the pre-test with only 77.27% correct answers. Two examples still achieved a score over 95% with L4 (100%) and L3 (95.45%). While most others remained around 80% of correct answers, ranging from 77.27% (L5) to 86.36% (L7).

## 4.2 E-Learning

In the past, many phishing awareness measures were passive or conditionally interactive learning. For example, videos were shown or content was conveyed in text form. However, the user receives less feedback about his or her own level of awareness and performance. There are also fewer opportunities to apply what they have learned directly and thus deepen their awareness. Therefore, an e-learning was developed, which is based on the findings of the NoPhish concept. The training was adapted to the university environment for a study. Important components of the training are sections with detailed information on possible facets in the area of phishing in combination with small tasks after each section, which are intended to deepen and test what has just been learned. This also gives users the opportunity for self-reflection on their own level of learning. Each exercise had to be successfully completed before moving on to the next section.

The overarching goal of this section is therefore to test whether such extensive training can effectively increase the ability to distinguish between legitimate and phishing messages over an even longer period of time compared to a very short measure such as a video and similar to those of the NoPhish Android game [33]. To this end, the e-learning was evaluated in an online study with participants from the university environment. Participants had to complete a test at three measurement points (pre-test, post-test, and retention-test). The pre-test had to be completed immediately before taking the e-learning and the post-test immediately after taking the e-learning. The retention-test then took place after five-months. The results of the study can be summarized as follows: 1) the training significantly increased the ability to distinguish between phish and legitimate messages in the pre-post comparison (prerequisite for longer-term observation) and 2) even after an interval of five months there is still a significantly better ability compared to the pre-test.

**Parts of the results described in this chapter have been published in:**

- Berens, B., Dimitrova, K., Mossano, M., & Volkamer, M. (2022). Phishing awareness and education-When to best remind?. In Workshop on Usable Security and Privacy (USEC).

### 4.2.1 Research Questions and Hypotheses

In the following, the concrete research questions and the related hypotheses is described. The general idea is to evaluate the effectiveness in terms of the sensitivity for the e-learning directly after the measure and in the long-term as well as to analyze the different effects for the specific phishing tricks. In previous research already the game that was based on NoPhish concept got a significant better phishing detection after five months. In Study 1 (see Section 3.3.1), a way shorter measure also achieved significant better results for sensitivity around eight weeks after the measure.

The question now is whether, especially with regard to sensitivity, a longer measure such as the e-learning over a period of five months can also achieve a significant improvement for sensitivity. Also, compared to study 1, more different examples should be used to cover an even wider range of manifestations for the phishing tricks. Based on these open questions, the first research question is:

**RQ**$_1$: *Is there a long-term effect of the e-learning as a phishing awareness measure?*

This leads to the following hypothesis or prerequisite:

**H**$_1$: Participants show significantly higher sensitivity immediately after e-learning compared to the pre-test.

Without this prerequisite, it is not meaningful to conduct a long-term evaluation. If there is no significant improvement directly after the e-learning, long-term effects (should they be measured) must necessarily be due to other external factors. Subsequently, the following main hypothesis needs to be tested:

**$H_2$**: Participants show significantly higher sensitivity five months after the e-learning compared to the pre-test.

In addition to this basic effectiveness in increasing sensitivity, there is also the question of whether the measure has a different effect on the different phishing tricks over time. Already after eight weeks of the video measure (see Section 4.1.3.2) different increases and decreases have been observed. It would now be interesting to know whether these trends for the various phishing tricks are repeated in another measure and whether they continue, level off or intensify over a longer period of time. So the second research question is as following:

**$RQ_2$**: *Does the long-term effect of the e-learning as a phishing awareness measure differ for specific phishing tricks?*

## 4.2.2 Methodology

This subsection describes the procedure for evaluating the above hypotheses. The focus is initially on the procedure of the study. The study design is divided into different phases and the content of each phase is described. Subsequently, the materials used for the evaluation of the sensitivity is described in detail (see Section 4.2.2.2). In there, the focus is on the creation of the email screenshots for the three tests. Finally, the ethical and privacy considerations that were necessary for the study were discussed (see Section 4.2.2.3).

### 4.2.2.1 Study Design

The study took the form of a repeated measures design. Also due to the restrictions imposed by the COVID-19 pandemic, an online study was chosen. Participants were contacted by email and received links to participate.

The study is roughly divided into two phases. The first phase consists of the invitation email with a lot of important information about the e-learning and duration of the study, as well as instructions on how to proceed if they agree to participate. The email also contained the link for the pre-test. After clicking on the pre-test link, participants were directed to the Sosci Survey platform. This platform was used for all three tests. Only the e-learning took place on another platform, the university's ILIAS system. On the first page of the pre-test, participants were again given a consent form explaining the procedure and other important aspects of the study (for the full study design see Figure 4.3). Next, participants were asked to generate a subject code (more on the subject code in Section 4.2.2.3).

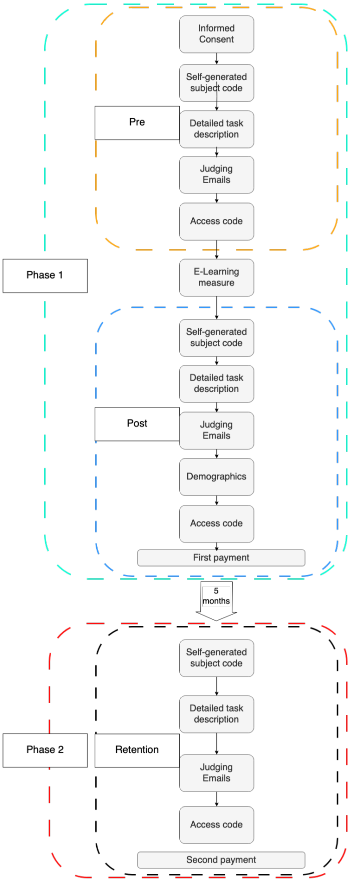


**Figure 4.3:** Study design for the three time points (pre, post and retention).

“Dear participants, since we want to anonymously assign your questionnaire, it is important that you generate your personal code. Because only this way your questionnaires can be assigned to each other without anyone being able to find out who filled out these questionnaires. So it is important that you still know the same code when you are asked next time. For this reason, we have formulated the following questions to help you remember your personal combination. Please name the first and last letter of your mother’s first name (e.g. Anne = AE) Please name the first and last letter of your father’s first name (e.g. Thorsten = TN) Please name the first and last letter of your first name (e.g. Hannah = HH) Please name your mother's birthday (e.g. July 17, 1950 = 17)" [7]

This was followed by the description of the scenario. The scenario was necessary so that the examples or more precisely their organization would not be classified as phishing due to lack of familiarity or personal preference. This should ensure that all participants have at least a roughly similar understanding of how to evaluate the examples and do not use factors other than the desired ones for evaluation. The scenario was as follows:

> "To determine how well you can distinguish messages with dangerous content from real messages, the following are examples with messages in different contexts. Knowing that you do not know or use all senders, service providers, operating systems and programs, we present received messages for different senders, service providers, operating systems and programs. In order not to directly declare implausible the senders, service providers, operating systems and programs to which you have no connection in reality, please assume in the following that... you are Martin Müller. you speak the languages German and English. your work colleague is Jonas Schmidt. you use all services that are used in this questionnaire. you use the different operating systems (e.g. Microsoft Windows, Apple OSX, Google Android, Apple iOS) and applications (Thunderbird, Apple Mail, Google Mobile Mail, GMX Mobil Client) used in this questionnaire."

After the baseline scenario had been presented, the participants were shown a series of screenshots. These screenshots showed an email from an organization. The order of the screenshots was completely random for each participant. More details on the structure, the process of creating the screenshots, and an example can be found in Section 4.2.2.2. Each screenshot was shown one at a time on a single page. Participants then had the task of deciding whether the screenshot showed a phishing or a legitimate email. In total, participants saw 34 different screenshots. The participants did not have the opportunity to change their decision after they had made a decision and moved on to the next screenshot. After all decisions were made, participants were given an access code for the e-learning. The code was sent to the investigators and intended to ensure that only participants who actually completed the pre-test would continue to participate in the study. In doing so, participants were kept unaware that the code was the same for all study participants. This was to ensure that there was no way for the investigator to link the data to any study participant.

Afterwards, the participants received a link to the e-learning from the investigator. The participants were informed that after completing the e-learning they would need the automatically issued certificate as proof of successful completion. The e-learning is the publicly accessible version [1]. In this version, the test administration did not have access to the environment and thus no linkage of data to test participants was possible.

The post test started after the participants sent the necessary certificate to the test management. Subsequently, participants again received a link to the Sosci Survey platform for the post-test. The test had many of the same components as the pre-test. This time, there was no consent form, but participants again started generating their own subject code. Then the participants were again presented with the same scenario as in the pre-test. Now the actual test followed again with the same 34 examples, whereby these were presented in a new random order. Participants were then asked a few more socio-demographic questions (age and gender) and were again given an access code. The code then had to be sent back to the experimental administration for confirmation of the successful post-test. Again, the code was the same for all participants, but different from the code from the pre-test.

The retention test then started five months later, and all participants who had completed all phases by then were again contacted for participation. If participants agreed to participate for the retention phase, they were sent the link to the platform Sosci Survey. The retention-test was also similar in structure to the pre-test. It started with the generation of the subject code to match all data of the two phases. Subsequently, the study scenario was presented a third time. Now the participants had to take part again in the test with the same 34 examples. Additionally, the examples were presented in a random order. At the end of the study, participants then received the final access code to receive the final payout.

#### 4.2.2.2 Study Material

This subsection deals with the study material used to measure sensitivity in the pre-test, post-test and retention-test. These are screenshot examples that are based on a collection of emails that were compiled in advance from various organizations. In particular, examples were collected from large, well-known organizations, e.g. Lufthansa, PayPal or DHL. All emails were written in German. Previous research has shown that the language of the email has an influence on the evaluation of this [93]. However, this influence should not be considered in this study.

The e-mails used as examples have been sent in this form by the relevant organizations in the past. This was to ensure that no “errors” in the creation of artificial examples would have an influence on the participants’ decision. Both the desktop and a mobile context were selected as the environment for the emails. For the desktop context, the URL was presented in the form of

[1] `https://secuso.aifb.kit.edu/betr-nachrichten-schulung-buerger`

both a tooltip and a status bar. An equal number of phishing and legitimate examples were chosen for the study. By creating the examples for the study, these authentic emails were taken and then assigned to either the legitimate email category or to a phishing trick. If the example was assigned to a phishing trick, the necessary adjustments were made. This was to ensure that only this change would be decisive for the actual decision. Then either the sender address, the URL or the format of the attachment was adjusted. The adjustments should cover as wide a spectrum of e-learning as possible. It must be mentioned that not all sections of the e-learning were suitable for this form of evaluation. For example, the basic introduction to the topic (Chapter 1) or the request to make bank transfers (Chapter 3), could not be meaningfully covered by the methodology of testing with examples and the question whether an example is phishing or legitimate. The other considerations is discussed in more detail in the Section 4.2.4 of the limitations.

Finally, as described in the Section 4.2.2.1, 34 examples were created based on the phishing tricks as in Section 3.1. Accordingly, the following tricks were selected: 1) Content - simple examples where the sender or the text content was changed, 2) Non-Brand related Domain - where the link URL was changed - with a non-sender domain (random domain), 3) Non-brand related Domain + Brand Outside - where the link URL was changed with a non-sender domain and the sender in the path or subdomain 3) Small Deviations in the Domain - where the link URL was changed - small changes in the domain like typo or letter swap, 4) Special Link Manipulation - where the link URL was changed and the email was changed - mismatch between link text and link URL or a wrong tooltip, and 5) Attachment - where the file format was changed to a potential malicious. For the detailed list of all examples in phishing Tricks, URL of the link and chapter in e-learning see Table 4.7 and Table 4.8.

#### 4.2.2.3 Recruitment & Ethical / Data Protection Considerations

The participants were recruited from students at the Karlsruhe Institute of Technology. In this process, students were primarily recruited through a snowball system via three different routes: Posts on social networks (Twitter and Facebook), leaflets in various places, e.g. on campus or student dormitories, and by personal invitation. Prestudies recorded an average of 108 minutes for the measure, plus about 15 to 20 minutes per test. This resulted in a total study duration of approximately 180 minutes. All participants who successfully completed at least the first two tests received 20€. If the participant has also successfully completed the third phase, the amount increases to a total of €40. The calculation of the payment was based on the German minimum wage of 9.50€ per hour at that time.

The Sosci Survey platform was used for data collection [94]. As a platform, it meets the requirements regarding the European Data Protection Regulation (GDPR).

**Table 4.7:** Overview of the phishing URLs for the e-learning study. "TT" stands for tooltip and "ST" stands for status bar.

| Name | Phishing Trick | More specific Phish | Sender | Subject | TT or ST | Indicator |
|---|---|---|---|---|---|---|
| P1 | Content | Sender | `amazon@host547.ru` | Your Amazon.de order with "Magnets, assorted colors..." has been shipped! | | |
| P2 | Content | Nonsense | Vodafone Team <nichtantworten@kundenservice.vodafone.com> | Confirmation of your cancellation | | |
| P3 | Non-brand related Domain | Random URL | Lufthansa <newsletter@lufthansa.com> | From March 2019 we will introduce a new system for awarding award miles | Status | `https://www.dtrdtcbj.com/de/en#blade` |
| P4 | Non-brand related Domain | Random URL | Lufthansa <newsletter@lufthansa.de> | From March 2019 we will introduce a new system for awarding bonus miles ein | Tooltip | `https://www.hisoliajo.host547.com/web3/HoEv/ksokGkd=ad3/kol45G` |
| P5 | Non-brand related Domain + Brand Outside | Subdomain | `versandbestaetigung@amazon.de` | Your Amazon.de order with "Magnets, assorted colors...." has been shipped! | Status | `https://amazon.de.kolwerg.com/596ksokGkd89=adweb3/HoEv` |
| P6 | Non-brand related Domain + Brand Outside | Subdomain | `versandbestaetigung@amazon.de` | Your Amazon.de order with "Magnets, assorted colors...." has been shipped! | Tooltip | `https://amazon.de.kolwerg.com/596ksokGkd89=adweb3/HoEv` |
| P7 | Non-brand related Domain + Brand Outside | Path status | Vodafone Team <nichtantworten@kundenservice.vodafone.com> | Confirmation of your cancellation | Status | `https://www.qpglljhotg.com/vodafone.com.596ksokGkd89=adweb3/HoEv` |
| P8 | Non-brand related Domain + Brand Outside | Path tool | `paypal@mail.paypal.de` | PayPal account overview for August | Tooltip | `https://www.qpglljhjotqgg.com/paypal.de` |
| P9 | Small Deviations in the Domain | Typo status | Lufthansa <newsletter@lufthansa.de> | From March 2019 we will introduce a new system for awarding bonus miles | Status | `https://www.luftthansa.com/de/newsletter-information` |
| P10 | Small Deviations in the Domain | typo toolt | Vodafone Team <nichtantworten@kundenservice.vodafone.com> | Confirmation of your termination | Tooltip | `https://vodafon.de/fwlink/?LinkID=462932` |
| P11 | Small Deviations in the Domain | swap dialog | Jonas Schmidt | Job offer farmers market | Dialog | `https://www.baurenmarkt.de/job-angebot/1eqds321h1w34UHA` |
| P12 | Special Link Manipulation | Mismatch status | `paypal@mail.paypal.de` | PayPal account overview for August | Status | `https://www.hisoliajo.host547.com/web3/HoEv/596ksokFkd89=ad3/kol45G5Hwerg?32` |
| P13 | Special Link Manipulation | mismatch tool | DHL Paket <info@dhl.de> | New password for dhl.de | Tooltip | `https://www.host547.com/verify/\T1\sskey=kw2RtU_5dsh` |
| P14 | Special Link Manipulation | mismatch dialog | Jonas Schmidt | Notes | Dialog | `https://husjukuila-torgibut/com/join` |
| P15 | Special Link Manipulation | fake tooltip | `paypal@mail.paypal.de` | PayPal account overview for August | Tooltip | `https://www.hisoliajio.host547.com/web3/HoEv/596ksokFkd89=ad3/kol45G5Hwerg?32` |
| P16 | Attachment | | Vodafone Team <nichtantworten@kundenservice.vodafone.com> | Confirmation of your cancellation | | Rechnung.exe |
| P17 | Attachment | | DHL Paket <info@dhl.de> | Convenient parcel receipt with the DHL parcel box | | Informationen.pdf.exe |

**Table 4.8:** Overview of the legitimate URLs for the e-learning study. "TT" stands for tooltip and "ST" stands for status bar.

| Name | Sender | Subject | TT or ST | Indicator |
|---|---|---|---|---|
| L1 | Amazon.de <versandbestaetigung@amazon.de> | Your Amazon.de order with "Magnets, assorted colors..." has been shipped! | None | Amazon |
| L2 | Vodafone Team <nichtantworten@kundenservice.vodafone.com> | Confirmation of your cancellation | None | Vodafone |
| L3 | Lufthansa <newsletter@lufthansa.com> | From March 2019 we will introduce a new system for awarding award miles | Tooltip | https://www.lufthansa.com/de/miles-and-more/meilenvergabe |
| L4 | paypal@mail.paypal.de | PayPal account overview for August | Statusleiste | https://www.paypal.de/webapps/mpp/aq?utdled-notificamai&s=ci&mail=sys |
| L5 | DHL Paket <info@dhl.de> | New password for dhl.de | Tooltip | https://dhl.de/account |
| L6 | paypal@mail.paypal.de | PayPal account overview for August | Statusleiste | https://www.paypal.de/konto-uebersicht.php |
| L7 | paypal@mail.paypal.de | PayPal account overview for August | Statusleiste | https://www.paypal.de/konto-uerbersicht |
| L8 | versandbestaetigung@amazon.de | Your Amazon.de order with "Magnets, assorted colors..." has been shipped! | Statusleiste | https://amazon.de/596ksokGkd89=adweb3/HoEv |
| L9 | versandbestaetigung@amazon.de | Your Amazon.de order with "Magnets, assorted colors..." has been shipped! | Tooltip | https://amazon.de/596ksokGkd89=adweb3/HoEv |
| L10 | Lufthansa <newsletter@lufthansa.com> | From March 2019 we will introduce a new system for awarding award miles | Tooltip | https://www.lufthansa.com/meinlufthansa/service |
| L11 | Vodafone Team <nichtantworten@kundenservice.vodafone.com> | Confirmation of your cancellation | Statusleiste | https://www.vodafone.com/meinvodafone/services/Ihre-rechnungen |
| L12 | Lufthansa <newsletter@lufthansa.com> | From March 2019 we will introduce a new system for awarding award miles | Statusleiste | https://www.lufthansa.com/de/en/newsletter-information |
| L13 | DHL Paket <info@dhl.de> | New password for dhl.de | Tooltip | https://dhl.de/go/12/3ETWV5HW-2PSB2P6G-2RFZRFKW-17YA1VS-o.html |
| L14 | Vodafone Team <nichtantworten@kundenservice.vodafone.com> | Confirmation of your cancellation | None | Rechnung.pdf |
| L15 | DHL Paket <info@dhl.de> | Convenient parcel receipt with the DHL parcel box | None | Informationen.pdf |
| L16 | J Schmidt | Gifts | Dialog | https://www.amazon.de/dp/B00Li0KY52/ref=sr_1_4?keywords=sohn+kriminalroman&qid=1581586169&sr=8 |
| L17 | Jonas Schmidt | Video KickOff | Dialog | https://www.youtube.com/watch?VGhijvA4XTU&list=RDA44i5rzysu |

Regarding the ethical requirements, the necessary points of the university were met. On the one hand, the participants should be informed about the objectives of the study, which was ensured via the invitation email as well as via the start page of the pre-test. They were also informed about their rights and possibilities to terminate the study at any time without negative consequences (without giving any reasons). They did not have to actively object to this, but simply stop the study at a certain point, e.g. by no longer forwarding the access codes to us. In addition, participants had to explicitly agree to this in order to participate beyond the first page of the pre-test. As mentioned in Section 4.2.2.1, a subject code system was used to ensure the pseudonymity of the participants. In each phase, participants had to generate a subject code that only they could know, based on a set of instructions. This code was based on information provided by the participants, and with the same instructions, this code should always lead to the same result. The instructions for generating the code were as follows:

> Dear participant, since we want to assign your questionnaire anonymously, it is important that you generate your personal code. This is the only way to match the questionnaires to each other without anyone being able to find out who filled out these questionnaires. So it is important that you still know the same code when you are asked next time. For this reason, we have formulated the following questions to help you remember your personal combination. Please name the first and last letter of your mother's first name (e.g. Anne = AE). Please name the first and last letter of your father's first name (e.g. Thorsten = TN) Please name the first and last letter of your first name (e.g. Hannah = HH) Please name your mother's birthday (e.g. July 17, 1950 = 17)
>
> The code presented here as an example would thus be AETNMA17.

### 4.2.3 Results

Of the original 46 participants in the study, 20 completed questionnaires were collected for the retention-test, i.e. these participants both completed the retention-test and sent in the access code. Unfortunately, the method of maintaining pseudonymity with the subject codes carries a risk that participants may make errors in code generation despite the same instructions, and therefore the codes may read differently at different phases. In the case of the study, only 16 of 20 participants could be clearly assigned all three tests and thus evaluated for further analysis. Of these, nine were male and seven female, and all were 24.69 years old on average. Before moving to the sections with the hypothesis tests, the descriptive statistics of the average detection rates for phishing, legitimate and all examples as well as sensitivity and criterion for all three tests are reported (see Table 4.9).

**Table 4.9:** Overview of the average rates for all three points in time in % for the correct answers for phishing and legitimate examples as well as overall. Also adding the values for the signal detection theory.

| Example Type | Pre | Post | Retention |
|---|---|---|---|
| Phishing | 63.24 | 90.44 | 84.93 |
| Legitimate | 80.02 | 91.18 | 93.75 |
| Overall | 71.63 | 90.81 | 89.34 |
| Values Signal Detection Theory | | | |
| Sensitivity | 1.28 | 2.66 | 2.47 |
| Criterion | 0.24 | 0.02 | 0.25 |

Following the repeated measures ANOVA [95], paired t-tests were performed as part of the post-hoc tests to compare the measurement time points against each other. This was done conservatively to counteract alpha error accumulation. Therefore, the Bonferroni correction was applied to the multiple testing.

For the calculation of the effect size, the measure $\eta_p^2$ was used. The measure has the advantage that the effect size from different studies can be compared with each other [96] and is therefore particularly suitable for looking at the results from multiple studies. Effect sizes are interpreted here according to Ellis [97] as follows: .01 = small effect, .06 = medium effect, and .14 = large effect.

#### 4.2.3.1 RQ1 - Long-term Effect

The repeated measures ANOVA showed a significant difference between pre-test, post-test, and retention-test for sensitivity with F(2,30) = 31.037, $p < .0001$, $\eta_p^2$ = .674. Following Ellis' interpretation, this is a large effect. For the study, sensitivity is evaluated, ranging from +3.38 to -3.38. Paired t-tests for the comparisons between the three measurement time points showed that both pre-test and post-test ($p < .0001$), and pre-test and retention-test ($p < .0001$) were significantly different. In contrast, the p value for the comparison between post-test and retention-test was .292. Combining this with the descriptive figures, the pre-test e-learning sensitivity was low with ($d'$ = 1.28, SD = .935). For the post-test, the sensitivity then increased ($d'$ = 2.66, SD = .76) and showed a slight decrease until the retention-test ($d'$ = 2.47, SD = .422).

Thus, H1 “Participants showed significantly higher sensitivity immediately after the e-learning compared to the pre-test” can be accepted. The same was true for the main hypothesis H2

"Participants show significantly higher sensitivity five months after the e-learning compared to the pre-test".

Now, the Signal Detection Theory (see Figure 3.13) gives another measure that, in conjunction with sensitivity, can make an important statement about the ability of the participants. The criterion describes the tendency to select either phishing or legitimate examples.

The repeated measures ANOVA showed a significant difference between pre-test, post-test, and retention for the criterion with F(2,30) = 3.709, $p = .037$, $\eta_p^2$ = .101. Following Ellis' interpretation, this was a large effect. Post-hoc analysis for the comparison between the three measurement time points showed that there is no significant difference between any of the group comparisons. The comparison between $C_{Pre}$ and $C_{Post}$ was not significant with $p = .133$, same for $C_{Pre}$ and $C_{Retention}$ with $p = .959$ and $C_{Post}$ and $C_{Retention}$ with $p = .27$ (see Figure 4.4 for a boxplot of the three groups with post-hoc comparison).

When looking at the descriptive values for the three points in time, it was noticeable that the participants initially achieved a $C = .241$. this means that they had a tendency to classify examples as legitimate (regardless of whether it is legitimate or phishing). After the e-learning, this value changed in a better direction, moving towards neutral 0 with $C = .021$. Then, at the time of the retention-test, a tendency towards legitimate classifications could be seen again with $C = .246$. Thus, the participants became again a bit "less cautious" after five months.

With regard to the criterion, it can now be stated that no significant deterioration occurred as a result of the measure i.e. people neither became overcautious (tendency towards phishing) nor overly uncautious (tendency towards legitimate). But in the sample there seemed to be a tendency to make rather uncautious decisions, which could only be neutralized in the short term by the e-learning.

#### 4.2.3.2 RQ2 - Phishing Tricks

This subsection focuses on the differentiated evaluation of the individual examples from Section 4.2.2.2. For this purpose, both the phishing tricks as a combination of several examples and the individual examples were analyzed for special trends. This should form the basis for subsequent recommendations for the further development of the measure or for specifying the focus for refreshing awareness.

**Phishing Tricks** By looking at the phishing tricks (all details in Table 4.10), it was noticeable that both "Small Deviations in the Domain" (45.83%) and "Attachment" (46.88%) performed by far the worst in the pre-test. They were even just below the guess probability of 50%. In contrast, the "Special Link Manipulation" (64.06%), "Non-brand related Domain" (65.63%) and

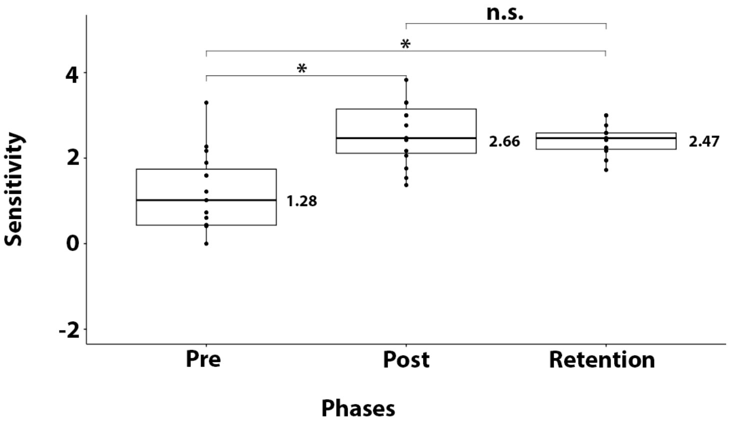


**Figure 4.4:** Boxplot for sensitivity for all three points in time. The mean sensitivity is on the right side of the boxplots. "*" marking those post-hoc comparison that are significant ($p < .05$ and "n.s." those that are not significant.

the “Non-brand related Domain + Brand Outside” (73.44%) were in the midfield. At the top were the “Content” with 81.25%.

Within the post-test rankings, there were a few changes. The “Small Deviations in the Domain” improved by 25%, but was still at the bottom of all tricks. All other tricks now reached values above 90% with “Content” (90.63%), “Attachment” (93.75%), “Special Link Manipulation” (93.75%), “Non-brand related Domain + Brand Outside” (95.31%) and “Non-Brand related Domain” (100%). In the retention-test there were now again very different developments e.g. some tricks remained almost at a similar level like “Non-Brand related Domain” (100%), “Special Link Manipulation” (96.88%) or “Non-brand related Domain + Brand Outside” (98.44%). Others dropped slightly like “Content” (84.38% with -6.25%) or “Attachment” (84.38% with -9.37%). And again “Small Deviations in the Domain” (51.67% with -19.16%) brought up the rear of all tricks.

**Phishing Examples** Going more into the details and looking at the individual examples of phishing tricks (see Table 4.11), some tricks achieved very similar results within the corresponding examples, while others achieved very different results. The “Content” examples often got similar values in the very high range of 75% and more. For the “Non-Brand related Domain” examples, it was noticeable that initially P3 achieved only 50% in the pre-test compared to 81.25% for P4. From the post-test on, however, both examples always reached 100%. In the examples for “Non-brand related Domain + Brand Outside” the two subdomain (P5 and P6) examples performed much worse than the path examples with only 62.5%. But also in the post-test with more than 93.75% correct answers the majority of the examples were recognized correctly. The examples

**Table 4.10:** Percentages of correct answers per phishing trick for all two phases.

| Phishing Trick | Pre | Post | Retention (5 months) |
|---|---|---|---|
| Content | 81.25 | 90.63 | 84.38 |
| Non-brand related Domain | 65.63 | 100 | 100 |
| Non-brand related Domain + Brand Outside | 73.44 | 95.31 | 98.44 |
| Small Deviations in the Domain | 45.83 | 70.83 | 41.67 |
| Special Link Manipulation | 64.06 | 93.75 | 96.88 |
| Attachment | 46.88 | 93.75 | 84.38 |

for “Small Deviations in the Domain” also showed a very different and volatile picture. First, P9 scored well below average at 18.75% compared to P10 (56.25%) and P11 (62.5%), although the other two did not achieve particularly good scores either. In the post-test, all examples improved, although P9 with 37.5% (+18.75%) was still worse than the other examples (P10 = 81.25% with +25% and P11 = 93.75% with +31.25%). For the retention-test, all examples dropped again (partly extremely). P9 reached only 12.5% (-25%), P10 only 37.5% (-43.75%) and P11 75% (-18.75%). Both P9 and P10 thus dropped to a level below the initial level. The “Special Link Manipulation” examples, on the other hand, showed a similar picture as the other tricks. The examples started at a very different level with P12 (56.25%), P13 (81.25%), P14 (87.5%) and P15 (31.25%). However, as of the post-test, all examples reached values from 81.25% to even 100%. The “Attachment” examples also started near or below the guess probability ((P16 with 56.25% and P17 with 37.5). Then both rose to 93.75% in the post-test and dropped slightly to the retention-test (P16 = 81.25% and P17 = 87.5%).

**Legitimate Examples** First of all, it was noticeable that only two examples achieved 100% correct answers in the pre-test (see Table 4.12). This was the first Amazon example with the status bar and the Youtube example with a dialog. Most of the examples were in the range between 80% and 90%. But there were also a few outliers downwards, e.g. the DHL example with the long URL that was only answered correctly by 50% (L13). However, also an example where the participants saw an email with a PDF with further information about the receipt of parcels from DHL was only correctly recognized in 67% (L15) or also the example with status bar for the account overview only achieved a value of 69% (L6) in the pre-test. Looking at the post-test results, most of the examples improved, apart from the two examples that had already reached 100% and an example from Lufthansa with a tooltip for awarding miles, which even dropped by 13% (L3). The two examples regarding the content and sender did not change either and remained at 81% (L1 and L2). The other twelve examples improved by 17% in the range from 6% to 25%. Thus, in the

**Table 4.11:** Percentages of correct answers per phishing email with phishing trick information for all two phases.

| Name | Phishing Trick | Pre | Post | Retention (5 months) |
|---|---|---|---|---|
| P1 | Content | 75 | 100 | 93.75 |
| P2 | Content | 87.5 | 81.25 | 75 |
| P3 | Non-brand related Domain | 50 | 100 | 100 |
| P4 | Non-brand related Domain | 81.25 | 100 | 100 |
| P5 | Non-brand related Domain + Brand Outside | 62.5 | 93.75 | 100 |
| P6 | Non-brand related Domain + Brand Outside | 62.5 | 93.75 | 100 |
| P7 | Non-brand related Domain + Brand Outside | 75 | 100 | 93.75 |
| P8 | Non-brand related Domain + Brand Outside | 93.75 | 93.75 | 100 |
| P9 | Small Deviations in the Domain | 18.75 | 37.5 | 12.5 |
| P10 | Small Deviations in the Domain | 56.25 | 81.25 | 37.5 |
| P11 | Small Deviations in the Domain | 62.5 | 93.75 | 75 |
| P12 | Special Link Manipulation | 56.25 | 100 | 100 |
| P13 | Special Link Manipulation | 81.25 | 93.75 | 100 |
| P14 | Special Link Manipulation | 87.5 | 100 | 100 |
| P15 | Special Link Manipulation | 31.25 | 81.25 | 87.5 |
| P16 | Attachment | 37.5 | 93.75 | 81.25 |
| P17 | Attachment | 56.25 | 93.75 | 87.5 |

post-test, all examples achieved at least 75% correct answers, ten of them over 90% and six even 100%.

### 4.2.4 Limitations

Not all phishing tricks from the e-learning were used for the evaluation [2]. URLs that can only be correctly detected with a tool because they contain a redirect or use a short URL were not applicable for the study context. With the use of such a tool, these examples would then again look similar to the examples used and would thus not provide any added value. Without the (fictitious) use of such a tool, participants would have no chance to make an informed decision. Such a survey would only make sense in the context of qualitative research if the participants could be asked about their perception of the URL. However, this was not the research question

[2] These are only the study specific limitations, those related to all three studies are discussed at the end of the chapter

**Table 4.12:** Percentages of correct answers per legitimate email for all two phases. In contrast to the phishing examples, there is no phishing trick included in the legitimate examples, so the column is not present.

| Name | Pre | Post | Retention (5 months) |
|---|---|---|---|
| L1 | 81.25 | 81.25 | 93.75 |
| L2 | 81.25 | 81.25 | 100 |
| L3 | 87.5 | 75 | 93.75 |
| L4 | 81.25 | 100 | 81.25 |
| L5 | 75 | 93.75 | 100 |
| L6 | 68.75 | 87.5 | 81.25 |
| L7 | 87.5 | 100 | 100 |
| L8 | 100 | 100 | 93.75 |
| L9 | 75 | 93.75 | 100 |
| L10 | 81.25 | 100 | 100 |
| L11 | 75 | 100 | 100 |
| L12 | 81.25 | 87.5 | 100 |
| L13 | 50 | 75 | 81.25 |
| L14 | 87.5 | 93.75 | 87.5 |
| L15 | 66.67 | 87.5 | 81.25 |
| L16 | 81.25 | 93.75 | 100 |
| L17 | 100 | 100 | 100 |

of the study. The same applies to URLs that can only be recognized if one knows the structure of the actual domain (e.g. kit-shop.de). In the context of the study, these would only have been possible with a different form of sample survey if the participants had been screened in advance with regard to their prior awareness and then only these would have been eligible for participation. Due to the long duration of the study and further limitations due to the Covid-19 pandemic, the sample was already limited and would not have allowed such further reduction.

## 4.3 Workshop

It is extremely complicated for organizations to ensure an adequate level of awareness in the area of IT security. Even more complicated is the question of how long the trained awareness will last. In many cases, organization follow regularities, e.g. a measure is carried out every twelve months.

However, it is usually not known whether the awareness measures will even last 12 months or whether a refresher would be necessary in the meantime.

In the case that a measure does not last for the full twelve months, the question naturally arises how to deal with this situation. Does one rely on the fact that one somehow “achieved” the twelve months or does one plan another measure. And if one decides for a further measure, how does this measure look then. Does one choose the same measure again or does one choose something new. Of course, it is obvious (depending on the initial situation) to first choose a comprehensive measure, which provides the fundamental basis for the importance of the topic and helpful tips and tricks. If one has initially decided on an extensive measure that has consumed a lot of time and resources, the question arises as to whether these can be spent again. Also a “simple” repetition can lead to the fact that coworkers are less attentive, since they already know the measure. A new measure, which possibly conveys similar or even the same contents, but in the form of a different medium, can bind the attention by its novelty alone. Additionally, this can also possibly take up the most important points again in a condensed form without having to start from scratch. In the past, single measures or different measures were often tested against each other. However, a sequence of measures to find a long-term potential synergy effect has not been considered so far. Here the question arises, which kind of measure is suitable to serve especially as a refresher for a more detailed training.

In order to increase awareness in a larger corporate context, training courses are often an option. In the case of training courses, these were often conducted by teams that were also involved in creating the training. Recruiting experts to carry out such training is a circumstance that is not applicable to every organization, as this is often only possible as part of a study. The question therefore arises as to how such an existing measure can be made available to an organization so that it can carry out the process of implementation itself. And what influence does it have on the effectiveness of the measure if the organization takes on the implementation and execution itself?

**Parts of the results described in this chapter have been published in:**

- Reinheimer, B., Aldag, L., Mayer, P., Mossano, M., Duezguen, R., Lofthouse, B., Von Landesberger, T. and Volkamer, M. (2020). An investigation of phishing awareness and education over time: when and how to best remind users. In Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020) (pp. 259-284).

### 4.3.1 Research Questions and Hypotheses

To answer the questions raised so far, the following research questions were formulated with corresponding hypotheses.

**RQ**$_1$: *How long does the effect of the workshop last?*

First of all, the following prerequisite must apply for the evaluation to be meaningful over a longer period of time:

**H**$_1$: The workshop leads to a significantly improved sensitivity of the participants directly afterwards.

Previous research from the field of phishing awareness combined with sensitivity has still found significant values in the range from eight weeks [31] to five months [33]. Therefore, a conservative assumption was made that the bi-monthly measurements should start at four months after the workshop. Due to the limited sample size in the organization, a compromise had to be made here so that there would still be enough participants for the later surveys and the refresher groups.

Now the effectiveness of the workshop is to be evaluated over a longer period of time. To this end, the following hypothesis was formulated: **H**$_2$: The sensitivity of the participants is higher after $\Delta t = 4 + 2i$ months, where $i \in \{0, 1, 2, 3, 4\}$ than before the workshop.

As soon as the sensitivity is no longer significantly better compared to before workshop, a refresher should be conducted. For this purpose, the refresher measures should be distributed as soon as **H**$_2$ no longer applies. The second research question was:

**RQ**$_2$: *Which type of refresher would result in a significantly better sensitivity in the long term?*

First of all, the following prerequisite must apply for the evaluation of refresher to be meaningful over a longer period of time:

**H**$_3$: Participants' sensitivity will be significantly higher immediately after the refresher than before the workshop (for long text, video, interactive email example, and short text, respectively).

For all refresher groups for **H**$_3$ is valid, it should then be checked how long this refresher lasts. Here, the particular focus was on the period of twelve months. Thus, the question was whether any of the refreshers could maintain sensitivity significantly above the pre-workshop result over a further six-month period. To this end, the following hypothesis was formulated:

**H**$_4$: Participants' sensitivity will be significantly higher six months after the refresher than before the workshop (for long text, video, interactive email example, and short text, respectively).

For this hypothesis, there were again two important conditions that needed to be controlled. Firstly, this sample was tested twice (directly after the refresher and six months later) compared to the others groups. Accordingly, it should be excluded that the renewed participation has an influence on the sensitivity recording. Secondly, no parallel events during the study should have an influence on the entire sample of employees in the organization and thus for the recording after twelve months.

First, both the actual workshop and the refreshers were examined over a total period of twelve months, checking which measure showed a significant improvement over the pre-test over the entire period. Since the detection of sensitivity is based on different phishing tricks, the question arises whether these phishing examples develop differently over the twelve-month period and whether the refreshers have a different impact immediately afterwards as well as over a period of six months.

**RQ**$_3$: *How long do the effects of the workshop and the refreshers for specific phishing tricks last?*

## 4.3.2 Methodology

In this section, the basic methodology of the study is described in detail. First, the study design with its flow chart and the necessary considerations is explained (see Section 4.3.2.1). Then, the materials (see Section 4.3.2.2) used for the study are presented with a special focus on the emails used to capture the sensitivity regarding their creation and manifestation (design, URL, and phishing trick). Finally, the ethical and privacy considerations necessary for the study are presented.

### 4.3.2.1 Study Design

The study mainly took the form of a between-subjects design. This means that almost all measurement time points treat different groups of participants. Only the participants of one group, who were surveyed after six months, were surveyed again after twelve months and evaluated accordingly using a repeated measures design.

Due to the large geographic distribution and thus difficulties in surveying many employees in one place at one time, an online experiment with survey was chosen. This was to ensure that the geographical distance would not influence the probability of participation and thus influence the results. It also had the advantage that it was more resource-efficient for the organization, as the distribution of study materials and the mobility of staff would otherwise have cost the organization even more. At the same time, the employees' work was less interrupted by participation in the study, which was also considered an important factor for a higher probability of participation. Based on the workshop, the study would have had the opportunity to query the overarching awareness regarding phishing. This could have been done by asking participants for a definition and examples of possible attacks. Here, the question would have arisen whether not having a complete list of all the attacks presented is equivalent to not knowing or whether it was only at the moment of questioning that this attack did not have the highest presence. This is contrasted with testing with actual phishing and legitimate examples. Here, an example is presented and the participant must decide whether the example is phishing or legitimate. The advantage here is that the participant can recognize different types of phishing without having to name them. Recognizing different phishing tricks is to be weighted higher than remembering the names conveyed. Also, phishing tricks can be recognized when they are presented that the participant might not have remembered before. After weighing the pros and cons, it was decided to test with examples. There are two main options for evaluating sensitivity with examples. Either send participants emails (which are either phishing or legitimate). To be in an ethical framework, one would have to announce such an action in advance with the indication in case of suspicion of phishing to report it to a special place. Or use an artificial environment with different examples similar to a quiz and give the participants the task to decide for each example whether it is phishing or a legitimate message.

While the first approach has a higher external validity, because it is closer to the actual behavior in everyday life, the second approach has a higher internal validity. Without gaining complete control over the participants' environment, the validity of the decision in the first case is very limited. And complete control in such a study in an organization means both ethical and labor law boundary crossings [98] - which were not justified by the added value of the study. Otherwise, however, it ultimately remained unclear whether an unrecognized phishing message was simply overlooked, for example, or whether other behaviors (such as asking a colleague) were resorted to. In the latter approach, the decision about phishing or legitimate message is the main task with the main focus of the participants. This represents a shift in focus compared to everyday life. Accordingly, this form of survey represents the best possible case and should not be interpreted as meaning that the results can be transferred 1:1 to everyday life. In addition, it is possible to examine many different phishing tricks in combination with comparable legitimate emails with one test and thus make a statement about these different tricks. When sending emails, much fewer different forms could be used and thus a reduced bandwidth could be tested, so that the everyday work is not massively disturbed. The alternative would be to spread the emails out over a longer period of time. However, this would get in the way of the research question regarding the time factor of how long the awareness remains, as this would possibly mean that the ranges would no longer be so clearly separated and very different time periods would also arise for the different phishing tricks. Since this time aspect was the main focus of the study, based on the technical requirements and the serious problems with sending phishing emails, it was decided to proceed with an “artificial” test.

Each group was then assigned to a measurement point at the beginning of the study, so that it was clear in advance that a similar number of employees in a similar composition was available for each measurement point. All groups then received the information about the study and its procedure together. Each group then received only the emails intended for them regarding the study invitation.

Once participants received the study invitation, they had one week to participate in the study. Following this, a reminder email was sent to the entire group. These were always sent to the entire group due to the anonymous nature of the survey. Now the participants had another seven days to participate before the survey was closed.

Eleven groups were defined based on the existing employees. These included seven groups for the normal retention of knowledge survey and four groups for each of which a form of refresher was provided (for details see Section 3.3.4). Due to the large number of employees, several workshop sessions were conducted by the organization. Also, employees who at the time of the first workshop (in October 2018), did not have the opportunity to participate (vacation or illness) were invited to attend another workshop at a later time (end of 2018 / beginning of 2019). To avoid introducing greater complexity and potential effects into the timing of the later surveys, all participants in the later workshop were assigned to either the pre-test or post-test. Thus, the majority of the measurement time points consisted of a unique group of people.

The workshop in the context of the study was an extended form of the training from Section 3.3.3. The decision for the extension was made together with the organization in order to also convey current developments and important aspects specifically for the organization. Accordingly, the workshop consisted of three sections. Section 1 represented the basic sensitization and creation for the topic of IT security awareness. Section 2 dealt with the topic of phishing based on the training from Section 3.3.3. Section 3 covered best practices around the topic of passwords. The first section and the third section were created by the organization’s CISM (Chief Information Security Manager). Section 2 represented an adapted or modified version of the training from Section 3.3.3 for the organization. It was discussed with the organization that section 2 is essential and therefore should be collected exclusively in the study. At this point, sections 1 and 3 are briefly explained for the study and what adjustments were made for section 2 based on Section 3.3.3. Once the first version of the measure was adapted to the organization context the State Office for Geoinformation and State Survey (SOGSS) CISM invited the “persons of contact for information security concerns” (PoC-InfoSecs) to an information and tutorial event. PoC-InfoSecs were needed as these were later on the trainers for the employees for the organization. So a train-the-trainer methodology was used in the SOGGS’s CISM being trained by the researchers and trained the trainers in the organization themselves. In total four such events were necessary to cover all the PoC-InfoSecs. Their feedback and expertise regarding their individual target audiences led to further adaptations of the tutorial. For instance, general security awareness topics, such as ’risk management’ were taken out and the choice of examples was refined to relate more directly to employees daily work. Based on their feedback, the tutorial was improved and sent to the Poc-InfoSecs. It was agreed to use exactly this presentation for their tutorial. Altogether the development of the entire tutorial took six months, covering the first ideas for the structure, the development of the first iteration of the tutorial, organizing the meetings with the PoC-InfoSecs, and the iterative improvement of the PowerPoint presentation.

Section 1 represented a basic introduction to the topic of IT security. For this purpose, general information about previous incidents was presented and behaviors were explained on how to deal with such incidents within the organization. In addition, various statistics about typical dangers and sources of danger were presented, e.g. emails. To convey the importance of this topic, anonymized incidents were presented to show what vulnerabilities continue to exist in the organization.

Section 3 focused on the importance of good passwords. This was done by raising awareness of the risk of using weak passwords. Accordingly, behaviors and methods were also provided on how to generate and use strong passwords. This section also includes a short interactive quiz in which examples were shown with the question how their password strength would be estimated.

To adapt section 2, it must first be said that the training from Section 3.3.3 was initially designed as self-study material. This means that it was a PDF with complete sets that was not suitable to be presented in this form. Accordingly, an adaptation for the PowerPoint format with reduction of the texts to the usual presentation length had to be carried out. In addition to the form of the workshop, the content was also slightly adapted to the context. The entire design of the workshop was adapted to the context of study 2, i.e. the slide design corresponded to the corporate design of the organization. Due to the particular relevance for the organization, the focus of the whole workshop was on the email context. However, in principle most of the content can be transferred to any other context like short messages. To prevent participants no longer following the workshop due to inappropriate information, all content that was inappropriate was deleted or heavily shortened. For example, the use of social media or mobile devices is heavily restricted and only allowed for a small portion of the workforce. As a result, the focus of the workshop was on desktop clients and email in the desktop context. All examples from the original workshop were modified to be more relevant to the everyday life of the employees. For this purpose, examples were used that were created from actual reported phishing emails. These were anonymized for this purpose so as not to allow backtracking. In addition to the form and content of the workshop, the workshop version itself was also modified to meet the train-the-trainer approach.

For this purpose, it was additionally fed with hints that should help the trainers in the implementation. This includes specific requests for behavior (such as asking a specific question at a certain point in time) or assistance to counteract possible problems that may arise at certain points. For example, concrete tasks included showing a sample message and asking the trainer to ask the audience for their opinion on phishing or legitimate by a show of hands and then selecting a person to justify the decision. Afterwards, the trainer was to present the correct solution and address any positive or remaining errors in the justification. This should both hold the audience's attention by alternating between more passive and active elements. It should also increase the awareness through direct feedback.

Additionally, it was clearly defined in advance that five groups would also be surveyed a second time. The group surveyed at the time when sensitivity was no longer significantly improved was, by definition, to be surveyed again after twelve months. Likewise, four additional groups were to be introduced at this time to receive the refresher measure and then also be surveyed again after twelve months. The full study design with all groups and comparisons can be found in Figure 4.5. For this, it was important that these participants generate a subject code as part of their study. This should then allow the individual results of the participants to be linked between the measurement at six months and twelve months without breaking anonymity. This leads to the possibility that different sample sizes could be reported for the six months and twelve months time points for these groups. If not all participants followed the code generation instructions, it could lead to problems linking the data, making the sample larger at the six months time point than at the twelve months time point.

The online survey was conducted as follows: When the participants opened the link from the invitation, they were directed to the Sosci Survey platform. The entire online survey was designed to match the corporate design of the organization. This was to prevent participants from not taking the survey because of the design. In addition, the participants were already informed about the study in advance through several channels, so they knew that the survey platform was legitimate. There, participants were again given clarification regarding the content and process of the study and the context of the organization's collaboration with the university. The participants were asked not to use any external tools to answer the survey and to focus only on them during the survey. To this end, participants were educated that this was not a performance survey of them to the organization. Rather, they were to view the survey as a self-evaluation for their own level of awareness. It was also made clear to them that they did not have to fear any consequences for poor results. As in other common practice, participants were told that participation was completely voluntary and could be terminated at any time without giving a reason. To do so, participants simply had to close the survey. Subsequently, all participants except those in the pre-test and post-test saw a page with to generate the subject code. This was because it was unknown in advance which group was no longer significantly improved in their sensitivity and would be surveyed

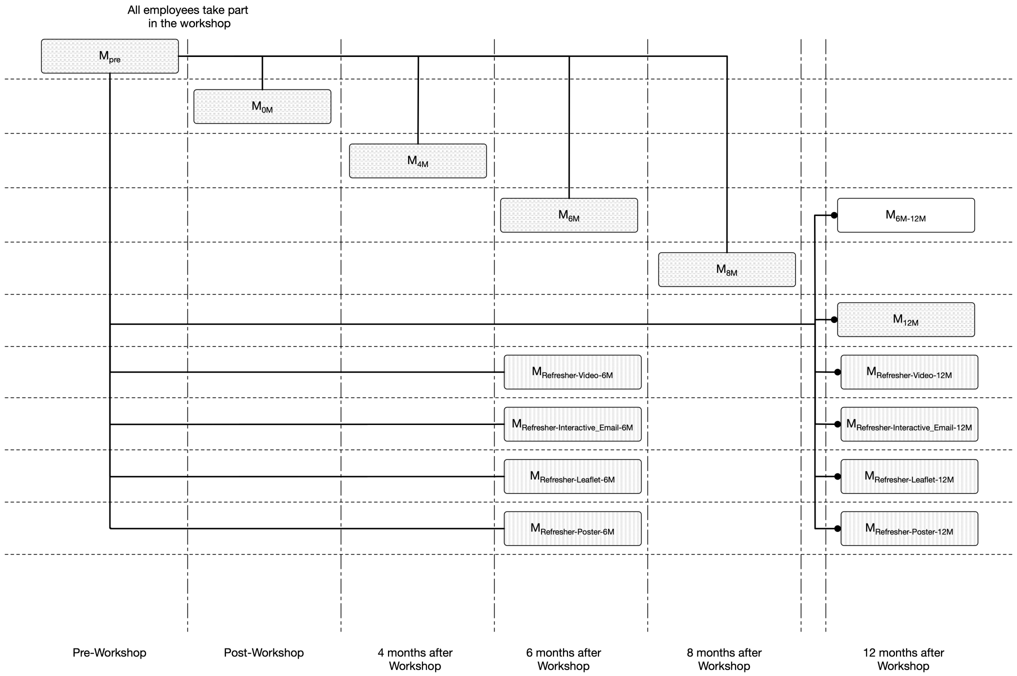

All employees take part in the workshop
$M_{\text{pre}}$
$M_{\text{0M}}$
$M_{\text{4M}}$
$M_{\text{6M}}$
$M_{\text{8M}}$
$M_{\text{6M-12M}}$
$M_{\text{12M}}$
$M_{\text{Refresher-Video-6M}}$
$M_{\text{Refresher-Interactive\_Email-6M}}$
$M_{\text{Refresher-Leaflet-6M}}$
$M_{\text{Refresher-Poster-6M}}$
$M_{\text{Refresher-Video-12M}}$
$M_{\text{Refresher-Interactive\_Email-12M}}$
$M_{\text{Refresher-Leaflet-12M}}$
$M_{\text{Refresher-Poster-12M}}$
Pre-Workshop
Post-Workshop
4 months after Workshop
6 months after Workshop
8 months after Workshop
12 months after Workshop

**Figure 4.5:** Study design with all groups and comparisons. The dotted groups belong to research question 1 and the vertically striped groups to research question 2. The group without markings is the control group for the effect of the double measurement.

accordingly later. Afterwards, all groups received an introduction to the scenario at hand. For this purpose, a small role play approach was used. The participants were asked to put themselves in the role of a person who received the emails and now had to assess whether they were phishing or legitimate. The full text of the scenario was as follows:

> To help you determine how well you can distinguish messages with dangerous content from genuine messages, the following are examples of messages in different contexts. Knowing that you may not know or use all senders, service providers, operating systems, and programs, we present received messages for a wide variety of senders, service providers, operating systems, and programs.
>
> In order not to declare the senders, service providers, operating systems and programs to which you have no connection in reality directly implausible, please assume in the following that...
>
> you are Martin Müller and have two email addresses: martin.mueller.77@web.de. and martin.mueller.77@posteo.de. you speak German and English. your supervisor

> is Wolfgang Lange. His email address is wolfgang.lange.54@web.de. Your work colleague Jonas Schmidt is. He has two email addresses: jonas.schmidt.77@web.de and jonas.schmidt.77@posteo.de. You use all the services that are used in this questionnaire. They use the different operating systems (e.g. Microsoft Windows, Apple OSX, Google Android, Apple iOS) and programs (Skype, Thunderbird, Chrome, Outlook, Apple Mail, Gmail, WhatsApp) used in this questionnaire.[8]

After the participants read the scenario, the tests began to record sensitivity. For this purpose, each participant was presented with 20 examples in a random order. On each page, the participant was presented with one example and a questions below the screenshot: Is this message fraudulent (with the answer options Yes, fraudulent or No, not fraudulent)?

For the groups selected for the refresher measure, the respective measure was presented (video, interactive email example,short text or long text) before the presentation of the 20 examples.

#### 4.3.2.2 Study Material

For the assessment of the effectiveness regarding the sensitivity, a test was chosen as described in Section 4.3.2.1. For the test, examples had to be created that fulfilled various requirements. The examples had to be implemented in the online survey i.e. they had to be images of emails, as due to technical limitations no .xml or .eml files could be used. In order to represent all factors regarding phishing type, operating system, email client or web browser in each combination, an unrealistically large number of examples would be used. For the study, phishing tricks (Content, Non-Brand related Domain, Non-brand related Domain + Brand Outside, Small Deviations in the Domain, Special Link Manipulation and Attachment), Operating System (Windows and macOS), email clients (Outlook, Thunderbird, and Apple Mail) and web browsers (Firefox, Chrome, and Safari) were set up for phishing and legitimate examples, resulting in 288 examples. Since this is a utopian number of examples for the study, it was decided that the focus should be on the phishing tricks. Accordingly, all phishing tricks were to be present and an even distribution was to be found for the other measures. This resulted in 10 phishing examples and one legitimate equivalent each. Since there is still the misconception that http is more insecure than https, all examples were created with “https” in the URL, so that this factor does not have an influence on the decision. The screenshots are based on actual emails sent from authentic senders. Some examples are from fictitious organizations e.g. Explore Friends, Shopping Total, SecurePay or Mein Paketservice or Bauernmarkt total. Others are from real existing organizations like Amazon or from fictitious persons like Jonas Schmidt.

The examples were designed to be static images of the emails. In the examples with a link, the cursor is already over the link. Thus, in this case, the participants could directly recognize the

URL behind the link. In the case of implausible content or the attachment, no cursor was visible so as not to misdirect attention here.

An overview of the phishing tricks used, associated URLs, senders and subject can be found in Table 4.13 and Table 4.14. For more details on the specific tricks see Section 3.1.

#### 4.3.2.3 Recruitment & Ethical / Data Protection Considerations

All communication regarding study implementation, study invitations and reminders of study participation was carried out via the organization's CISM. Care was taken to ensure that the distribution of the groups was as heterogeneous as possible across the entire organization, i.e. that there was not a strong concentration of different departments or geographical locations in one group. As far as this was possible within the framework of the available staff. The allocation of the employees to the different groups was done by the CISM and the information about the exact composition also remained exclusively within the organization.

Only participation in the workshop was mandatory for all employees. This was to ensure that during the study period no incident in the organization would undermine trust in the study. For example, if an untrained employee in a "control group" had fallen for a phishing message during this period, which he or she would have (likely) detected with the workshop - the study would have had a disproportionate negative impact on the organization. In contrast, participation in the sensitivity surveys was voluntary for each employee and could be conducted during working hours. Based on the organization's requirements and in order to exclude identifiers, no socio-demographic data of the participants were collected.

### 4.3.3 Results

In this subsection, the results divided among the different research questions or hypotheses is reported. First, it was checked how long the workshop significantly increased the sensitivity (see Section 4.3.3.1). Subsequently, the research question 2 is to be answered, which of the refresher measures achieves a significantly better sensitivity again after the drop (see Section 4.3.3.2). Afterwards, the systematic long-term evaluation of both the original sample and the refreshment groups were conducted. For this purpose, the research question is answered whether significantly better scores for one of the refresher measures can still be achieved twelve months after the initial workshop. The final research question involves analyzing the differences of the main measure and the refresher measures for the different phishing tricks. Before proceeding to the sections with the hypotheses testing, the descriptive statistics of the average detection rates for phishing, legitimate, and all examples, as well as sensitivity and criterion are reported (see Table 4.15).

**Table 4.13:** Percentages of correct answers per phishing email. “TT” stands for tooltip and “ST” stands for status bar.

| Name | Phishing Trick | More specific Phish | Sender | Subject | TT or ST | Indicator |
|---|---|---|---|---|---|---|
| P1 | Content | Implausible | explore friends <anzeige30294@explore-friends.de> | Job offer at Sicher-Online | None | attachment doc |
| P2 | Non-Brand related Domain | Random 1 | shopping-total.de <kundenbetreuung@shopping-total.de> | Pay invoice | Status | https://www.uhszhiklo.cz/easymoney |
| P3 | Non-Brand related Domain | Fake Tooltip | explore friends <register@explore-friends.com> | Please confirm registration | Status | https://www.jak-shemekuli123.de/registerierungsbestaetigung//mueller/b5244018565445f84458ce29a46485c0a |
| P4 | Non-Brand related Domain | Random 2 | SecurePay24 <no_reply@email.securepay24.de> | Monthly invoice March 2018 | Tooltip | https://www.uhszhiklo.de/ |
| P5 | Non-brand related Domain + Brand Outside | Mismatch | SecurePay24 | Confirmation of your registration for the newsletter | Status | http://www.nshikta-babkr.de/djk=1854525962466=hj2563=sdkkl-dlja-nas364t_hjha.164482=1566 |
| P6 | Small Deviations in the Domain | IP | info@mein-paketservice.de | Convenient parcel receipt with the MPS parcel box | Status | https://www.130.83.167.22/secuso.org.secure-login.de/ |
| P7 | Small Deviations in the Domain | Subdomain | explore friends <accounts@explore-friends.com> | New device | Tooltip | https://www.explore-friends.com.host547.com/account |
| P8 | Special Link Manipulation | Similar Letter | Bauernmarkt total <accounts@bauernmarkt-total.de> | Reminder: Use bonus points | Tooltip | https://www.bauemmarkt-total.de/login |
| P9 | Special Link Manipulation | Extension | Bestellung - Amazon Payments | Your invoice for the Amazon order 028-0903817-1477128 from Dearing [Important] [Important] [Importante] | Status | https://amazon-secure.de/voucher-id=231ksa?/254933-g5ZB6 |
| P10 | Attachment | | hausmeister.schilling@posteo.de | Repair work in your office | None | attachment exe |

**Table 4.14:** Percentages of correct answers per legitimate email “TT” stands for tooltip and “ST” stands for status bar.

| Name | Sender | Subject | TT or ST | Indicator |
|---|---|---|---|---|
| L1 | Jan Horscheidt | Job offer Test Manager near Fürstenfeldbruck | Status bar | `https://www.explore-friends.com/jobs/fuerstenfeldbruck-test-manager-automotive-2785644128?ijt=18&paging_context=search_query%5Bkeywords%5Dautomotiv&search_query` |
| L2 | None | Your My Parcel Service - Rate change ordered | Status bar | `https://control-center.mein-paketservice.de` |
| L3 | Mein Paketservice | Booking information for your airmail shipment | Tooltip | `https://buchung.mein-paketservice.de/serviet/cc?soHHKH7880.2/jdh58h4GLA5.585hdhsbz5sji_LINK:https:DVMDredirectht.2mein-paketservice.2ecomDVMdefault.2easpxDWMd.3http:DVM` |
| L4 | `jonas.schmidt.77@posteo.de` | Fwd: Photos of the Christmas party | Status bar | `https://dropbox.com/pic-lkoi/592xtu499` |
| L5 | Jonas Schmidt | Highlights of the conference | Tooltip | `https://www.faz.net/aktuell/technik-motor/computer-internet/eckonferenz2016/highlights.htm` |
| L6 | Jonas Schmidt <`jonas.schmidt.78@web.de`> | Research assistant at the TU Darmstadt | Tooltip | `https://www.tu-darmstadt.de/forschen/id=324438213` |
| L7 | Wolfgang Lange | Your summary | None | attachment txt |
| L8 | MERCURE Hotel Regensburg F01 <`H6639-F01@accor.com`>[ | AW: Your order 42343224 at HRS.de | None | attachment pdf |
| L9 | SecurePay24 | Confirmation of your newsletter subscription | Tooltip | `http://www.securepay24.de/Online-Anmeldung/nlconfirmation.html?doi=152564322452&mid=f1718-0` |
| L10 | explore-friends <`kundenbetreuung@explore-friends.com`> | User account added | Both | `https://www.explore-friends.com/login/mueller/b5244018565445f84458ce29a46485c0a` |

**Table 4.15:** Overview of the average rates for all groups including the refresher groups (LT = long text, IE = interactive email example, V = video and P = short text) in % for the correct answers for phishing and legitimate examples as well as overall. Together with the values for the signal detection theory (SDT). N = total number of participants per group.

| Example Type | Pre | 0M (Post) | 4M | 6M | 8M | 6M LT | 6M IE | 6M V | 6M ST | 12M | 12M Double | 12M LT | 12M V | 12M IE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N | 72 | 35 | 34 | 37 | 54 | 31 | 43 | 35 | 32 | 36 | 20 | 17 | 17 | 12 |
| Phish | 61.82 | 80.27 | 70.89 | 73.26 | 66.30 | 74.51 | 81.87 | 76.29 | 71.88 | 69.71 | 68.50 | 74.99 | 75.30 | 81.66 |
| Legitimate | 68.75 | 79.14 | 74.72 | 69.75 | 72.78 | 67.74 | 63.72 | 70.29 | 70.61 | 74.44 | 76.00 | 79.36 | 74.71 | 70.83 |
| Overall | 65.45 | 79.68 | 72.90 | 71.42 | 69.69 | 70.96 | 72.36 | 73.15 | 71.21 | 72.19 | 72.43 | 77.28 | 74.99 | 75.99 |
| SDT | | | | | | | | | | | | | | |
| Sensitivity | 1.11 | 2.13 | 1.6 | 1.45 | 1.39 | 1.61 | 1.73 | 1.8 | 1.56 | 1.55 | 1.44 | 1.93 | 1.77 | 1.96 |
| Criterion | 0.12 | -0.05 | 0.09 | -0.06 | 0.15 | -0.27 | -0.40 | -0.07 | 0.01 | 0.12 | -0.08 | -0.11 | -0.08 | -0.37 |

When comparing sensitivity between measurement time points, the data should be examined for differences between groups. Next, the comparisons need to be differentiated into: 1) data from the different measurement time points to no longer significant difference e.g. $M_{pre}$, $M_{0M}$, $M_{4M}$, $M_{6M}$, $M_{8M}$, $M_{12M}$ or the $M_{Refresher-6M}$ and 2) data collected multiple times within the same group e.g. $M_{Refresher-6M}$ and $M_{Refresher-12M}$. The former are independent data, whereas the latter are dependent data. Accordingly, different tests must be applied for the comparisons. Therefore, one-way ANOVAs were selected for the comparisons of independent data and repeated measures ANOVAs for the comparisons with dependent data. For both, the corresponding prerequisites were checked.

#### 4.3.3.1 RQ1 - Long-term Effect

The following research question was how long the effect of the significantly better sensitivity lasts. For this purpose, it was checked after each measurement time whether the values are still significantly better. This was to ensure that the timing of the refresher measures was optimal. No significantly better sensitivity was measured after six months. In order to avoid a random effect, the group after 8 months was still recorded as a control. The one way ANOVA showed a significant difference for the sensitivity between the groups with F (4,227) = 5.457, $p < .001$. The effect size was $\omega^2$ = .093, which corresponds to a medium effect [95]. LSD post-hoc comparison showed that sensitivity from pre-test ($d' = 1.11$, SD = 1.12) was significantly lower than post-test ($d' = 2.13$, SD = 1.15). Also in comparison to time point 4M ($d' = 1.60$, SD = 1.01), there was still a significantly lower sensitivity for the pre-test. In contrast, there was no significant difference for the pre-test and 6M ($d' = 1.45$, SD = 1.01). According to the results, the prerequisite that the workshop leads to a significantly better sensitivity holds. For the first main hypothesis, the hypothesis could be accepted for the measures 0 (post-test) and 1 (four months).

For all subsequent time points, the hypothesis must be rejected. The one-way ANOVA showed no significant difference for the criterion between the groups with F (4.173) = 0.684, $p = .631$.

#### 4.3.3.2 RQ2 - Refresher Retention

Based on the results of RQ1, the sixth month was selected for the time of the refreshment. The question therefore arises as to whether there is significantly better sensitivity after the 6-month refresher than in the pre-test. For this purpose, the pre-test and the four groups with refresher measure were compared using a one-way ANOVA. The test showed a significant difference in sensitivity between the groups F (5.244) = 2.140, $p = .037$). The effect size was $\omega^2$ = .027, i.e., a small effect. The LSD post-hoc comparison found a significant difference from the pre-test ($d' = 1.11$, SD 1.12) for both refresher long text ($d' = 1.61$, SD =1.18), refresher video ($d' = 1.80$, SD = 1.42), and refresher interactive email example ($d' = 1.73$, SD = 1.19). A comparison of the different refresher measures showed no significant difference.

After the refresher measures had been distributed after six months and three of them (long text, video, and interactive email example) again achieved a significantly better sensitivity than in the pre-test, it was now necessary to check whether the effect still existed after twelve months (see Figure 4.6 for a full overview of all comparisons between pre-test and the respective groups). Accordingly, it was decided that no survey should be conducted for the short text group at the twelve-month point. This would only cause unnecessary costs for the organization in view of the fact that the refresher had already failed to meet the requirements at six months. The method of linked sampling by subject codes unfortunately resulted in a reduced sample. Not all subject codes could be clearly assigned to both time points and accordingly these participants had to be excluded from the evaluation. There remained 20 participants for the M$_{6-12}$ survey, 17 participants for refresher$_{video}$, and 12 for refresher$_{interactiveemailexample}$. One-way ANOVA revealed a significant difference with F (5.172) = 2.721, $p = .022$ for sensitivity. An LSD post-hoc comparison showed that the sensitivity of the pre-test ($d' = 1.11$, $SD = 1.12$) was significantly different from that for refresher$_{longtext}$ ($d' = 1.93$, $SD = 1.17$) with $p = .009$, for refresher$_{video}$ ($d' = 1.77$, $SD = 1.32$) with $p = .031$ and refresher$_{interactiveemailexample}$ ($d' = 1.96$, $SD = 1.34$) with $p = .016$). The effect size yielded a value of $\omega^2$ = .047, corresponding to a small effect.

The one-way ANOVA also showed a significant difference for the criterion after the 6-month refresher with F (5,248) = 2.981, ($p = .013$). The LSD post-hoc comparison found a significant difference for the refresher long text ($C = -.23$, $SD = .59$) with $p = .043$ and refresher interactive email examples ($C = -.43$, $SD = .65$) with $p < .001$ compared to the pre-test ($C =$ .12, SD = .84). An optimal value for the criterion would be neutral and therefore 0. Accordingly, the negative values for refresher long text and refresher interactive email example compared to the pre-test mean that these groups had the tendency to make a decision more in the direction of phishing. The one-way ANOVA also showed a significant difference for the criterion after the 12-month refresher with F (4,136) = 1.527, $p = .198$.

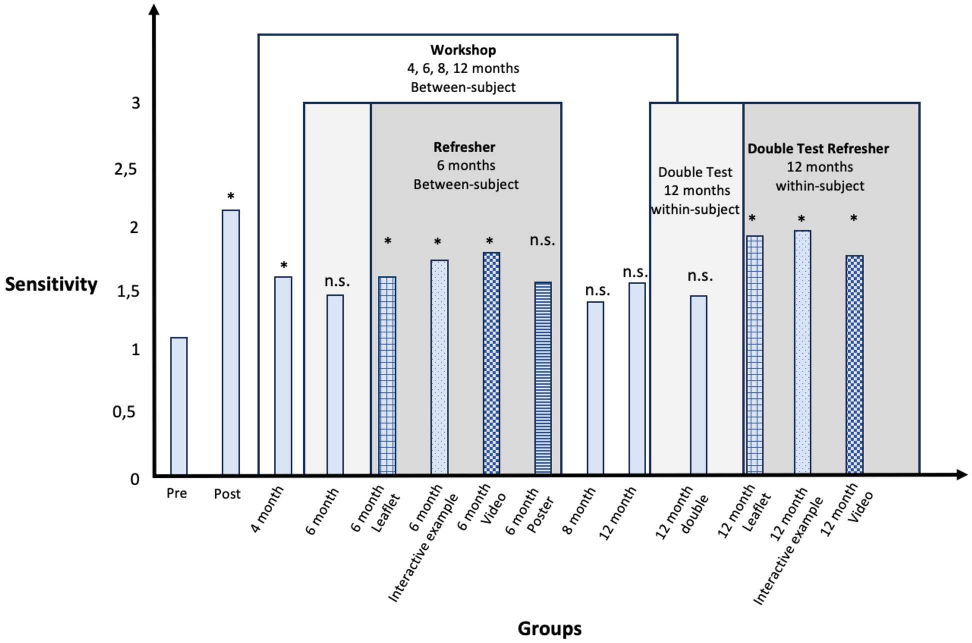


**Figure 4.6:** Study Groups comparing significance to the pre-test. "*" marking those post-hoc comparison that are significant and "n.s." those that are not significant.

#### 4.3.3.3 RQ3 - Phishing Tricks

In this section, the results of each individual example (phishing or legitimate) or phishing trick is discussed. First, the phishing tricks is discussed in more detail. Then continuing with the phishing and legitimate examples. The reported data always refers to the percentage of correct answers across all participants of the respective group or their improvement or deterioration over time in percent (marked with plus or minus).

**Phishing Tricks**

Starting with the overall phishing tricks (see Table 4.16), there seemed to be a clear divide between two groups each consisting of three phishing tricks each. While “Non-brand related

Domain"(78.27%), "Special Link Manipulation"(67.40%) and "Content" (75%) achieved values above the guessing probability for the pre-test, the other three tricks "Non-brand related Domain + Brand Outside" (45.80%), "Small Deviations in the Domain" (46.55%) and "Attachment" (34.70%) achieved below guessing probability. All rates for all phishing tricks improved from the pre-test to the post-test, ranging from 10.70% for "Content" (post-test = 85.70%) to 25.60% for "Non-brand related Domain + Brand Outside" (post-test = 71.40%). In total the "Attachment" phishing trick was still identified merely above the guessing probability (post-test = 54.30%).

Four months later the gap for the correct answers between the phishing tricks increased even further. Some fall back to their respective pre-test level such as "Non-brand related Domain + Brand Outside" (47.10% - only 1.3% higher), "Small Deviations in the Domain" (47.05% - only 0.5% higher) and "Content" (73.50% - even -1.5% lower). Where others still got identified more often like "Attachment" with 61.80%, "Non-brand related Domain" with 89.23% and "Special Link Manipulation" with 82.35%. Six months later the ranking of the phishing tricks remained similar with the "Non-brand related Domain"(91.00%) leading closely followed by the "Content" (89.20%) and the "Special Link Manipulation" (78.40%). The other three phishing tricks remained around the guessing probability 'Non-brand related Domain + Brand Outside" (56.80%), "Attachment" (56.80%) and "Small Deviations in the Domain" (50%).

The refresher after six months lead to an increase for most of the phishing tricks again. There are only two exceptions. The video (-23.60%) and the short text (-6.20%) lead to an decrease of correct answers for the phishing trick "Content". Also the other two refreshers had only small effects on this phishing tricks with increases of 5.60% (long text) and 4.10% (interactive email example). The most improvement could be achieved for the phishing trick "Attachment" where all groups improved from 16.90% (long text), 27.80% (short text), 42.00% (interactive email example) to 56.70% (video). Especially the video lead to an improvement for the "Non-brand related Domain + Brand Outside" with +34.20%, more than all the other groups.

Also six months later the trends remained similar with the "Attachment" benefiting the most from the refresher with improvements from 21.50% (long text), 30.00% (video) and 40.30% (interactive email example) compared to the pre-test. The phishing trick "Content" showed for two of the three groups no change at all and for the third group only a slight improvement of 7.40% (video) was measured. For all three refresher groups the "Non-brand related Domain" phishing trick achieved the highest percentage of correct answers. For the other phishing tricks there was no clear trend across the groups with every refresher having strength and weakness for one or another of the remaining phishing tricks.

**Phishing Examples** Next the phishing tricks are separated into the individual examples (see Table 4.17 for the details). Among the phishing examples, it was noticeable that five examples reached values in the range of the guess probability of 50% or even below (P3 with 54.2%), P9

**Table 4.16:** Percentage of correct answers for phishing tricks for all groups. Also with the data for the refresher groups (LT = long text, IE = interactive email example, V = video and P = short text).

| Name | Pre | Post | 4M | 6M | 8M | 6M LT | 6M IE | 6M V | 6M P | 12M | 12M Double | 12M LT | 12M V | 12M IE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Content | 75.00 | 85.70 | 73.50 | 89.20 | 74.10 | 80.60 | 79.10 | 51.40 | 68.80 | 86.10 | 80.00 | 75.00 | 82.40 | 75.00 |
| Non-Brand related Domain | 78.27 | 91.40 | 89.23 | 91 | 80.27 | 86.03 | 93.03 | 91.43 | 89.57 | 84.23 | 81.67 | 91.67 | 82.33 | 94.43 |
| Non-brand related Domain + Brand Outside | 45.80 | 71.40 | 47.10 | 56.80 | 37 | 61.30 | 69.80 | 80 | 46.90 | 63.90 | 45 | 56.20 | 82.40 | 66.70 |
| Small Deviations in the Domain | 46.55 | 67.10 | 47.05 | 50 | 46.30 | 64.50 | 65.15 | 51.45 | 54.70 | 44.45 | 50 | 62.50 | 61.75 | 70.80 |
| Special Link Manipulation | 67.40 | 91.45 | 82.35 | 78.40 | 78.70 | 82.25 | 91.85 | 81.45 | 81.25 | 72.20 | 77.50 | 81.25 | 76.50 | 87.50 |
| Attachment | 34.70 | 54.30 | 61.80 | 56.80 | 61.10 | 51.60 | 76.70 | 91.40 | 62.50 | 61.10 | 60.00 | 56.20 | 64.70 | 75.00 |

with 50%, P7 with 45.8%, P8 with 43.1% and P10 with 34.7%). As the only example over 90% P2 reached a value of 93.1%.

Comparing the results of the pre-test with the post-test, it was noticeable that most examples improved. Some examples improved strongly e.g. P3 (by 28.4% to 82.9%), P9 (by 27.1% to 77.1%), P7 (by 25.6% to 71.4%), P6 (by 23.3% to 91.4%), P5 (by 19.4% to 100%) and P4 (by 17.8% to 91.4%).

Looking at the period from pre-test to measurement after four months, it was noticeable that there were three categories of development: 1) The example continued to be recognized better; 2) The example were recognized comparably well to the pre-test; and 3) The example were recognized even worse than in the pre-test. Six examples fall into the first category: P10 with +27.1%, P4 with +20.5%, P9 with +17.6% P5 with +16.5%, P6 with +14.3% and P3 with 13.4%. Three examples fall into the second category: P7 with +1.3%, P1 with -1.5% and P2 with -1.9%. And an example P8 with -16.6% made up the third category.

To summarize the results over time, it can be said that most examples were initially improved by workshop and then lost (to varying degrees) over time. In particular, the simple examples P1 and P2 remained fairly constant to slightly improved over time after the workshop.

P4 and P5 initially improved to near-optimal levels and then dropped back down slightly to the range around 80%. Still other examples such as P3 or P6 initially increased and then dropped off more sharply into the 60% to 70% range. The complicated examples such as P7, P8, P9, and P10, on the other hand, initially rose without reaching the good values of the other examples in the 80%+ range and then actually dropped off again very quickly over time.

Looking at the refreshing measures and their effect on the various examples, it was noticeable that the effects differ greatly from example to example (sometimes around 40%). There was very little difference for the example P2, where each refreshing measure achieved a value of more than 85% of correct answers (front runner short text with 100%). The situation is similar for P4, P5 and P6, where values above 80% correct answers are always achieved by each measure. But there are also examples where there are clear discrepancies, e.g. P8 with 40% for video and 64.5% for long text. Also in the example P1 the video achieved only a value of 51.4% compared to the long text with 80.6%. Conversely, video achieved the best value for P7 with 80% compared to short text with 46.9%. And the largest margin is for P10 example with 91.4% for video and only 51.6% for long text.

Now the same comparisons should be made again for twelve months after the workshop and six months after the refresher. The basic trend was similar to the one after six months, i.e. the simple examples P2, P3 as well as P5 achieved the best values across all measures. Nevertheless, there were sometimes significant differences, both between the measures and between those who received the workshop twelve months earlier. Especially for the P6 example with 70.6% for the video and 100% for interactive email example or the P7 with 56.2% for long text and 82.4%, differences were between 25% to almost 30% between the refresher measures. Furthermore, even compared to those who received only the workshop, significant differences were evident. While the simple examples still achieved high rates, there was only a difference of 13.9% for the P3 even with the best measure (12M with 61.1% to interactive email example with 75%). Similarly, the "Attachment" had a difference of only 13.9% with 61.1% for 12M and 75% for interactive email example. In contrast, the examples P9 (12M = 61.1% to interactive email example = 83.3%), P6 (12M = 75% to interactive email example = 100%) and P8 (12M = 27.8% to interactive email example = 83.3%) were very different from the group with only one measure compared to the one with the interactive email example refresher.

**Legitimate Examples** The complete overview of all examples and the number of correct answers can be found in Table 4.18. Looking at the legitimate examples, it was noticeable that three examples only reached a value close to 50% in the pre-test (L1 = 40.3%, L6 = 54.2%, and L8 = 58.3%). However, three examples also reached a value above 80% (L4 = 88.9%, L5 = 86.1%, and L7 = 81.9%). The rest of the examples ranged in between those groups. By the post-test, eight of ten examples improved to a value in the range between 85.7% and 97.1%. Only the two examples L6 (48.6%) and L8 (40%) remained in the range below 50%. None of the examples reached a value of 100% correct answers at the time of the post-test. Three different trends were then seen over the first six months. Some examples remained at their level from the post-test like the examples L5 and L7 in the range of about 90% correct answers or L6 in the range of 50%. Other examples lost slightly around 10% like L2, L4, and L10. The remaining examples lost over 10% up to 28.9% including L1, L3, and L9. A basic trend was that over all measurement times and group paths, the example L6 achieved only about 43% correct answers and never gets above 60%. In contrast, the examples L4, L5, and L7 consistently achieved good values on average between 80% and 90% (with only one outlier for long text at six months with 71% for L7). Also striking was the discrepancy within the interactive email example group at both six months and twelve months. At both time points, the group partially reached values above 75% for part of the examples, at twelve months even 100% for L7 and L10, but also values of only 30% for other examples. Except for the example L6, there was no such discrepancy in any of the other groups or measurement time points.

### 4.3.4 Limitations

First of all, the results of the samples have to be considered with regard to the following limitations: 1) the assignment of participants from different areas of the company was as balanced as possible, 2) the distribution of participants between the different workshop sessions was not controlled, i.e.

**Table 4.17:** Number of correct answers in percent for phishing examples for all groups. Also with the data for the refresher groups (LT = long text, IE = interactive email example, V = video and P = short text).

| Name | Phishing Trick | Pre | 0M | 4M | 6M | 8M | 6M LT | 6M IE | 6M V | 6M P | 12M | 12M Double | 12M LT | 12M V | 12M IE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P1 | Content | 75 | 85.7 | 73.5 | 89.2 | 74.1 | 80.6 | 79.1 | 51.4 | 68.8 | 86.1 | 80 | 75 | 82.4 | 75 |
| P2 | Non-Brand related Domain | 93.1 | 91.4 | 91.2 | 89.2 | 88.9 | 87.1 | 93 | 94.3 | 100 | 94.4 | 90 | 93.8 | 88.2 | 100 |
| P3 | Non-Brand related Domain | 54.2 | 82.9 | 67.6 | 64.9 | 66.7 | 71 | 90.7 | 74.3 | 71.9 | 61.1 | 60 | 62.5 | 70.6 | 75 |
| P4 | Non-Brand related Domain | 73.6 | 91.4 | 94.1 | 89.2 | 85.2 | 83.9 | 97.7 | 94.3 | 87.5 | 83.3 | 80 | 100 | 88.2 | 83.3 |
| P5 | Non-brand related Domain + Brand Outside | 80.6 | 100 | 97.1 | 91.9 | 90.7 | 93.5 | 93 | 88.6 | 90.6 | 83.3 | 95 | 100 | 82.4 | 100 |
| P6 | Small Deviations in the Domain | 68.1 | 91.4 | 82.4 | 94.6 | 66.7 | 87.1 | 88.4 | 85.7 | 81.2 | 75 | 75 | 81.2 | 70.6 | 100 |
| P7 | Small Deviations in the Domain | 45.8 | 71.4 | 47.1 | 56.8 | 37 | 61.3 | 69.8 | 80 | 46.9 | 63.9 | 45 | 56.2 | 82.4 | 66.7 |
| P8 | Special Link Manipulation | 43.1 | 57.1 | 26.5 | 27 | 33.3 | 64.5 | 51.2 | 40 | 46.9 | 27.8 | 35 | 43.8 | 52.9 | 58.3 |
| P9 | Special Link Manipulation | 50 | 77.1 | 67.6 | 73 | 59.3 | 64.5 | 79.1 | 62.9 | 62.5 | 61.1 | 65 | 81.2 | 70.6 | 83.3 |
| P10 | Attachment | 34.7 | 54.3 | 61.8 | 56.8 | 61.1 | 51.6 | 76.7 | 91.4 | 62.5 | 61.1 | 60 | 56.2 | 64.7 | 75 |

**Table 4.18:** Number of correct answers in percent for legitimate examples for all groups. Also with the data for the refresher groups (LT = long text, IE = interactive email example, V = video and P = short text). In contrast to the phishing examples, there is no phishing trick included in the legitimate examples, so the column is not present.

| **Name** | **Pre** | **0M** | **4M** | 6M | 8M | 6M LT | 6M IE | 6M V | 6M P | 12M | 12M Double | 12M LT | 12M V | 12M IE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| L1 | 75 | 85.7 | 73.5 | 89.2 | 74.1 | 80.6 | 79.1 | 51.4 | 68.8 | 86.1 | 80 | 75 | 82.4 | 75 |
| L2 | 93.1 | 91.4 | 91.2 | 89.2 | 88.9 | 87.1 | 93 | 94.3 | 100 | 94.4 | 90 | 93.8 | 88.2 | 100 |
| L3 | 54.2 | 82.9 | 67.6 | 64.9 | 66.7 | 71 | 90.7 | 74.3 | 71.9 | 61.1 | 60 | 62.5 | 70.6 | 75 |
| L4 | 73.6 | 91.4 | 94.1 | 89.2 | 85.2 | 83.9 | 97.7 | 94.3 | 87.5 | 83.3 | 80 | 100 | 88.2 | 83.3 |
| L5 | 80.6 | 100 | 97.1 | 91.9 | 90.7 | 93.5 | 93 | 88.6 | 90.6 | 83.3 | 95 | 100 | 82.4 | 100 |
| L6 | 68.1 | 91.4 | 82.4 | 94.6 | 66.7 | 87.1 | 88.4 | 85.7 | 81.2 | 75 | 75 | 81.2 | 70.6 | 100 |
| L7 | 45.8 | 71.4 | 47.1 | 56.8 | 37 | 61.3 | 69.8 | 80 | 46.9 | 63.9 | 45 | 56.2 | 82.4 | 66.7 |
| L8 | 43.1 | 57.1 | 26.5 | 27 | 33.3 | 64.5 | 51.2 | 40 | 46.9 | 27.8 | 35 | 43.8 | 52.9 | 58.3 |
| L9 | 50 | 77.1 | 67.6 | 73 | 59.3 | 64.5 | 79.1 | 62.9 | 62.5 | 61.1 | 65 | 81.2 | 70.6 | 83.3 |
| L10 | 34.7 | 54.3 | 61.8 | 56.8 | 61.1 | 51.6 | 76.7 | 91.4 | 62.5 | 61.1 | 60 | 56.2 | 64.7 | 75 |

employees from different areas may have attended the same workshop session, 3) due to the train-the-trainer approach, the workshop sessions were not conducted by someone who had been directly trained from an expert, and 4) the workshop sessions were not supervised or controlled. This approach has the potential that, despite the workshop of the respective trainers, the workshop may have differed although all had the same slides. The individual personalities and prior knowledge of the trainers may have had an impact on the workshop itself. Nevertheless, this is a very realistic approach that adds value to the practical implementation of such a long-term combination of measures. Similarly, the distribution of participants across the groups was designed to prevent a single workshop session from having a large impact on any one group.

Phishing training was one of three components of the workshop. It cannot be ruled out that awareness of the other components overrides the phishing aspects, at least to some extent. Accordingly, the positive effect of the workshop could be even longer lasting if it is carried out as a single measure. Even though there might be an influence from the other three components, it's realistic scenario that such similar topics are combined to a single workshop in the organization context.

The study was conducted in a German public sector organization. Accordingly, the sample of participants in the study is not representative of the entire German population. Participation in the study was also voluntary and no socio-demographic data was collected. Therefore, no statement can be made about the representativeness of the sample for the organization. This is a clear limitation, as the total workforce contains a large variance and the sample could be, for example, particularly young, middle age or old. However, it was decided to take this step in order to ensure anonymity and thus trust in the persons conducting the study. The anonymity of the employees, with the certainty that there would be no negative consequences in the event of poor performance, was intended to ensure that their behavior was as authentic as possible, thus making the data of the study reliable, and was therefore judged to be of greater value than the informative value of the socio-demographics.

The measure of the interactive email example was supposed to be presented for at least eight minutes before participants could continue. Due to technical difficulties, this was only the case for three minutes (similar to the long text and video). This means that participants spent less time with the measure than intended. Looking at the times for this page, the average time recorded was 7.3 minutes, which is close to the target time. However, this group scored very well in many areas such as sensitivity and various phishing tricks. Based on the available data, there is no indication that an even longer time with the measure would have resulted in worse scores. It is possible that a longer time would have resulted in better results, but in the case of the study this would not have affected the overall score. Accordingly, the impact of this limitation is considered to be small.

## 4.4 Discussion Part I

This section discusses the results and limitations of the three studies. First, the results for the three main objectives are discussed and linked to the corresponding work. From this, suggestions for future improvements to the measures are derived and further possible alternatives are discussed.

### 4.4.1 Research Questions

The studies of Chapter 4 were designed to answer three main objectives: 1) Do the phishing awareness measures have a long-term effect and how long does it last? 2) What kind of refresher is helpful to maintain the awareness in the long-term? 3) Does the long-term effect of the phishing awareness measures differ for specific phishing tricks? The three studies showed that sensitivity (to distinguish between phishing and legitimate) remained significantly improved both with very short measures such as a video over a period of eight weeks and with a more extensive measure such as the e-learning over a period of five months. The workshop with the train-the-trainer approach led to a significant improvement over a period of at least four months and with the three refresher measures (long text, interactive email example, and video) after six months, the improvement remained significant over a total period of twelve months. In addition, the results of the three studies clearly indicated that the phishing trick “Small Deviations in the Domain” requires the most attention when improving phishing awareness measures.

### How long does the effect of the phishing awareness measures last?

All three studies from Chapter 4 showed that the used phishing awareness measures (video, e-learning, and workshop) can significantly improve the participants sensitivity directly after the measure (see Table 4.19 and Table 4.20). All three measures also achieved significant results over an eight-week-period (video), but also for four months (workshop) or five months (e-learning), respectively. These results complement previous research, which showed above all that phishing detection remains significantly better over relatively short periods of ten days [29] to eight weeks [32]. The fact that the phishing awareness measures examined in this study were able to maintain awareness over a longer period of time may indicate that the measures are designed to be more specific or more comprehensible and therefore provide a deeper understanding so that awareness can be maintained for longer.

A second concern within the first research question was to determine the point in time at which the awareness gained through the measure decreases to such an extent that it is no longer significantly better than before the measure. Due to the study design, this can only really be determined for the workshop that was evaluated in the third study, where the critical loss of awareness was between four and six months. Since the e-learning in the second study lasted five months and this period also applied to the game in another study [33], it can be assumed that at least for more intensive measures such as the e-learning, workshop and games a period of around five months seems to be the sweet spot for a needed refreshment of awareness. A newly published study [34] that examined the same video as the first study for a five-month period was able to show that at least phishing detection was still significantly improved. However, in this new study, the overall correct answers were not maintained for five months, because the classification as legitimate does not improved sufficiently. Overall, the workshop and the e-learning program might be slightly superior to the video in terms of long-term performance, although it is surprising that the five-minute video as a passive intervention achieves almost the same long-term effects. Organizations and users have to carry out a cost-benefit analysis to determine whether the shorter and less resource-intensive intervention of the video is superior to more time-consuming measures like e-learning and workshop with a longer-term effect.

Due to these results, it is important to finally discuss that the way how measures are used can have an effect on the participants performance. While both workshop and e-learning are similar in such factors as involving the participants in an active way, the workshop performed slightly worse in terms of sensitivity (2.13 to 2.66). There may be various reasons for this, e.g., the trainers were only indirectly trained by experts and therefore cannot be described as experts on the topic themselves in the narrower sense. Previous research has already shown that training by experts is more effective than the same training without experts [99]. Another reason could be that not all participants were necessarily involved in the exercise as part of the workshop. Thus, the influence

**Table 4.19:** Sensitivity overview for all studies of Part I for pre, post and the respective retention phases.

| Study | Workshop | | | | | | Video | | | E-Learning | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Group | Pre | Post | 4M | 6M | 8M | 12M | Pre | Post | 2M | Pre | Post | 5M |
| $d'$ | 1.11 | 2.13 | 1.6 | 1.45 | 1.39 | 1.55 | 0.599 | 2.23 | 1.83 | 1.28 | 2.66 | 2.47 |

**Table 4.20:** Sensitivity overview for all refresher groups from study 3 (LT = long text, IE = interactive email example, V = video and P = short text).

| | Workshop refresher | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Group | 6M LT | 6M IE | 6M V | 6M ST | 12M | 12M LT | 12M V | 12M IE |
| $d'$ | 1.61 | 1.73 | 1.8 | 1.56 | 1.55 | 1.93 | 1.77 | 1.96 |

of active participation in the exercise and the implementation by experts could be investigated in more detail in the future to determine these influencing factors more precisely. Also having the workshop on the phishing topic together with other similar topics, might be efficient for the organization, but also be confounding in retaining all the necessary knowledge.

**Which type of refresher would result in a significantly better sensitivity in the long term?**

The second objective was to analyze, what kind of refresher can refresh the awareness, which has dropped over time, so that a significant improvement can be seen again compared to before the phishing awareness measure? In the Section 4.3, four very different refresher measures (short text, long text, video, and interactive email example) were selected. The study in Section 4.3 showed that of the four refreshment measures used, three could achieve another significant improvement after six months compared to before the first measure. Here it is to be noted that the short text as most compressed form with only the most important tricks to recognize, was not sufficient. Future research should try to answer the question what factors lead to the results, e.g., if it was due to the compressed content or also the very sparse presentation as pure text with some examples. In contrast, the longer text with similar content (the same tips and tricks for recognition), but additional information that was actually irrelevant for the test, achieved a significantly better result. It could be investigated in the future whether the longer text with backgrounds and further context increases the intrinsic motivation [100, 101] and thus promotes the processing of the content and therefore longer retention. The backgrounds and context are provided, albeit in a different form, but basically in all three other refreshment measures (video, long text, and interactive email example). From this it can be concluded that it is not simply that the shorter the focus on the most important information, the easier it is to remember, but that the embedding and preparation of the information seems decisive. It is noticeable that the interactive email example was able to maintain its achieved values more consistently, even though the video could occasionally also achieve good or even better values. The long text often showed similar trends to the interactive email example, only usually slightly worse. This could mean that the interactivity with the information where it is needed, such as the tooltip, helps to improve awareness. In contrast, the video often seemed to show contrary trends to the interactive email example related to the phishing tricks, e.g., while video performed better for the “Non-brand related Domain + Brand Outside”, the interactive email example performed better for the “Special Link Manipulation”. Accordingly, either the video or the interactive email example performed better over the 12-month period. Future studies could re-examine the refresher in both directions, e.g., to shorten or lengthen the individual refreshers to distinguish which effects are due to the kind of the refresher and which are due to the amount of information. Another possibility for future research could be that even the request for a specific reminder of the content of a past measure could contribute to refreshed awareness without repeating the actual content. Alternatively, if sufficient resources are available, is of course to investigate whether repeating the measure makes a significant difference instead of a slimmed-down refresher in the long-term.

**How long do the effects of the phishing awareness measures for specific phishing tricks last?**

Not only do the phishing awareness measures have different overall effects in the long term, but they may also differ in terms of which phishing tricks they are effective or less effective for. Therefore, the third objective was to find the strengths and weaknesses of the measures for different phishing tricks. There are two different dimensions in which the phishing tricks can be evaluated, one being the influence of the different measures on the phishing tricks and second being the development over time. Overall, it can already be stated across all measures that there were major differences in their impact on various phishing tricks and that no measure could achieve optimal scores for all tricks over an extended period of time. It is therefore worth taking a closer look.

Basically all three measures could achieve a good result for the “Non-brand related domain” phishing trick between 90% and 100% directly after the measure. So there is basically no need to doubt that the measures for this trick achieve very good results in the short term. For the other three phishing tricks (“Non-brand related Domain”, “Non-brand related Domain + Brand Outside” and “Special Link Manipulation”), all measures at least achieved significant improvements in the post-test. But looking at the results in detail, one notices that not every measure reached an optimal level for every trick. Each of the measures seemed to have at least one phishing trick in the post-test that performed worse than the other phishing tricks and also worse than the other measures. For example, the video seemed to have its strengths in “Small Deviations in the Domain” (95.45%). A possible explanation could be that the other two tricks are either not included in the video explicitly “Special Link Manipulation” and therefore only synergy effects of the explanations can address this trick, while “Non-brand related Domain + Brand Outside” could be influenced by the one example where the mouse cursor is directly over the part where the impersonate domain is placed, which could have mislead the participants. Additionally, “Small Deviations in the Domain” even though with examples and an example how easy such small “errors” can be overlooked, could not have performed as good due to the examples being directly presented with highlights and therefore were too obvious to create the necessary awareness. The e-learning had its strengths in “Non-brand related Domain + Brand Outside” (95.31%) and “Special Link Manipulation” (93.75%). A likely reason could be that the part for the “Small Deviations in the Domain” is relatively late in the e-learning and therefore it might be that participants are already kind of overloaded with information. The workshop, achieved very good values for “Special Link Manipulation” (91.45%) but performed worse than the other two measures for the other two phishing tricks just mentioned. A possible explanation could be that “Non-brand related Domain + Brand Outside” are only indirectly addressed with tips to check the who-area, but no specific example for either the subdomain or path attacks are mentioned. Combining the findings from different studies, the video for “Non-brand related Domain + Brand Outside” achieved a value of approx. 80% in study 1 in the post-test, as well as a similar value in study 3 as a refreshment. On the other hand, the value of video as a single measure drops below 70% already after eight weeks, but as a refresher the value can even be kept constant at approx. 80% over six months. A possible explanation could be that the video itself with the lack of exercises and lack of interactivity does not reach the same level of user involvement as the other measures. When the video is used as a refresher, where the needed awareness has been already achieved in the past, it still displays the important aspects and is therefore enough to refresh the awareness. In this case, the video in combination with the workshop seems to have a positive effect for a longer period of time. Interestingly, this effect also shows up for “Non-brand related Domain” to a very similar extent. In contrast, the trend looks different for the other two phishing tricks. For the “Non-brand related Domain” very similar values of around 80% were measured in study 1 and study 3 at least in the post-test , but there was a slight drop in the video twelve months later. Additionally, for the “Small Deviations in the Domain”, noticeably lower values were already measured in the post-test of study 3 than of study 1 (50% to 95%).

The question now is what caused this and how best to address it. One cause could be that the measures point out the important aspects to the participants, especially the URL. But whether this was used as the final basis for decision-making was not checked in the studies. Most importantly, the influence of the URL as a basis for decision-making would have to be discussed, since the URL was already visible to the users. In total, combining the results from all studies the “Special Link Manipulation” and “Non-brand related Domain + Brand Outside” seem to need some attention, but especially looking at the refreshments with vastly different results. “Special Link Manipulation” seems to be addressed rather well by all refreshments with some room for improvement over the full twelve months. Where “Non-brand related Domain + Brand Outside” results differ greatly, depending on the refreshment, ranging from pretty bad (56.20% for long text) to good (82.40% for video). It needs further analysis why all the other refreshments perform far from the optimum for this trick. The main focus should lay with the “Small Deviations in the Domain” even though still pretty good after eight weeks in study 1, the trick performs the worst in the two other studies and therefore should be in the focus of further research. It seems to be very difficult to help people detect such a phishing trick over a longer period of time. As this needs the most attention and focus, this trick probably is also the most difficult one to detect in the real life.There is a possibility that one phishing awareness measure can only go part of the way and other solutions are needed to tackle the problem in the long term.

One possible explanation for the different results of the various phishing tricks could be that everyday life is insufficient for all phishing tricks to apply and consolidate the awareness learned to the same extent. For example, there are too few points of contact with fake tooltips. On the other hand, the essential awareness and important fundamentals remain in place in the long term. The remaining important fundamentals can be seen from the fact that even in the 12M group, there was a slight drop in the number of simple phishing tricks - but there was still an improvement compared to the pretest. The participants, therefore, have learned over such a long period that they should look at the URL behind the link.

In addition to the different results of the phishing tricks, one particular trend stands out. Both in the three studies (with the exception of e-learning for the “non-brand related domain”) and in the related work [3, 33], no measure achieved an optimal value of 100% at any time. This trends is especially interesting taking into account that the measures emphasize the importance of the URL respectively domain and that, unlike in reality, the participants did not have to hover over the link to display the status bar or tooltip. In addition, the studies represent a “best-case” scenario, as the participants’ main task is to differentiate between legitimate and phishing. Despite all these factors, a value of at least 80% correct answers is only measured for the phishing trick “Non-brand related Domain” for each measurement time. Accordingly, it should be noted that the measures do create a significant improvement compared to the previous point in time and can also keep this constant over different periods of time. However, there is still potential for further improvement for all measures and points in time. In addition to awareness measures, there is at least one other way to help people distinguish between legitimate and phishing messages. Another option would be to use a tool support to provide information and technical features directly at the moment when this decision has to be made. Therefore, in the following Chapter 5, a phishing awareness measure is compared with tool support.

### 4.4.2 Future Work

The phishing awareness measures from Part I all lead to significant improvement of the participants sensitivity and most of them could also maintain this improvement over a longer period of time (four to six months for e-learning and workshop). Still there was a gap for even further improvements as e-learning scored the best value across all measures and studies with $d' = 2.23$ for the post-test (maximum being $d' = 3.38$). There are various reasons for the non-optimal values which can provide helpful clues for further research.

One main reason for not optimal performance could be that from other areas it is already known that only parts of the knowledge can be acquired with the first run through [35, 36]. Therefore it needs to be repeated to give the chance to acquire all the knowledge. This could be done by two ways either repeating the same measure or doing a repetition with a different measure.

Such a refresher represents an important milestone to remember in the long term. Based on the current results, three out of four different refresher measures have been able to maintain a significant increase over a period of another six months (twelve months in total). These three measures were all based on the same concept in terms of content and just had different ways to deliver the content. In there, the video (with weaknesses in the “Small Deviations in the Domain”) and the interactive example (with weaknesses in the “Non-brand related Domain + Brand Outside”) seem to complement each other best. The question here is, whether one of the measures should be adapted to its weakness or whether the measures should simply be combined. Even a combination of the measures with a duration of five to ten minutes each still represents a significant reduction in time compared to the original measure.

Another reason for the lower results for the measures video and workshop, especially for the post-test, could be that participants are not forced to test their awareness and do not get feedback on their performance, e.g., their strengths and weaknesses for different phishing tricks. A very short measure (such as the video), without the possibility to practice one’s own awareness and where the participants are only passively taught the content, will probably need a faster refreshing than a longer measure where the participants also test their own awareness (the e-learning). Nevertheless, the workshop from study 3, for example, also has interactive elements with exercises. But compared to the online training in study 2, not every participant necessarily had to complete the exercises successfully. In such a workshop setting similar to a class room only a sample of the participants will complete the exercises in the public group. Accordingly, if participants incorrectly assess their own performance and thus omit these exercise opportunities, they may also need a quicker refresher in such a measure. This should be explored more systematically in the future. First, it could be investigated whether the active practice and especially different intensities of exercises, have an influence on the long-term effect. Second, both the workshop (with practice) and the online training provide an opportunity for feedback to the participant. Therefore, both can directly address mistakes that participants make. Future studies could systematically examine whether general feedback (e.g., total number of correct answers) to specific feedback (examples made incorrectly and the strategies needed to address them) have an impact on the long-term effectiveness of the intervention. Furthermore, more passive measures (such as the video) could be evaluated against various forms and manifestations of more active measures.

Study 3 (see Section 4.3) evaluated the factor repetition with two measures (workshop + either video, interactive email example or long text) spanning over twelve month in total. A first foundation has been laid by systematically recording effectiveness over twelve months, with the findings that the three measures show a significant effect. However, the question arises, for example in the organizational context, what needs to happen after this period, e.g. should the first measure be repeated, or is it enough to repeat the second refresher measure (video, interactive email example or long text)? Another possibility would be to use a completely new measure, i.e., crossed with the second measure, or a completely different one, e.g., a short training. Here, it remains to be clarified how many different measures are needed before a repetition can create the same results again.

Another clue would be whether a single training intervention may not achieve an optimum at all. Perhaps it is necessary to combine such a measure with further technical support, which makes it easier to recognize certain particularly complex and in everyday life rare sophisticated phishing tricks. Here, of course, it should be checked to what extent the effect on phishing tricks can be attributed to the measure or the technical support. It is also possible that the measure only provides the basic awareness for the topic of phishing and the technical tool supports or decides for the user. Due to the fact that all the selected measures did not consistently reach an optimal value for sensitivity, it should be discussed what could be possible alternatives. On the one hand, a measure could be selected that targets the vulnerabilities of individual users even more specifically and accordingly trains only these vulnerabilities as part of the measure. This could be carried out until all phishing tricks reach an optimal level. Then, as part of smaller measures, repetition could be performed where the correct answers are deferred to a later repetition and the incorrect answers are repeated in a more timely manner. This is consistent with the principle of learning by repetition with flashcards that is often suggested for students on a wide variety of topics as well [100]. It could also be investigated whether the combination of an awareness measure in combination with a technical measure could lead to the desired results of optimal sensitivity to all phishing tricks. For this purpose, one of the tested measures could be tested individually or in parallel with a technical solution and then in combination. Thus, it could be verified whether the effects are due to the measure, the technical solution or the combination of both (which is done in part two of this thesis).

### 4.4.3 Overall Limitations

In this section, the limitations of the general embedding of the study, e.g. story and study design, is first explained and then the limitations to the content of the studies with a focus on the study material, in particular the messages used, is discussed.

Similar to other studies in the field of phishing detection [53, 3], the primary task in this study was to decide between phishing and legitimate messages, and there were no other tasks to divert attention. Therefore, it is always a “best case scenario” statement that would not apply in reality. However, the survey does add value and can be seen as a basic requirement for a measure to work. Even in this “best-case” scenario, for some examples only a value close to the rate probability could be measured in the pre-test (see Table 4.5) and thus the prerequisite of a not too high pre-test result is given. However, almost perfect recognition rates for e.g. `https://www.dtrdtcbj.com/de/en#blade` with 100% (as in Table 4.11) probably cannot be achieved in reality. In contrast, other types provide evidence that they require some other form of support. For example, the “Small Deviations in Domain” examples (see Table 4.11) showed results that are barely acceptable in the post-test, even less in the retention-test. Another factor that leads to the studies being more of a “best-case” scenario is that in the examples used, the mouse cursor was already over the link and the URL could be recognized. Thus, the important step that the participants must first know how to get to the URL was omitted. Therefore, no statement can be made about the awareness of this aspect. Design and plausibility would probably have a higher value in the evaluation if the URL was not directly visible. However, since the examples were all based on real existing messages, this effect should not have been more relevant in this case. Future studies should investigate whether a solution can be found with other technical tools and check whether this has an effect on the results.

Alternatively to the decision between phishing and legitimate examples, one could have asked other questions and used a cover story to deceive the participants of the actual goal of the studies. This would have improved the external validity of the results. However, there are several difficulties here, which mainly concern internal validity. On the one hand, this cover story must be credible and also be maintained over the entire duration of the experiment. Second, it is needed to determine when this is no longer the case and therefore decided when decision are not made under the cover story anymore. In an online experiment, this is not easy to ensure. Most importantly, given the video as an intervention, it would have been difficult to establish another subject as authoritative. Instead, one could have tried to evaluate the video in the context of a field study. However, even field studies have idiosyncrasies that should be considered more relevant to later studies given the initial evaluation over a longer period of time. If a message is not classified as phish in a field study (not clicked on), then this can have various reasons, e.g. the message was intercepted by a filter, simply “overlooked” or sorted out on the basis of a non-matching company (i.e. classified as spam rather than phish). Also in the case of the opposite, if the message or the URL was clicked, then it could be because the participants knew that it was a message part of the study and were curious what could happen. One could also try to capture this in the context of an additional survey, but the social desirability or shame could very strongly influence the answers here. Also basically the control over receiving the messages or the respective channel (desktop PC, laptop or smartphone) is very difficult to control and has an enormous influence on the results.

In all within-subject design studies, the same messages were used, albeit in random order. Thus, one factor for the good post-test scores could be that participants recognized the examples. However, several factors speak against this having a major impact. First, participants did not receive feedback on whether their own assessment of the example was correct or incorrect. So if they recognized the examples, they would have to remember their answers at the same time and could then have given them again, but without knowing whether they were correct. Second, the focus of the study was on the retention-test, and it is highly unlikely that the study was relevant enough for the participants to retain the examples and their responses over the eight-week period. It is more likely that they recognized the examples per se, but then made a new decision again.

With regard to the study material there are also some limitations, in particular with the examples used and the distribution of phishing and legitimate examples. Two of the major influential factors for the amount of examples used is the distribution between phishing and legitimate examples and the amount of phishing tricks included. Every included trick increases the amount of examples, and when done properly even more than by one example. Especially if the phishing tricks are essential for the evaluation, it should be considered using more than one example per phishing trick. Otherwise, the answer is based on only one example and other factors, such as the appearance of the email, may distort the effect. Furthermore using a lot of examples in a study can lead to other effects that need to be controlled, e.g the learning effect and the fatigue of the participants. In such a case, the study design has to be adapted accordingly. For example, the order of the examples should be more strongly controlled and, if in doubt, manipulated. Another possibility is to increase the sample size in order to obtain as many permutations of the distributions of the examples as possible and to be able to make corresponding statements about learning effects and fatigue over the duration of the study. Some research has already been conducted on the distribution between phishing and legitimate examples, as unfortunately, there is a lack of reliable information on the actual distribution in real life. When the phishing percentage was very low, such as 1% (as in [102]), the detection rate for phishing was also very low and even on a level of the guessing probability (around 50%). While between 5% and 20% there seemed to be no difference for the phishing detection rate. This research did not include a measure that trained users and therefore lacks the information about the influence of such low rates when participants were trained to detect phishing and instead was more based on their pre-study awareness. Another study included a measure together with pre- and post-testing and investigated higher rates such as 25%, 50% and 75% [103]. In general, their training neither improved phishing detecting nor lowered the false positive rate by a large margin, especially for low percentages of phishing examples. Still they found that a higher rate of phishing increased the hit rate, but also increased the false alarm rate. While the confidence of the participants increased more, when the phishing percentage was higher. Concluding the phishing rate especially for studies with training should not be too low, as participants are lacking the option to test their newly acquired awareness to detect phishing and their confidence is not boosted, too. Therefore, it can be concluded from the previous studies that although the distribution of phish and legitimate messages is probably not 50:50 in reality, this is a suitable distribution for the learning and study context.

Another aspect of the examples is the expected probability that they is correctly classified as phishing or legitimate. In the context of an experimental setting, with reference to classical test theory [104], it is essential to integrate different so-called item difficulties. In this case, this means to include a mixture of some messages that can be correctly classified as legitimate or phish with a moderate probability, but also messages that are correctly recognized with either a very high or very low probability. The detection probability was not systematically varied in the studies and could be a further focus in subsequent studies. However, it can be noted that the distribution of legitimate examples in the studies already covered the spectrum of guess probability to perfect detection (50% to 100%). In addition, the general focus on examples that can only be correctly classified by URL makes them more difficult than those studies that also use messages that can also be classified by aspects such as context. Also such emails could be more difficult than the average phishing message that users are confronted with in their life. It would need to be verified whether the rates found can be confirmed in a scenario with higher external validity, such as a field experiment. If a measure does not achieve optimal values even in such a scenario, then the measure should first be optimized before a more elaborate study such as a field study is conducted.

Finally, both in the studies Section 4.1 and Section 4.2 each involved samples with a very limited number of participants. The first study had a small sample size of only 16 participants, the sample was not representative for the retention-test. Thus, the findings of the study should be confirmed with a larger and more diverse sample. At this point, further findings regarding the phishing tricks could also be included to gain additional new insights. For the second study the participants were not paid and participation in the retention test was voluntary, it was to be expected that the sample would decrease as a result. Accordingly, only around half of the participants gave their e-mail address for further contact. Thus, only about half of the participants could be made aware of the second phase by email, of which in turn another half responded to the call for participation. Consequently, the sample only reached a size of 22 participants. With such a small sample size a generalization for the whole population can not be made. Nevertheless, the results indicate that the video can effectively increase sensitivity over a more extended period than eight weeks. As the values from the post-test to the retention hardly dropped, this provides an indication that a longer period could also be assumed for the increase in awareness. Due to the decision to initially evaluate an 8-week period for the short measure, future studies with more participants can extend this period of eight weeks and thus consolidate the results of this study.

# 5 Part II: Phishing Awareness and Tool Support: A Synergistic Approach

So far, it has been shown that many of the awareness measures can contribute to an improvement in discriminating phishing and legitimate messages among users. However, as can be seen in Chapter 4, none of the measures achieve the optimum in terms of sensitivity; neither immediately after the first measure nor after the refresher. A possible solution to improve the sensitivity could be to support the user with a technical tool to help differentiating between the phishing and legitimate messages.

Previously, there has been little research on how these two interventions, phishing awareness measures and tool support, compare directly. Theoretically, it is possible to compare the results from different studies based on their metrics, e.g. effect sizes. However, there are always differences in the study design that make an indirect comparison only partially robust. For example, the choice of examples can have a large impact on the effectiveness of an intervention (as seen in Section 4.4 with different results for different measures for the same phishing trick). Also, there may be examples that are better addressed by one or the other intervention. Examples that are less likely to be addressed by an intervention are also less likely to be included in a study, leading to the comparability difficulties described above. Accordingly, there is still little knowledge about which strengths and weaknesses the respective intervention reveals in a direct comparison. Most importantly, there is little or no knowledge about the synergistic effects and improvements that can be achieved by combining two interventions.

To avoid constantly reinventing the wheel, existing, state-of-the-art interventions were used to evaluate the two types: Phishing awareness measures and tool support. For this purpose, the video from the NoPhish concept as an awareness measure [6] and TORPEDO [76] as a tool support are compared against each other. These two interventions have been selected as representative of their type, as both are ready-to-use measures for both awareness ([6]) and tool support ([76]). Both represent interventions that are very easy to introduce and also easy to distribute to a large number of staff or users. Additionally, both interventions should require about the same level of effort to implement.

The chapter is structured into four sections each on a specific aspect of one single study. First the research questions and hypotheses of this part (see Section 5.1) are described in detail. Then the methodology to evaluate the four different variants of TORPEDO and the two status quo options with and without the NoPhish video is reported (see Section 5.2). The results of the evaluation (see Section 5.3) and the discussion (see Section 5.4) are closing this part of the thesis.

**Contributions described in this chapter:**

- Confirmation of previous research on a phishing awareness measure (video) and a phishing tool support (link-centric warning) to be effective in increasing the ability to distinguish between phishing and legitimate messages in a single, between-subjects study design [9].
- Novel comparison between a phishing awareness measure (video) and a phishing tool support (link-centric warning) showed the tool support to be a better single intervention [9].
- Novel combination of a phishing awareness measure (video) and a tool support (link-centric warning) has proven to be more effective than a single intervention [9].
- Novel evaluation of specific phishing tricks revealed that only the combination achieved near-optimal effectiveness with regard to the dangerous phishing trick „Small Deviations in the Domain“ category [9].

## 5.1 Research Questions and Hypotheses

The present study is intended to answer a series of five research questions. First, results of previous studies should be confirmed. This is to ensure that the present study does not lead to (unconscious) changes in the results for already evaluated measures and technical solutions due to slight deviations in the study design. Accordingly, the first two research questions aim at confirming the results of previous studies and are considered as basic prerequisites for further analysis. The first research question is aimed at the video from the NoPhish concept used in the study and in the past:

**RQ**$_1$: *Does the video as a phishing awareness measure influences the viewers sensitivity?*

Accordingly, the hypothesis to be tested is as follows:

**H**$_1$: Participants show significantly higher sensitivity immediately after the video compared to those without the video.

The second research question deals with the link-centric warning under investigation (TORPEDO). TORPEDO has also shown in the past that it can lead to a significant improvement in phishing detection. Accordingly, the second research question is.

**RQ**$_2$: *Does the link-centric warning as a phishing tool support influences the users sensitivity?*

And the associated hypothesis to be tested in the study is as follows:

**H**$_2$: Participants show significantly higher sensitivity when using the link-centric warning compared to those without using the link-centric warning.

Following this, the research questions that form the focus of the study is addressed. The first question is whether, in a direct comparison, one of the two interventions is better in terms of increasing sensitivity. Accordingly, the third research question is:

**RQ**$_3$: *Is there an option to be preferred between the phishing awareness measure (video) and the tool support (link-centric warning) in terms of sensitivity*?

This results in the following hypothesis to be tested:

**H**$_3$: There is a difference in sensitivity between participants with a video and participants with a link-centric warning.

Subsequently, the question is whether the combination of a phishing awareness measure and a tool support still offers additional added value compared to the respective individual intervention. It could be that the two intervention each have particular strengths in a different area of phishing tricks. Thus, the combination could have synergistic effects that go beyond the individual intervention. This leads to the next research question:

**RQ**$_4$: *Does the combination of a phishing awareness measure and a tool support improve sensitivity better that a single intervention?*

Accordingly, the hypothesis is as follows:

**H**$_4$: There is a difference in sensitivity between the combination of interventions compared to the individual intervention.

Similar to the individual studies in Chapter 4, the individual interventions are also examined for their influence on the various phishing tricks. The phishing tricks require very different behavioral cues as well as different forms of technical support to be detected by the user. In the course of this, the different risk levels of the link-centric warning are also examined more closely in terms of their influence.

From this, the last research question is derived:

**RQ**$_5$:*What effects do the interventions or their combination have on the different phishing tricks?*

Due to the low number of examples per phishing trick and so far unknown effects, exploratory investigations are conducted. In comparison to the previous hypotheses, only the hit rates and miss rates for the phishing examples are explored here, so no specific hypothesis is formulated for this research question.

## 5.2 Methodology

This section describes the methodology of the study in detail. First, the basic study design is discussed in depth (see Section 5.2.1). In there, it is discussed which groups are addressed in the user study. After that, the materials used for the study is reviewed again. This includes a brief description of the interventions, the questionnaire used to measure sensitivity, and the messages used.

Due to the very different status quo situations of potential users, two control groups have been used for the study. These two groups intended to represent the current status quo for users in the area of e-mail. One was the display of the URL behind a link in a status bar, as used in the Thunderbird e-mail client, for example.

This resulted in the first (control) group 1 - Status bar: When the link was touched with the mouse, a status bar similar to the one in the web browser or Thunderbird was displayed at the bottom left of the e-mail. Another often used display is a simple tooltip. In this tooltip, there is either only the URL or a very short information how to interact with this tooltip. This representation is common, e.g. in the Microsoft Outlook e-mail clients or in Apple Mail.

This was the basis for the second (control) group 2 - Tooltip: When the link was touched, a tooltip was displayed with the URL of the link. The display was very close to the one in Microsoft Outlook. Above the URL a short text was displayed, which reads: “Click or tap to follow the link”.

Next, moving on to the different intervention groups, which included one or both of the interventions. First, it was the aim of confirming the previous positive results about the video. This had to be checked for the status bar as well as for the tooltip individually in each case.

This resulted in each case in the first (intervention) group 3 - Video + Status bar: This group had the same display as group 1, i.e. a status bar at the bottom left of the email. The difference was that before this group got to the task of deciding on the emails, it first saw a video as an intervention.

And it also resulted in the second (intervention) group 4 - Video + Tooltip: This group had the same presentation as group 2, i.e. a tooltip right at the link. The difference was that this group still saw the same video as group 3 as an intervention before the decisions were made.

Next was the confirmation of previous results on the link-centric warning. The link-centric warning overrode any kind of other tooltips and at the same time still allowed the display in the status bar. Accordingly, only one intervention group was needed at this point to represent the scenario of the deployment.

Accordingly, the third (intervention) group was group 5 - Link-Centric Warning: this group received a replicated representation of the link-centric warning for each example. Whenever the cursor was moved over one of the links, a corresponding tooltip was displayed with the appropriate representation of the link-centric warning. The link-centric warning also included a matching tutorial during installation, which presented the most important functionalities and representations in a short form. This tutorial was presented on a page directly before the decisions. The link-centric warning was an adapted version of TORPEDO. The difference between the low risk, unknown risk and unknown risk with indicator are the most important features to test, e.g., testing if the unknown risk with indicator leads to more false positive decisions. Even though it would be interesting to understand the perception of the low risk history level, the introduction seems more difficult as these represent highly individual URLs. Also compared to the other three included risk level this risk level should not be handled differently to the other low risk level (see Section 3.4.2 for the background of TORPEDO and Section 5.2.1 for the adaption for the study). The reduced scheme for the study can be found in Figure 5.1.

Furthermore, it was examined whether also the combination of the two interventions has a larger effect than the single interventions. Accordingly, this became the (intervention) group 6 - Video + Link-Centric Warning: This group received both the video similar to groups 3 and 4 for the introduction, but also the tutorial for the link-centric warning. Subsequently, a corresponding tooltip of the link-centric warning was also displayed for each example in this group. This group thus combined all possible interventions and represented the maximum support in the context of the study.

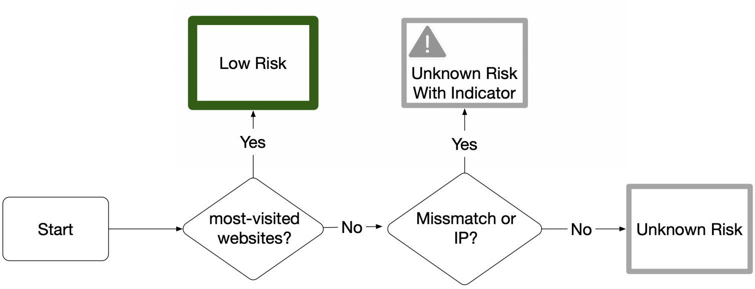


**Figure 5.1:** Adapted link-centric warning (TORPEDO) scheme for the study.

The last two (intervention) groups once again focused on a special feature of the link-centric warning. This was about the tutorial, which has so far been seen as part of link-centric warning. The question arose as to how necessary this tutorial is and whether an effect of the tutorial can be measured at all. In order to systematically investigate the effect of the tutorial, two groups were necessary. The first was to investigate whether the addition of the tutorial makes a difference to the reduced version of the link-centric warning without tutorial.

This resulted in the fifth (intervention) group 7 - Link-Centric Warning without Tutorial: This group, similar to groups 5 and 6, received the link-centric warning with supplementary information for the decision about the examples. But it did not receive an introductory tutorial about the functionality and different risk levels before the decisions were made. Accordingly, the participant had to understand the entire information from the tooltips itself.

In addition, it was to be examined whether the video can also be a compensation for the tutorial. Although the specific functions of the link-centric warning were not explained, at least the attention for the URL and its structure was increased and could thus form further synergy effects in combination with the link-centric warning.

Consequently, the sixth and last (intervention) group was group 8 - Video + Link-Centric Warning without Tutorial: This group received the video at the beginning of the study, similar to groups 3, 4 and 6. Subsequently, however, no tutorial was presented and for the decision about the examples the participant received the link-centric warning and could only understand the risk level distinction of the link-centric warning on the basis of the information from the tooltips.

### 5.2.1 Study Design

The purpose of this subsection is to provide a more detailed description of the entire process of the study with all the necessary steps and details.

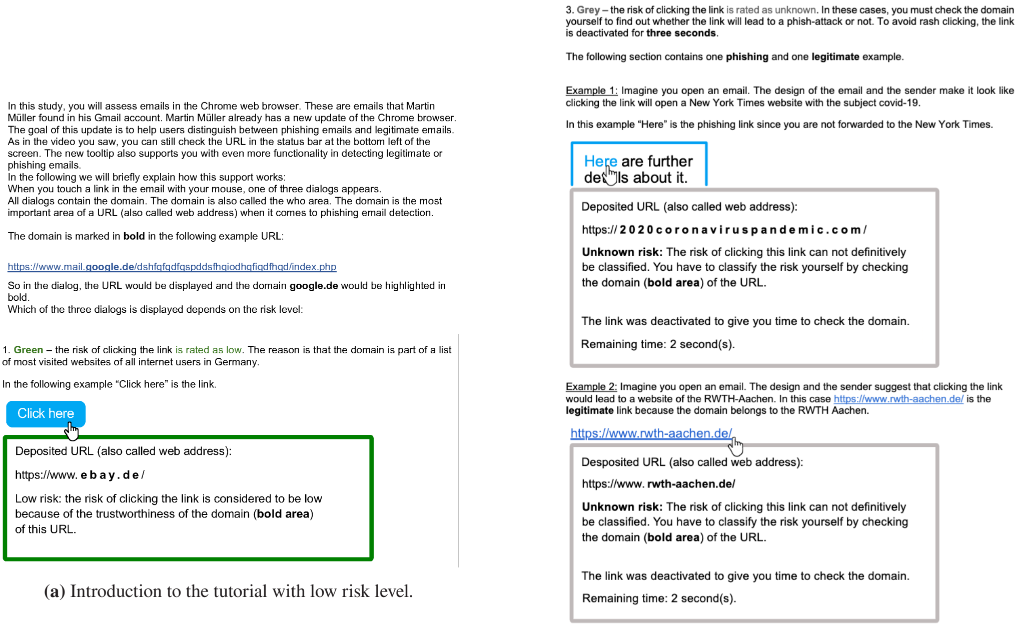
The study was a between-subjects design. This means that each group received only one form of intervention. The groups were designed to answer the five research questions. The study was conducted in the form of an online survey. Participants were first transferred from an online panel to the online survey platform Sosci Survey. More about recruitment of participants in Section 5.2.3. This platform has the advantage of being compliant with GDPR rules [94].

In the following, the entire process of the study is described in detail. First, the participants were forwarded to the survey platform via a link. Then, in the first step, the participants received a basic introduction to the procedure of the study and their associated rights. This included that the participants' data would be stored anonymously and would not offer any possibility of attribution to their person. They were also informed that participation could be terminated at any time without giving a reason. All the participant had to do was close the tab or the browser. The first page of the survey ended with the participant's consent. By clicking the "Continue" button, they gave their consent.

Page 2: The participants were now informed that a small exercise was waiting before the actual study started. This exercise was to help them to understand how to interact with the examples in the study environment. More about the study environment in the subsection Section 5.2.2. The interactive task was intended to ensure that participants did not exhibit different behavior than they would otherwise do to check email because of the static examples and non-given environment of a real email client. Likewise, they were given the information that this is important for the further task of classifying 20 emails into either legitimate or phishing.

Next, the five-minute NoPhish video (see Section 3.3.1) was shown to the groups 3, 4, 6 and 8. The video taught the most important tips and tricks for recognizing phishing messages, especially those with links. The participants were asked to start the video themselves. The participants were informed that an activated sound is necessary. Therefore, the participants were first asked to take the precautions and then play the video on their own. Afterwards, the participants still had to answer the question how they had watched the video. This was a closed question with different answer options ranging from (completely without pause to not at all). Since this question also included the risk of social desirability, four more attention questions were asked. To proceed with the study, participants had to answer three out of four questions correctly. This was to check that the participants who were later be surveyed in the video groups had shown at least a basic level of attention to the video.

In the next step, the groups 5 and 6 received the tutorial belonging to TORPEDO. The tutorial and TORPEDO were not introduced as such. Participants were told that they were using the Chrome web browser. During the course of the study, there was an update to the browser, whereby the participants were told that this update should help them to detect phishing emails. The update included a new tooltip with additional functionality, which is explained below. The tutorial took up an entire page (see Figure 5.2 and Figure 5.3). The tutorial was an adapted version from the original tutorial. All the information were presented on a single page. The content of the tutorial was reduced to the information necessary for the study and additional information such as the blue risk level were cut to not confuse participants.



In this study, you will assess emails in the Chrome web browser. These are emails that Martin Müller found in his Gmail account. Martin Müller already has a new update of the Chrome browser. The goal of this update is to help users distinguish between phishing emails and legitimate emails. As in the video you saw, you can still check the URL in the status bar at the bottom left of the screen. The new tooltip also supports you with even more functionality in detecting legitimate or phishing emails.
In the following we will briefly explain how this support works:
When you touch a link in the email with your mouse, one of three dialogs appears.
All dialogs contain the domain. The domain is also called the who area. The domain is the most important area of a URL (also called web address) when it comes to phishing email detection.

The domain is marked in **bold** in the following example URL:

https://www.mail.**google.de**/dshfqfqdfqspddsfhqiodhqfiqdfhqd/index.php

So in the dialog, the URL would be displayed and the domain **google.de** would be highlighted in bold.
Which of the three dialogs is displayed depends on the risk level:

1. **Green** – the risk of clicking the link is rated as low. The reason is that the domain is part of a list of most visited websites of all internet users in Germany.

In the following example "Click here" is the link.

Deposited URL (also called web address):

https://www.**ebay.de**/

Low risk: the risk of clicking the link is considered to be low because of the trustworthiness of the domain (**bold area**) of this URL.

**(a)** Introduction to the tutorial with low risk level.

3. **Grey** – the risk of clicking the link is rated as unknown. In these cases, you must check the domain yourself to find out whether the link will lead to a phish-attack or not. To avoid rash clicking, the link is deactivated for **three seconds**.

The following section contains one **phishing** and one **legitimate** example.

Example 1: Imagine you open an email. The design of the email and the sender make it look like clicking the link will open a New York Times website with the subject covid-19.

In this example "Here" is the phishing link since you are not forwarded to the New York Times.

Here are further details about it.

Deposited URL (also called web address):

https://**2020coronaviruspandemic.com**/

**Unknown risk:** The risk of clicking this link can not definitively be classified. You have to classify the risk yourself by checking the domain (**bold area**) of the URL.

The link was deactivated to give you time to check the domain.

Remaining time: 2 second(s).

Example 2: Imagine you open an email. The design and the sender suggest that clicking the link would lead to a website of the RWTH-Aachen. In this case https://www.rwth-aachen.de/ is the **legitimate** link because the domain belongs to the RWTH Aachen.

https://www.rwth-aachen.de/

Desposited URL (also called web address):

https://www. **rwth-aachen.de/**

**Unknown risk:** The risk of clicking this link can not definitively be classified. You have to classify the risk yourself by checking the domain (**bold area**) of the URL.

The link was deactivated to give you time to check the domain.

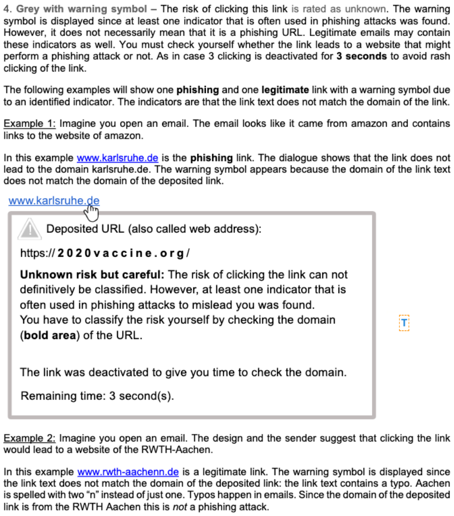
Remaining time: 2 second(s).

**(b)** Explanation of the unknown risk level with legitimate and phishing example.

**Figure 5.2:** Adapted version of the tutorial used for the study Part1 and Part2.

These groups were also given a series of control questions. A total of three questions were asked, all of which had to be answered correctly in order to continue participating in the study. Again, the goal was to ensure that the group later counted as including the tutorial actually focused attention to the tutorial so that a basic understanding was reached. Otherwise, the results of the group could be distorted and lead to wrong conclusions about the comparison between the groups with tutorial and the groups without tutorial.

Next, all groups were given a practice task to prepare for the later task of deciding between phishing and legitimate examples. Each exercise task was always already customized to the group, i.e. participants were given either a status bar, a tooltip or the link-centric warning in the example. First, it was explained to the participants that links can hide behind many different forms, e.g. buttons, texts or even images. The corresponding text is added to the graphic for each example. It was explained to the reader that one only has to move the mouse over the link and then the



4. **Grey with warning symbol** – The risk of clicking this link is rated as unknown. The warning symbol is displayed since at least one indicator that is often used in phishing attacks was found. However, it does not necessarily mean that it is a phishing URL. Legitimate emails may contain these indicators as well. You must check yourself whether the link leads to a website that might perform a phishing attack or not. As in case 3 clicking is deactivated for **3 seconds** to avoid rash clicking of the link.

The following examples will show one **phishing** and one **legitimate** link with a warning symbol due to an identified indicator. The indicators are that the link text does not match the domain of the link.

Example 1: Imagine you open an email. The email looks like it came from amazon and contains links to the website of amazon.

In this example www.karlsruhe.de is the **phishing** link. The dialogue shows that the link does not lead to the domain karlsruhe.de. The warning symbol appears because the domain of the link text does not match the domain of the deposited link.

www.karlsruhe.de

Deposited URL (also called web address):

https://**2020vaccine.org/**

**Unknown risk but careful:** The risk of clicking the link can not definitively be classified. However, at least one indicator that is often used in phishing attacks to mislead you was found.
You have to classify the risk yourself by checking the domain (**bold area**) of the URL.

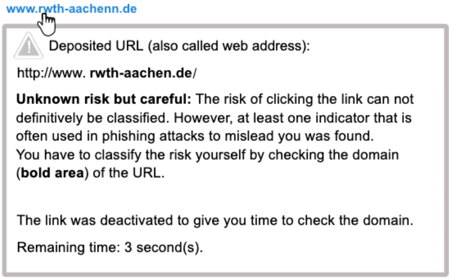
The link was deactivated to give you time to check the domain.

Remaining time: 3 second(s).

Example 2: Imagine you open an email. The design and the sender suggest that clicking the link would lead to a website of the RWTH-Aachen.

In this example www.rwth-aachenn.de is a legitimate link. The warning symbol is displayed since the link text does not match the domain of the deposited link: the link text contains a typo. Aachen is spelled with two “n” instead of just one. Typos happen in emails. Since the domain of the deposited link is from the RWTH Aachen this is *not* a phishing attack.

**(a)** Explanation for unknown risk with indicator level with phishing example.



www.rwth-aachenn.de

Deposited URL (also called web address):

http://www. **rwth-aachen.de/**

**Unknown risk but careful:** The risk of clicking the link can not definitively be classified. However, at least one indicator that is often used in phishing attacks to mislead you was found.
You have to classify the risk yourself by checking the domain (**bold area**) of the URL.

The link was deactivated to give you time to check the domain.

Remaining time: 3 second(s).

**(b)** Legitimate example of the unknown risk with indicator level.

**Figure 5.3:** Adapted version of the tutorial used for the study Part3 and Part4.

cursor changes into a hand. To this end, participants were told that the links within the study were disabled for demonstration purposes. Accordingly, redirection to dangerous pages were excluded. Further down the page, an example then followed that gave them an understanding of the task - which later had to be carried out several times for the study. The user received an example e-mail in the e-mail inbox web environment. The e-mail was addressed to Mr. Martin Müller with his Gmail account. This e-mail was to be checked for the number of existing links. The goal here was on the one hand that the participants learn how to interact with the e-mails in the context of the study. For this, they had to learn that the links were just like in a mail program with the cursor over the link showing the URL. They should also learn that there are many different forms of links and that they can hide behind images or even buttons, for example. Thus, all possibilities that had been used later in the study should already be known. The participants should not make a wrong decision because they were unfamiliar with the study environment. Next, participants were asked to indicate the number of links in the image. If the answer was correct, participants were redirected to the following page.

Then on the next page, all participants were again given an introductory scenario to assess the examples. The study participants were told that they would receive 16 examples in the following and that they should examine these e-mails. In doing so, they must determine whether an email is phish or legitimate. To minimize rejection based on their own experience, participants were asked to imagine a fictitious scenario. The description was as follows:

> In order not to directly declare the senders, service providers and programs you have no connection to in reality as fraudulent, please assume in the following that: * You are Martin Müller and have the email address: martin.mueller.77@gmail.com,.
>
> * You use all services used in this questionnaire (Amazon, Lufthansa, Google, LinkedIn, DHL, Netflix, GMX, PayPal, Microsoft, and Apple).
>
> Note: In this study, the links are disabled, so redirection is excluded.

Then came the main part of the study. Study participants were shown 16 examples of different emails. More on the structure of the emails in the subsection Section 5.2.2. One example was shown on each page. The examples were presented in a random order for each participant. On the page was then an example of the email with the question: this email is a: phishing email ... legitimate email. Due to technical restrictions, it was decided to use screenshots of the emails. In addition, the functionality was implemented on the platform that if the mouse was moved over a certain area of the page, a tooltip or status bar was displayed. What exactly was displayed depended on the respective group membership (see Section 5.2.1). After the participants had made a decision for an example and moved on to the next page, this decision could not be reversed by returning to the example. After participants had worked through all of the examples, participants in the intervention groups were still educated with TORPEDO. They were told that it was a fictitious update of Chrome. The function, on the other hand, already existed in the add-on TORPEDO and could have been downloaded via a link. On the next page, the socio-demographic questions were asked. These included questions about gender, age (with various gradations) and the highest professional education. On the last page, participants were informed about the end of the study and received the code for payment on the participant platform.

### 5.2.2 Study Material

In the following, the process for creating the examples is described in more detail. First, the basic design (sender, subject, content) of the e-mail examples with the copy of an authentic e-mail is discussed. Then, the adaptation of the emails for the study context is explained. This is about the authentic emails being converted into phishing emails using various phishing tricks.

In order to perform the sensitivity detection, it was necessary to use different examples. To create these examples, different approaches can be chosen. On the one hand, these examples can be built from scratch. This has the advantage that one can decide on every aspect of the email, e.g., what the sender address is, what the subject line is, how many graphics are included, and what the layout of the text is. The difficulty here is that while theoretically these emails can be designed to look legitimate except for the indicator to be manipulated (sender, URL or attachment), in practice, subscribers may unconsciously perceive some indicator that was deliberately not designed as phish as such and therefore reject the email. They may also come to the conclusion that the organization (if known to them) would not design or phrase such emails in such a way. Another option is to create a collection of emails that have already been sent by the organizations in exactly this form. In this way, it is still not possible to completely exclude unconscious wrong decisions based on other indicators. But at least it is not an influence introduced by the study methodology and thus even with such a "systematic" error the external validity of the study increases. This error would then accordingly also happen outside the study. The examples were all written in the German language. Admittedly, some of the organizations used are not originally from Germany and thus an English email would also have been authentic in a certain sense. Nevertheless, there are two reasons in particular for using a different language here. Firstly, the organizations in Germany mainly send German emails, so that all users are included. Secondly, the influence of the language on the decision should not be the focus. An English email could have led to rejection due to unfamiliarity with certain formulations.

The issue now is the selection of organizations for the context of the emails. To increase the likelihood that there is some familiarity with the emails in addition to the introductory scenario, a list of organizations from the most visited domains in Germany was chosen. This was to ensure that participants would at least know the name of the organization and its context and thus not make the decision out of complete ignorance. A query of the familiarity with the organization was omitted for various reasons. The participants should not be primed in advance on a possible distribution of examples and their organization, especially since two examples were used per organization (more on this later in this section). In principle, it would have been possible to list more organizations that were not used in order to disguise the duration. However, this could also lead to confusion if organizations do not appear later. Therefore, the added value of the influence of familiarity on the individual examples was seen as less important than the existing priming. Eight organizations were then selected from the list of most visited domains, which should be known to most Germans at least by name (Amazon, Lufthansa, Google, LinkedIn, DHL, Netflix, GMX, PayPal). The organizations should also come from different areas of life, e.g. shopping, mobility, communication, entertainment.

The study always used the same number of phishing and legitimate examples per participant, i.e. each participant saw eight phishing examples and eight legitimate examples at a rate of 50%. It is possible that this probably does not correspond to the ratio of phishing to legitimate messages in reality. However, there has also been too little study to date on what the actual correct ratio would be. It can be assumed that this ratio is also highly individual, e.g. how much one moves on the internet, how often one has already given one's contact details (email or cell phone number) and whether services one uses have already been the target of a data dump. And even in the case that the correct ratio would be known, study methodology or feasibility and external validity are at odds here. Research is still lacking to be able to say concretely what influence the total number of examples has, e.g. with respect to habituation but also training effects, on the recognition rate in the course of the study. But an excessive total number of examples will definitely have an influence. However, to be able to make a statement about different phishing tricks (more about phishing tricks), a certain number of examples is still needed. Therefore, it was decided to include 16 examples so that eight organizations and four phishing tricks can be used twice.

The examples all contained a so-called “call-to-action”, i.e. the readers were asked to make a decision in some way. Without “call-to-action” the user could read the emails, but there would be no need to make a decision between phishing or legitimate. Thus, when in doubt, the emails would simply remain undecided, as there would be no negative consequence. Accordingly, this call is necessary for the need for a decision to exist.

Hence, the base of the emails consisted of authentic emails. The next step was to create the phishing samples from a portion of the emails. Past studies had often examined phishing in a rather simple form [105, 106]. Even though this is similar to reality that phishing messages can be recognized already by the bad grammar. Thus, phishing messages are becoming more and more sophisticated [107, 108] and these indicators (grammar and content) can be replaced very efficiently and effectively by other strategies (manipulating the URL). Then, the only reliable indicator that remains is the one that applies to both the simple and complicated ones. According to this, users should always follow the behavioral advice and thoroughly examine the URL. However, even in the area of the URL, there are a variety of many ways phishers can modify the URL to convince the reader of its trustworthiness. These strategies are based on the fact that the reader of the URL has no or very little knowledge about the URL and its structure.

In the past, various ways to change the URL have been proposed in groups of phishing tricks (see Section 3.1). In this study, the similar phishing tricks to the studies from Chapter 4 had been used. This was to ensure that a wide range of ways to change the URL had been represented. Especially, also those based on the strengths, but also the weaknesses of the individual measures or tools. This is also in contrast to the individual studies of the interventions used (NoPhish Video and TORPEDO), each of which mapped only a portion of these phishing tricks in their evaluations. Based on this, the following strategies were used for the study:

- Non-brand related Domain
- Non-brand related Domain + Brand Outside
- Small Deviations in the Domain
- Special Link Manipulation

Next the full overview of all the examples used with information such as the phishing trick, sender, subject, TORPEDO risk level and URL is provided. See Table 5.1 for the phishing examples and Table 5.2 for the legitimate examples.

### 5.2.3 Recruitment & Ethical / Data Protection Considerations

This section discusses both the recruitment of participants and the further ethical considerations, at recruitment and during the course of the study.

Participants were recruited using a panel. In this process, a study is advertised and a suitable description is deposited. Only three criteria were defined as a reason for exclusion: 1) There must not have been any participation in a phishing study in the past. This was ensured with the help of a block list, which included all previous participants of the research group from phishing studies; 2) The participants must not be under 18 years of age. This was ensured via a feature of the panel. It is also a requirement of the panel, as only people at least 18 years old can be paid; 3) Participants had to be German speaking. This was also ensured via a feature of the panel.

To reduce self-selection bias, the study was not advertised as IT security or phishing in the panel's section. Accordingly, the description spoke only very abstractly about the context of email interface and user experience. Specifically, the description read, "Participate in a survey about user experience with email interfaces." Power analysis was used to calculate the sample size. Because the interventions have not been directly compared before, it was not possible to refer to previous studies and their results. Since there was also little other evidence on how large an effect there might be between the interventions and their combination, the reasoning of Cohen was followed [92]. He suggested that if the information is unclear or insufficient, at least a medium effect ($f = .25$) should be assumed. From previous studies, however, conditions for a one-way ANOVA (at least with adjustment of the data) were assumed to be present. The test power was set at the level of .95 with an alpha error of .05. The analysis resulted in a necessary sample size of 360 participants to be able to prove a medium effect with the values shown. To prevent the number of participants from falling below this limit due to the exclusion of participants, e.g. by the attention questions, a 20% buffer of participants was chosen. In total, this resulted in a targeted sample size of 430 participants. Consistent with the suggestions for ethical research, participants should be paid at least minimum wage. For several reasons, participants had to be paid the same. First, participants were randomly assigned to groups upon entering the survey. Accordingly, no assignment of participants could be made through the panel. Second, no different payment could be made at the panel depending on group membership. Therefore, all participants were paid as if they were participants in the longest group. Based on internal pretesting, a study duration of 30 minutes was assumed. The German minimum wage at the time of the study (December 2021) of

**Table 5.1:** Phishing Examples with TORPEDO Risk Levels.

| Name | Phishing Trick | More specific Phish | Sender | Subject | TORPEDO Risk Levels | Indicator | Tooltip URL |
|---|---|---|---|---|---|---|---|
| P1 | Non-Brand related Domain | Random | Amazon.de<promotion5@amazon.de> | Information about your Amazon account | grey | https://telefon.host745.com/hinzufuegen | |
| P2 | Non-Brand related Domain | IP | Lufthansa<online@booking.lufthansa.com> | Information about your upcoming flight | grey with indicator | https://87.147.12.250/buchungsänderung | |
| P3 | Non-Brand related Domain + Brand Outside | Subdomain | Google<no-reply@account.google.com> | Help us increase the security of your Google account | grey | https://www.google.com.megahoust.ru/sicherheitscheck | |
| P4 | Non-Brand related Domain + Brand Outside | Path | LinkedIn<messages-noreply@linkedin.com> | Martin, people are searching for your LinkedIn profile | grey | https://login.linkyzt.com/www.linkedin.com/profil | |
| P5 | Small Deviations in the Domain | Letter Swap | DHLPaket<noreply@dhl.de> | Your package will be delivered today. Book drop off location now ... | grey | https://account.dlh.com/zustellung | |
| P6 | Small Deviations in the Domain | Typo | Netflix<info@mailer.netflix.com> | New login to your account | grey | https://www.netfllx.com/neuerlogin | |
| P7 | Special Link Manipulation | Mismatch | GMXKundenmanagement<mailings@system.gmx.net> | Storage space for emails almost full | grey with indicator | https://premium.host547.ru/speichervoll3 | https://premium.gmx.de/speichervoll |
| P8 | Special Link Manipulation | Mismatch | PayPal<paypal@mail.paypal.de> | Upcoming changes to PayPal's terms and conditions | grey with indicator | https://www.hokpurt.ru/AGB | https://www.paypal.com/AGB |

**Table 5.2:** Legitimate Examples with TORPEDO Risk Levels.

| Name | Sender | Subject | TORPEDO Risk Levels | Indicator | Tooltip URL |
|---|---|---|---|---|---|
| L1 | `Amazon.de<order-update@amazon.de>` | Delivered: Your Amazon.de order with order no. 616-619884949494-6161684 | green | `https://packet.amazon.de/paketverfolgung` | |
| L2 | `Lufthansa<service@your.lufthansa-group.com>` | Lufthansa Booking | grey | `https://www.lufthansa.com/buchungsanzeige` | |
| L3 | `Google<no-reply@account.google.com>` | Security warning for martin.mueller.77@googlemail.com | green | `https://www.google.com/neuesgerät` | |
| L4 | `LinkedIn<messages-noreply@linkedin.com>` | Martin, a free course on strategic thinking - for a limited time only | grey | `https://video.linkedin.com/kurs` | |
| L5 | `DHLPaket<noreply@dhl.de>` | Your Zalando parcel is at the desired drop-off location. | grey | `https://mailing.dhl.de/wunschort` | |
| L6 | `Netflix<info@mailer.netflix.com>` | We're updating our prices - and here's why | green | `https://www.netflix.com/neuepreise` | |
| L7 | `GMXKundenmanagement<mailings@produkt.gmx.net>` | Confirmation of the GTC update for your GMX mailbox | green | `https://bestaetigung.gmx.de/AGB` | `https://brief.gmmx.net/AGB` |
| L8 | `PayPal<paypal@mail.paypal.de>` | Martin Mueller 5 euros for testing the PayPal app | grey with indicator | `https://www.paypal.com/gutschein` | `https://www.paypall.com/gutschein` |

9.82€ was then used for payment. This minimum wage was rounded up to 10€ and then resulted in a payment of 5€ per participant for the approx. 30 minutes.

Participants were already informed on the panel page that there would be attention questions during the course of the study and that the study could only be completed and payment made if the answers were correct. After the participants were redirected from the panel to the survey platform, they were first presented with a brief introduction and all the necessary information about the study. There, the essential information about the collection and use of data, but also about the storage and deletion of (personal) data was explained. Participants were informed that participation is completely voluntary and can be canceled at any time without giving any reason. After cancellation of participation, however, no payment can be made. It was explained that the data is analyzed and processed anonymously and does not allow any traceability through publication. IRB/ERB consent was obtained for the study.

In particular, for the groups that received the fictitious browser update (TORPEDO), a special debriefing was created at the end of the study. This was to prevent them from behaving unsafely in the future because they thought the update existed in their browser. At this point, it must be noted that participants who read the introduction but left the study during the course did not receive the debriefing. This is a clear limitation. But due to the anonymous survey these participants could not be contacted afterwards. Despite in such a case a debriefing would have been desirable, the lack of debriefing is not expected to cause these participants a strong disadvantage in their everyday life.

**Table 5.3:** Overview of the average rates for all eight groups in % for the correct answers for phishing and legitimate examples as well as overall. Also adding the values for the signal detection theory (SDT).

| Value | Status bar | Tooltip | NoPhish video + Status bar | NoPhish video + tooltip | TORPEDO warnings | NoPhish video + TORPEDO warnings | TORPEDO warnings without tutorial | NoPhish video + TORPEDO warnings without tutorial |
|---|---|---|---|---|---|---|---|---|
| Phish | 54.31 | 56.60 | 81.63 | 90.69 | 85.52 | 97.12 | 62.93 | 91.37 |
| Legitimate | 81.68 | 83.73 | 83.67 | 81.91 | 98.41 | 97.44 | 86.21 | 89.29 |
| Overall | 68.68 | 69.71 | 82.09 | 85.22 | 92.42 | 97.29 | 75.48 | 90.48 |
| SDT | | | | | | | | |
| Sensitivity | 1.01 | 1.13 | 1.87 | 2.12 | 2.51 | 2.92 | 1.42 | 2.36 |
| Criterion | 0.41 | 0.4 | 0.03 | -0.13 | 0.27 | 0.01 | 0.37 | -0.03 |

## 5.3 Results

This section describes the results of the study divided into the five research questions. First, the two research questions that served as prerequisites were, whether the video (RQ1) or the link-centric warning (RQ2) lead to a significantly better sensitivity directly following the respective intervention, as was shown in past studies. This is followed by the main research questions, whether there is an option to be preferred between the video or the link-centric warning (RQ3) or whether the combination of both interventions leads to further improvement (RQ4). Finally, the question, whether there is a difference of the interventions on the different phishing tricks is answered (RQ5).

A single one-way ANOVA was calculated to evaluate the research questions one through four. The individual results of these ANOVAs are reported in the sections of the respective research question. The randomly assigned group is the independent variable and the sensitivity or criterion is the dependent variable.

For the analysis of research question five, the descriptive values are reported first (see Table 5.3. Due to the low number of examples per phishing trick, a hypothesis test was not performed. Also, the hit rate (number of correct answers for the phishing examples divided by the total number of answers) was used here rather than the sensitivity, since the sensitivity cannot be applied to the individual phishing examples or tricks by definition. For the legitimate examples, the correct rejection rate (number of correct responses for the legitimate examples divided by total number of responses) was reported instead.

The one-way ANOVA for sensitivity yields a significant difference for sensitivity between groups with $p < .0001$, F(7.401) = 50.867, $\omega^2$ = .46. Following Cohen's guidelines (effect sizes above

.14 = a large effect) this corresponds to a large effect. A significant difference for the criterion is also found for the groups with $p < .0001$, F(7,401) = 16.062, $\omega^2$ = .20. This also corresponds to a large effect. In each case, the post-hoc test numbers were based on an adjustment for multiple testing using the Games-Howell correction.

### 5.3.1 RQ1 - Video Effect

This section is about the first research question on the effectiveness of the video. Based on the significant results for the comparison of the sensitivity between the groups, post-hoc tests were performed between the status quo with and without video for status bar and tooltip, respectively. A significant difference was shown for both status bar ($d' = 1.01$) versus video + status bar ($d' = 1.87$), and tooltip ($d = 1.13$) versus video + tooltip ($d' = 2.12$). Following the results, the hypothesis H1, that the video improves the sensitivity, could be confirmed.

There was also a significant difference for the criterion comparison. The post-hoc tests showed that again the status quo with and without video differed significantly from each other. Status bar reached a value of $C = .41$ and video + status bar a value of $C = .033$. Tooltip reached a value of $C = .4$ and video + tooltip a value of $C = -.126$. In both cases, the combination of a significant effect in combination with a value closer to 0 showed that there was an improvement of the criterion by the video.

### 5.3.2 RQ2 - Link-Centric Warning Effect

This section deals with the second research question on the effectiveness of link-centric warning. To test the effectiveness of the link-centric warning, the post-hoc tests of the status quo (status bar and tooltip) with the link-centric warning group were considered, respectively. For both comparisons of link-centric warning ($d' = 2.51$) with the status bar ($d' = 1.01$) and tooltip ($d' = 1.13$), significant differences emerged. Following the results, the hypothesis H2 that link-centric warning improves sensitivity could be confirmed.

No significant difference was shown for the criterion comparison. The groups were all at a very similar level with link-centric warning ($C = .269$), status bar ($C = .41$), and tooltip ($C = .4$).

### 5.3.3 RQ3 - Video vs. Link-Centric Warning

This section deals with the third research question on the effectiveness of the video and link-centric warning compared to each other. For this comparison, three groups were compared. This time, the respective status quo in combination with the video was compared with the link-centric warning. For the consideration of the sensitivity it could be stated that there was a significant difference. The link-centric warning group ($d' = 2.51$) differed significantly from the two video groups with status bar ($d' = 1.87$) and tooltip ($d' = 2.12$). Following the results, the hypothesis H3 about the difference could be confirmed and additionally it could be stated that the link-centric warning improved the sensitivity more than the video.

There was also a significant difference for the criterion comparison. The video + link-centric warning ($C = -.096$) and link-centric warning ($C = .258$) groups were significantly different. In contrast, there was no significant difference for the video + status bar ($C = .042$) and link-centric warning ($C = .258$) groups.

### 5.3.4 RQ4 - Combination of Video and Link-Centric Warning

This section deals with the fourth research question on the effectiveness of the combination of the video and link-centric warning compared to the individual interventions. For this comparison, the four following groups are compared: 1) link-centric warning; 2) video + status bar; 3) video + tooltip; and 4) video + link-centric warning. Examination of the post-hoc tests showed a significant difference between both link-centric warning ($d' = 2.51$) and video + link-centric warning ($d' = 2.92$), and between video + status bar ($d' = 1.87$) and video + link-centric warning ($d' = 2.92$). The same was true for the comparison between video + tooltip ($d' = 2.12$) and video + link-centric warning ($d' = 2.29$). Following the results, the hypothesis H4 about the difference between the combination of interventions compared to the individual interventions could be confirmed.

There was also a significant difference for the comparison of the criterion. Video + link-centric warning group ($C = .005$) and the link-centric warning group ($C = .269$). Both, the video + link-centric warning ($C = -.126$) and video + status bar groups, were not significantly different from the video + link-centric warning group with respect to the criterion.

### 5.3.5 RQ5 - Phishing Tricks

This section deals with the fifth and final research question on the effectiveness of the different interventions in relation to the phishing tricks, phishing examples and legitimate examples, respectively. The full overview of all results is presented in three tables for the phishing tricks (see Table 5.4), phishing examples (see Table 5.5), and for the legitimate examples (see Table 5.6). Three out of four phishing tricks achieved very similar scores for status quo (status bar and

**Table 5.4:** Overview of the percentage of correct answers for phishing tricks. Grp1 = status bar, Grp2 = tooltip, Grp3 = video + status bar, Grp4 = video + tooltip, Grp5 = link-centric warning, Grp6 = video + link-centric warning, Grp7 = link-centric warning without tutorial, Grp8 = video + link-centric warning without tutorial.

| Phishing Trick | Grp1 | Grp2 | Grp3 | Grp4 | Grp5 | Grp6 | Grp7 | Grp8 |
|---|---|---|---|---|---|---|---|---|
| Non-brand related Domain | 65.52 | 66.98 | 89.8 | 94.68 | 93.65 | 96.15 | 71.55 | 94.05 |
| Non-brand related Domain + Brand Outside | 62.93 | 61.32 | 86.73 | 92.55 | 96.03 | 100.00 | 71.55 | 96.43 |
| Small Deviations in the Domain | 19.83 | 25.47 | 55.1 | 75.53 | 53.97 | 92.31 | 32.76 | 75.00 |
| Special Link Manipulation | 68.97 | 72.64 | 94.9 | 100 | 98.41 | 100.00 | 75.86 | 100.00 |

tooltip) in the range of 60% to 70% (Non-brand related Domain, Non-brand related Domain + Brand Outside and Special Link Manipulation). In contrast, "Small Deviations in the Domain" only achieved values of just under 20% to 25% here. Now looking at the effect of the video, for both status quo over all phishing tricks there was an increase of about 25% to even 50% ("Small Deviations in the Domain" from 25.47% to 75.53%). "Small Deviations in the Domain", however, remained at 55.10% for the video group with the status bar. The other three phishing tricks were in the upper 80% range for the video groups up to even 100%. When looking at the results for the link-centric warning, there were very similar values as for the video groups. "Small Deviations in the Domain" only reached a value of about 50% and the other three phishing tricks in the range above 90%. If the video and the link-centric warning were combined to one intervention, "Small Deviations in the Domain" improved to over 90% correct answers. Now the influence of the tutorial on the effectiveness of the link-centric warning was examined. The link-centric warning without tutorial and without additional video achieved only slightly better values than the status quo in the low 70% range for "Non-brand related Domain", "Non-brand related Domain + Brand Outside" and "Special Link Manipulation" and only about 30% for "Small Deviations in the Domain". The combination of video and link-centric warning without tutorial resulted in relatively similar values as the status quo with tooltip and video with 75% for "Small Deviations in the Domain" and 90% to 100% for the other three phishing tricks.

**Phishing examples** Upon examination, the answers for the phishing examples range from very few correct answers to a majority of correct messages. But also for the individual examples, the status quo (status bar and tooltip) showed very similar tendencies, with slightly higher values for the tooltip. Some phishing examples (random) already reached values of around 80% correct answers. Then followed for the two "Mismatch" examples in the range around 70% and the two "Non-brand related Domain + Brand Outside" with around 60%. The IP is then far behind compared to the other "Non-brand related Domain" example at around 50%. And both examples for "Small Deviations in the Domain" reached a value below 30%. This order of examples was almost the same for all interventions. However, it was still noticeable that both the path versus the subdomain and the typo versus the letter swap perform worse for some interventions (link-centric warning, link-centric warning without tutorial and video + link-centric warning without tutorial).

**Table 5.5:** Overview of the percentage of correct answers for the phishing examples. Grp1 = status bar, Grp2 = tooltip, Grp3 = video + status bar, Grp4 = video + tooltip, Grp5 = link-centric warning, Grp6 = video + link-centric warning, Grp7 = link-centric warning without tutorial, Grp8 = video + link-centric warning without tutorial.

| Name | Phishing Trick | Grp1 | Grp2 | Grp3 | Grp4 | Grp5 | Grp6 | Grp7 | Grp8 |
|---|---|---|---|---|---|---|---|---|---|
| P1 | Non-brand related Domain | 86.21 | 81.13 | 93.88 | 100.00 | 100.00 | 100.00 | 87.93 | 100.00 |
| P2 | Non-brand related Domain | 44.83 | 52.83 | 85.71 | 89.36 | 87.30 | 92.31 | 55.17 | 88.10 |
| P3 | Non-brand related Domain + Brand Outside | 56.90 | 60.38 | 93.88 | 100.00 | 98.41 | 100.00 | 74.14 | 100.00 |
| P4 | Non-brand related Domain + Brand Outside | 68.97 | 62.26 | 79.59 | 85.11 | 93.65 | 100.00 | 68.97 | 92.86 |
| P5 | Small Deviations in the Domain | 20.69 | 24.53 | 59.18 | 74.47 | 69.84 | 94.87 | 37.93 | 80.95 |
| P6 | Small Deviations in the Domain | 18.97 | 26.42 | 51.02 | 76.60 | 38.10 | 89.74 | 27.59 | 69.05 |
| P7 | Special Link Manipulation | 67.24 | 71.70 | 93.88 | 100.00 | 96.83 | 100.00 | 79.31 | 100.00 |
| P8 | Special Link Manipulation | 70.69 | 73.58 | 95.92 | 100.00 | 100.00 | 100.00 | 72.41 | 100.00 |

**Legitimate examples** These examples can only be divided into "groups" similar to phishing tricks to a very limited extent. There were only two groups (the normal legitimate and the mismatch legitimate). In general, as expected, the legitimate examples were on a higher level than the phishing examples (between around 80% and almost 100%). The "Status Quo" (status bar or tooltip) and the groups with the video showed quite similar results here around 80%. In contrast, the values for the link-centric warning increased, whereby the groups without tutorial were below the groups with tutorial (regardless of whether the video was also watched). The link-centric warning without tutorial group represented the fourth best value with about 85% followed by the video + link-centric warning without tutorial group with 89%. The two link-centric warning groups with tutorial (with and without video) were at the top with around 97% each. It is noticeable that the example L1 from Amazon performed worst by far among the the "normal" legitimates (at least for the not link-centric warning groups). The video groups only achieved an average of around 60%, even worse than the status quo at 75%. In contrast, the link-centric warning groups did not show a big difference to the other examples.

In addition to the individual phishing tricks and legitimate examples, another facet of the examples was particularly interesting for the groups with link-centric warning. Each example, whether phishing or legitimate, had a unique risk level distinction based on frame appearance and tooltip information text. In the following, the specifics of this other facet of the examples are discussed. The assignment of the individual examples to the respective risk level can be found in Table 5.1 for the phishing and Table 5.2 for the legitimate ones. Looking again at the examples focusing on the risk level distinction, it was noticeable that for the fraudulent examples there was not much difference between the "Grey" and "Grey with indicator" categories. In fact, the examples with indicator tended to score higher than the examples without indicator in each group, averaging 8% higher (from 0.5% to about 15%).

In contrast, the legitimate examples showed different results for at least some of the groups. The groups link-centric warning and video + link-centric warning consistently achieved high values in the range of over 90%. Therefore, no great distinction could be found here. In the groups without tutorial, on the other hand, one could see clearer differences. Here, the examples with indicator performed on average between 17% and 21% worse than the examples without indicator.

**Table 5.6:** Overview of the percentage of correct answers for the legitimate examples. Grp1 = status bar, Grp2 = tooltip, Grp3 = video + status bar, Grp4 = video + tooltip, Grp5 = link-centric warning, Grp6 = video + link-centric warning, Grp7 = link-centric warning without tutorial, Grp8 = video + link-centric warning without tutorial. In contrast to the phishing examples, there is no phishing trick included in the legitimate examples, so the column is not present.

| Name | Grp1 | Grp2 | Grp3 | Grp4 | Grp5 | Grp6 | Grp7 | Grp8 |
|---|---|---|---|---|---|---|---|---|
| L1 | 74.14 | 77.36 | 61.22 | 61.70 | 98.41 | 97.44 | 86.21 | 80.95 |
| L2 | 82.76 | 84.91 | 95.92 | 95.74 | 100.00 | 94.87 | 82.76 | 92.86 |
| L3 | 94.83 | 94.34 | 100.00 | 93.62 | 100.00 | 100.00 | 93.10 | 100.00 |
| L4 | 93.10 | 84.91 | 85.71 | 93.62 | 98.41 | 97.44 | 89.66 | 97.62 |
| L5 | 87.93 | 90.57 | 81.63 | 89.36 | 100.00 | 100.00 | 94.83 | 95.24 |
| L6 | 94.83 | 86.79 | 100.00 | 93.62 | 100.00 | 100.00 | 94.83 | 100.00 |
| L7 | 62.07 | 71.70 | 71.43 | 63.83 | 95.24 | 94.87 | 79.31 | 85.71 |
| L8 | 63.79 | 79.25 | 73.47 | 63.83 | 95.24 | 94.87 | 68.97 | 61.90 |

## 5.4 Discussion Part II

This section presents the discussion of the study results and limitations. First, the results for the main research questions are discussed and connected with the related work. From there, proposals for future improvement of the usage of phishing awareness measures and phishing tool support are derived and discussed.

### 5.4.1 Research Questions

First, this study provides further insight into the effectiveness of the video and link-centric warning. By considering sensitivity as the ability to distinguish between legitimate and phishing examples, both interventions showed significant improvements over the current status quo (RQ1 and RQ2). By comparing the two interventions with each other (RQ3), the link-centric warning performed better than the video. Thus, if a decision has to be made between the two, a clear tendency towards link-centric warning can be given. However, it has to be noted that the link-centric warning only achieved a rate close to the guess probability when it came to the phishing trick “Small Deviations in the Domain”. In this case, the video with the tooltip at least performed better with about 75%.

The current status quo was shown to be in particular need of improvement or insufficient to achieve a sufficient rate even in this “best case” scenario. Without any intervention, users correctly identified only about half of all phishing examples, and even for legitimate examples, they made the wrong decision about 20% of the time. Especially the “Small Deviations in the Domain” posed such a challenge to the users that even less than 25% of the examples were correctly recognized. So one needs to assume that the users with the status quo have almost no chance to recognize these tricks. Adding the video as an intervention, then both status quo groups improved on the phishing tricks.

The fourth research questions, was whether the combination of the two interventions adds value over the respective intervention. It can be clearly stated that in relation to each of the factors considered (sensitivity, criterion or individual phishing tricks) the combination provided an extraordinary added value. In this “best-case” scenario, the combination achieved almost the optimum in each area. If one still wants to look for the existing potential here, then this possibly lies in an even more optimized presentation of the domain and clear explanation of the IP in the video. The optimized representation of the domain could take different forms, e.g. it could be evaluated in further experiments whether a change of the distance between the letters leads to a still better recognition of the “Small Deviations in the Domain”. Especially to address the typo, one could also consider whether such duplication of letters should be highlighted again with a color. If you look at another aspect of the video that could be improved, it seems to be the IP address. Here, apparently the complete “other-ness” of the URL leads to the fact that the decision is not always correct also with the combination of video and link-centric warning. The IP address used is the university’s domain and thus could not have been known to the participants or perceived as matching the example, so this influence can be ruled out for the incorrect decision. Thus, it must be made clearer that if the destination of the IP is not known, fraudulent intent should be assumed. None of the organizations used would use such an IP address for a website when writing to their customers.

An additional focus that was part of RQ3 and RQ4 was about clarifying the effect of the tutorial. Both with and without video, the tutorial had a significant effect on the effectiveness of the link-centric warning. The groups with tutorial performed better than the groups without tutorial, and the link-centric warning without video performed even worse than the video. Thus, there is the clear recommendation that the link-centric warning should indeed only be used in combination with the tutorial. This is in line with the positive feedback the link-centric warning has received in the add-on store, where it is already available for the public (currently 4.7 of 5 stars on the 14th of January 2024). At this point, it would be worth discussing why some phishing tricks were not sufficiently recognized despite the link-centric warning and despite the introduction of the individual functions of the link-centric warning. Maybe the tutorial has to explain the function of the spaces between the letters as a support for the recognition of phishing tricks like the “Small

Deviations in the Domain". Simply including the function with the reference to the domain that appears in the tutorial does not seem to be enough to ensure that users are sufficiently aware of the need to read the domain carefully in the tooltip. Eye-tracking studies could be used in the future to investigate whether this hint in the tooltip is not perceived, meaning that attention does not linger long enough on this area. Another possibility would be that it is perceived, but the text was not formulated appropriately for every user. This additional future investigation could be done in a qualitative follow-up survey in the form of an interview following the eye-tracking study.

With regard to phishing tricks, it can be stated that there was a very clear trend. "Small Deviations in the Domain" is the phishing trick that was most poorly detected across all groups. As already mentioned earlier the phishing trick "Small Deviations in the Domain" makes use of phenomenons such as "word recognition" or "lexical processing". These phenomenons describe that words are not read letter by letter, but rather identify words by their shape and letter order [63, 64]. Therefore it was not surprising that the "Small Deviations in the Domain" was the most difficult trick to detect. Still the results made clear that there is a need for further investigation into how participants can be supported or taught even better to recognize this trick more reliably. Interestingly, no group achieved a perfect score for the comparatively "simple" phishing trick "Non-brand related Domain". This is mainly due to the IP example, which did not achieve a perfect value in any group. A possible reason could be that IP addresses are so far from what people relate with URLs. Especially as IP also went beyond the scope of the explanation of the URL with the domain being the second to last part of the URL before the third slash. Such a special case could lead to confusion and therefore not following the advice given. Due to the differences identified, examining the various phishing tricks even more closely in the future is fundamentally essential. A more extensive analysis of various phishing tricks can help to optimize the respective measures and tools, even more specifically for existing weaknesses, to ensure the best possible effectiveness.

### 5.4.2 Future Work

As a clear result of Part II, combining technical support and an awareness measure created synergies beyond the individual measure. The combination even reached a level of sensitivity previous measures did not achieve and reached a near-optimal value for sensitivity ($d' = 2.92$). However, this also clearly showed that the tool support is preferable to the awareness measure in the present study. It achieved similar values for phishing detection as the status quo (status bar or tooltip) with the video and reached an almost optimal value for legitimate examples. Only in the area of "Small Deviations in the Domain" this group showed slight weaknesses, but the combination with the video almost entirely offset these. This phishing trick should be clearly mentioned when looking for a starting point to improve this combination. Both examples, which can be assigned to “Small Deviations in the Domain”, were not detected with 100%. Future research could thus start with the highlight of the URL and the spacing. Research questions for the future could be whether modifying the highlight with different colors for specific letters or enlarging the spacing can help to recognize this trick even better. At this point, it has to be investigated when the readability is affected so much that it also causes “costs” for the user that outweigh the benefits.

### 5.4.3 Limitations

This section discusses both the methodological limitations of the study and the limitations of the results.

Regarding the methodological limitations, it must be clearly stated that the study dealt with so-called “best case” scenarios. The primary task of the participants was to distinguish between phishing and legitimate examples. As a result, while the internal validity of the various groups is high, the external validity can still be improved. For example, the introductory task promotes internal validity because each group has at least a basic knowledge of how to use the emails, even those without a tutorial or video. Therefore, the effect on all groups is comparable and should not affect the difference between the groups. However, the task does affect the comparability to reality and therefore lowers the external validity because in reality there is no such introduction before using an email client or web browser. This may be a form of priming, as participants focus even more on the URL for evaluation. Furthermore, this focus is important for determining the effectiveness of interventions that rely heavily on the URL as an indicator. However, neither the task nor the study design was such that participants were forced to make their decision based on the URL. In addition, it is already known from previous studies that even when URLs are checked, people have difficulties reading and interpreting them correctly [109]. Therefore, the task was necessary to make the comparison viable, but in the future it should be investigated if the results can be transferred to everyday life without such a task.

The study environment was adapted as closely as possible to a real web browser email environment without compromising the important variables of the study. Compared to the studies in Part 1, the study environment was again adapted so that static examples were no longer used where the participant already had to move the mouse over the cursor. Instead, the participant was required to interact in a manner similar to an email client or web browser. Also, the entire environment was used, not just the email itself, creating a more immersive experience for the participant. However, the scenario represented only one email environment and there are many different web browser email environments or even email clients with different interfaces. It was therefore decided to use the most common environment with Gmail [110]. Nevertheless, familiarity with this or other environments may have an impact on the study. Participants were also asked if they were already familiar with the environment. However, familiarity could not be tested as a confounding variable due to the highly unequal distribution within the groups. In most groups, the percentage of those who either currently had an account or did have an account was in the range of 10% to 20%. Only the status bar group (below 7%) and video + link-centric warning without tutorial (24%) had rates slightly outside this range. Overall, the distribution of familiarity was very similar across all groups. In addition, the groups were very uneven in size, so that a comparison did not seem very revealing, e.g., the ratio of eight to 41 participants for the Video + Status bar group. In the future, a more concrete and systematic study needs to be done to see if familiarity really does have an effect.

The results, even if they reached the optimum within the scope of the study (see Section 5.3.4), are not necessarily directly transferable to reality. Especially since the results of a previous small field study [75] were also based on an older version of the link-centric warning. Therefore, a larger study based on the results of this study should consolidate the results in an even more realistic environment. Considering the results, one would probably test the groups video + tooltip, link-centric warning, and video + link-centric warning against each other. One could discuss whether other groups might perform better in reality due to unknown factors than they did in the present study. The very systematic evaluation, the clear and comprehensible results raise the question of which group this could apply to. Actually, there is only the video + Status bar group, which is a complete alternative to the proposed groups and would not be represented in such a study design. It is possible that the chosen presentation of the status bar had an influence, or that the familiarity with one's own presentation in the status bar has a greater influence for this group. Omitting the tutorial, with its very clear results on effectiveness, is not recommended at this point. Since the link-centric warning already exists and is used by the users, feedback could be obtained in this context that the tutorial also subjectively provides a great added value.

# 6 Overall Discussion and Outlook

This chapter discusses the most important aspects of the thesis (Section 6.1) and concludes with a few final remarks (Section 6.2).

## 6.1 Thesis Discussion

As part of this thesis, progress was made in research to help people in discriminating between phishing and legitimate emails. In the following, the results, which can be compared across the two parts, are discussed. On the one hand, these are the different phishing tricks, which were evaluated across the studies in both Part I and Part II. On the other hand, the video as a phishing awareness measure represents an overlap, which was tested in Part I as an awareness measure and refresher in the long run and evaluated in Part II as a phishing awareness measure against and in combination with tool support.

By considering all the studies from Part I and Part II regarding phishing tricks, a clear trend stands out. “Small Deviations in the Domain” represented the most difficult phishing trick to detect in most cases. Regardless of which measure was used, in most studies this trick scored very low both up front and over time. In the case of e-learning and workshop, even in the range of four to five months only a value around the guess probability of 50% was achieved. Having received two measures (workshop and six month later the interactive email example) resulted in the best values from both the main measures and the main measures plus refresher. Still people with both the workshop and the interactive email example achieved only about 70% correct answers twelve month later. Only the combination of a measure (video) and tool support (link-centric warning) could have achieved a sufficient result of just over 90%. In this case, the question remains, whether these values persist over a longer period of time. For other phishing tricks like “Non-brand related Domain” and “Non-brand related Domain + Brand Outside” there was little difference between both parts. With the first phishing trick being easier to spot independent from the main or refresher measure and the second phishing trick being on a lower level around 80% after most of the measures. Accordingly, future measures should focus on “Small Deviations in the Domain” without neglecting the other tricks too much.

Over the course of three studies, the video has proven to be effective in increasing the sensitivity of those that watched it, by achieving a sensitivity of 2.23 (as a main measure in Section 4.1.3), 2.12 (as a main measure with a tooltip in Section 5.3.2) and 1.8 (as a refresher in Section 4.3.3.2). Comparing the video to other first time awareness measures it ranged in the middle of the e-learning ($d' = 2.66$) and the workshop ($d' = 2.13$) directly after the measures were conducted. Considering the time required to carry out the measure in relation to the results of the sensitivity achieved, the efficiency is very good. The other two measures require several hours, whereas the video only takes five minutes. If the results from Part I and Part II are combined with new research [34] on the effectiveness of the video, there is still potential for improvement for the phishing trick “Small Deviations in the Domain”, especially in the long-term. The video could focus even more on the surprise effect of the simplicity of reading over such changes to the domain, similar to workshop or e-learning, and show this type of trick less obviously. Another alternative would be to combine the video with at least a short exercise for the phishing tricks at the end. Extending the video with a short exercise would still set the video apart from the other measures in terms of length, but could help to anchor what has been learned even more deeply and make it more memorable. So with different samples and over different points in time, the video performance varied, but always lead to a significant improvement in the sensitivity of the users and showed very promising results especially for the short amount of time needed for an awareness measure.

## 6.2 Final Remarks

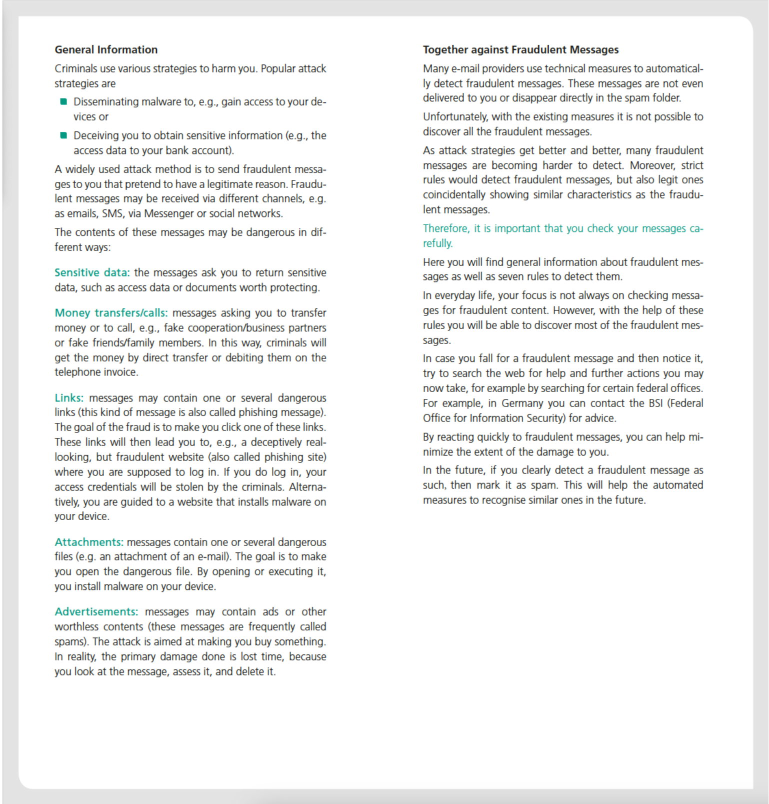
Phishing as an issue will probably be with society for many years to come, both as individuals and organizations. Any context in which messages are sent without sufficient verification of the recipient offers attackers the opportunity to trick people into unwanted behavior, be it entering sensitive data or executing files. Accordingly, research about phishing continues to be of particular relevance. It is also important to bring research even closer to people and to analyze their strengths and weaknesses in order to develop targeted measures. This includes both the development of awareness to differentiate between phishing and legitimate messages, as well as support through tools that can be integrated even more into everyday life. Previous studies that evaluate phishing training conducted pre-post-test comparisons and short follow-up measurements. However, in reality the topic of sustainability and refreshing awareness is relevant, which is why this work has a high practical relevance. By investigating different formats and length of phishing awareness measures, various points in time for refreshment and discriminating between different phishing tricks, this thesis has made an important contribution to the long-term impact of phishing awareness measures. By directly comparing a phishing awareness video and a link-centric warning, it was possible to show that instead of always looking at comparisons, there should be a greater focus on the combination of interventions. Such combinations could provide people with even better protection both in the context of the study and in everyday life through targeted support just in time and place of the decision situation.

A further advantage of this work is that the sensitivity and criterion from the signal detection theory were used throughout all the studies conducted. Even though sensitivity is the most important measure as it describes the ability to distinguish between phishing and legitimate examples, both are needed to paint a full picture. Especially because the improved sensitivity should not come at the expense of a deterioration of the criterion. Such a deterioration could mean that people’s false alarm rate increases as they suspect phishing more often, regardless of whether it is a phish or not. This methodological focus underlines the effectiveness of the interventions investigated in the studies and is an important unique selling point of this work. Overall, this thesis provides a systematic overview of the effectiveness of different phishing awareness measures, tools, and suitable refreshers and offers key insights into appropriate points in time to implement best-practice-recommendations for individuals and organizations to best protect against phishing attacks.

# A Appendix

## A.1 All pages of the leaflet

[1]



**General Information**

Criminals use various strategies to harm you. Popular attack strategies are

- Disseminating malware to, e.g., gain access to your devices or
- Deceiving you to obtain sensitive information (e.g., the access data to your bank account).

A widely used attack method is to send fraudulent messages to you that pretend to have a legitimate reason. Fraudulent messages may be received via different channels, e.g. as emails, SMS, via Messenger or social networks.

The contents of these messages may be dangerous in different ways:

Sensitive data: the messages ask you to return sensitive data, such as access data or documents worth protecting.

Money transfers/calls: messages asking you to transfer money or to call, e.g., fake cooperation/business partners or fake friends/family members. In this way, criminals will get the money by direct transfer or debiting them on the telephone invoice.

Links: messages may contain one or several dangerous links (this kind of message is also called phishing message). The goal of the fraud is to make you click one of these links. These links will then lead you to, e.g., a deceptively real-looking, but fraudulent website (also called phishing site) where you are supposed to log in. If you do log in, your access credentials will be stolen by the criminals. Alternatively, you are guided to a website that installs malware on your device.

Attachments: messages contain one or several dangerous files (e.g. an attachment of an e-mail). The goal is to make you open the dangerous file. By opening or executing it, you install malware on your device.

Advertisements: messages may contain ads or other worthless contents (these messages are frequently called spams). The attack is aimed at making you buy something. In reality, the primary damage done is lost time, because you look at the message, assess it, and delete it.

**Together against Fraudulent Messages**

Many e-mail providers use technical measures to automatically detect fraudulent messages. These messages are not even delivered to you or disappear directly in the spam folder.

Unfortunately, with the existing measures it is not possible to discover all the fraudulent messages.

As attack strategies get better and better, many fraudulent messages are becoming harder to detect. Moreover, strict rules would detect fraudulent messages, but also legit ones coincidentally showing similar characteristics as the fraudulent messages.

Therefore, it is important that you check your messages carefully.

Here you will find general information about fraudulent messages as well as seven rules to detect them.

In everyday life, your focus is not always on checking messages for fraudulent content. However, with the help of these rules you will be able to discover most of the fraudulent messages.

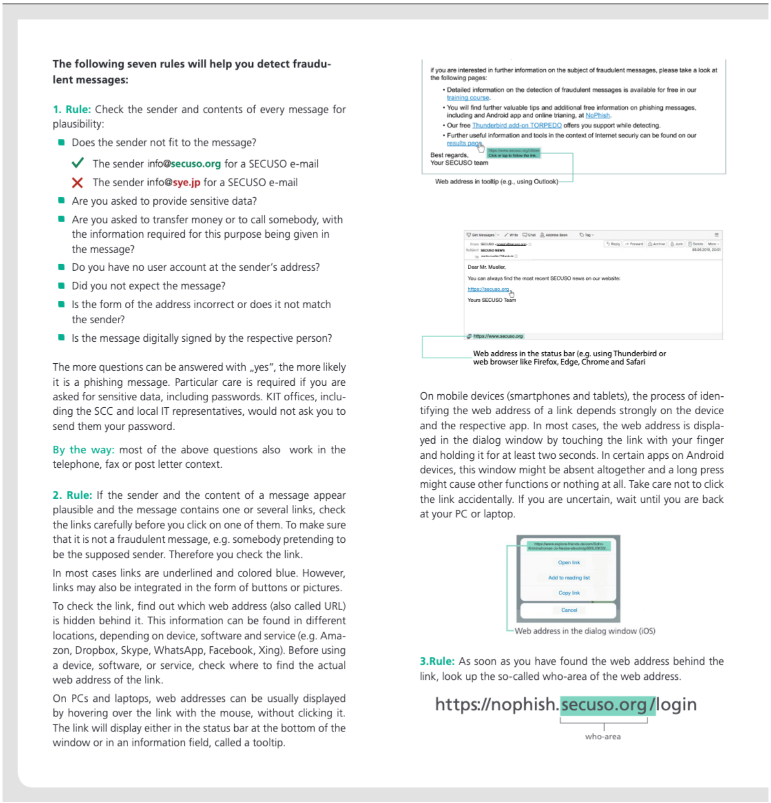
In case you fall for a fraudulent message and then notice it, try to search the web for help and further actions you may now take, for example by searching for certain federal offices. For example, in Germany you can contact the BSI (Federal Office for Information Security) for advice.

By reacting quickly to fraudulent messages, you can help minimize the extent of the damage to you.

In the future, if you clearly detect a fraudulent message as such, then mark it as spam. This will help the automated measures to recognise similar ones in the future.

**Figure A.1:** First part of the front pages of the leaflet from the original measure.

[1] current version available at `https://secuso.aifb.kit.edu/downloads/Flyer/NoPhish_betr.Nachrichten/Englisch/KIT_Faltblatt_BN_EN_2022v03.pdf`



**The following seven rules will help you detect fraudulent messages:**

**1. Rule:** Check the sender and contents of every message for plausibility:

- Does the sender not fit to the message?
  - ✓ The sender info@**secuso.org** for a SECUSO e-mail
  - ✗ The sender info@**sye.jp** for a SECUSO e-mail
- Are you asked to provide sensitive data?
- Are you asked to transfer money or to call somebody, with the information required for this purpose being given in the message?
- Do you have no user account at the sender's address?
- Did you not expect the message?
- Is the form of the address incorrect or does it not match the sender?
- Is the message digitally signed by the respective person?

The more questions can be answered with „yes", the more likely it is a phishing message. Particular care is required if you are asked for sensitive data, including passwords. KIT offices, including the SCC and local IT representatives, would not ask you to send them your password.

**By the way:** most of the above questions also work in the telephone, fax or post letter context.

**2. Rule:** If the sender and the content of a message appear plausible and the message contains one or several links, check the links carefully before you click on one of them. To make sure that it is not a fraudulent message, e.g. somebody pretending to be the supposed sender. Therefore you check the link.

In most cases links are underlined and colored blue. However, links may also be integrated in the form of buttons or pictures.

To check the link, find out which web address (also called URL) is hidden behind it. This information can be found in different locations, depending on device, software and service (e.g. Amazon, Dropbox, Skype, WhatsApp, Facebook, Xing). Before using a device, software, or service, check where to find the actual web address of the link.

On PCs and laptops, web addresses can be usually displayed by hovering over the link with the mouse, without clicking it. The link will display either in the status bar at the bottom of the window or in an information field, called a tooltip.

if you are interested in further information on the subject of fraudulent messages, please take a look at the following pages:

- Detailed information on the detection of fraudulent messages is available for free in our training course.
- You will find further valuable tips and additional free information on phishing messages, including and Android app and online trianing, at NoPhish.
- Our free Thunderbird add-on TORPEDO offers you support while detecting.
- Further useful information and tools in the context of Internet securiy can be found on our results page.

Best regards,
Your SECUSO team

https://www.secuso.org/citizen
Click or tap to follow the link.

Web address in tooltip (e.g., using Outlook)

SECUSO NEWS

06.06.2019, 20:51

Dear Mr. Mueller,

You can always find the most recent SECUSO news on our website:

https://secuso.org

Yours SECUSO Team

Web address in the status bar (e.g. using Thunderbird or web browser like Firefox, Edge, Chrome and Safari

On mobile devices (smartphones and tablets), the process of identifying the web address of a link depends strongly on the device and the respective app. In most cases, the web address is displayed in the dialog window by touching the link with your finger and holding it for at least two seconds. In certain apps on Android devices, this window might be absent altogether and a long press might cause other functions or nothing at all. Take care not to click the link accidentally. If you are uncertain, wait until you are back at your PC or laptop.

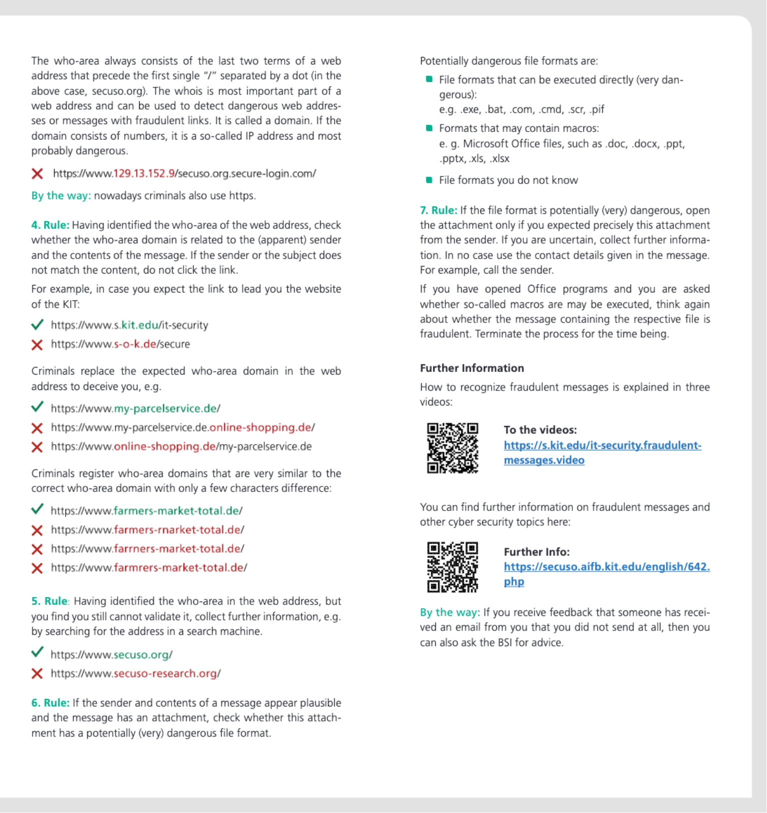
https://www.explore-friends.de/com/Sohn-Kriminalroman-Jo-Nesbe-ebook/dp/B00LI0K53/...
Open link
Add to reading list
Copy link
Cancel

Web address in the dialog window (iOS)

**3.Rule:** As soon as you have found the web address behind the link, look up the so-called who-area of the web address.

https://nophish.secuso.org/login

**Figure A.2:** First part of the back pages of the leaflet from the original measure.



The who-area always consists of the last two terms of a web address that precede the first single "/" separated by a dot (in the above case, secuso.org). The whois is most important part of a web address and can be used to detect dangerous web addresses or messages with fraudulent links. It is called a domain. If the domain consists of numbers, it is a so-called IP address and most probably dangerous.

✗ https://www.129.13.152.9/secuso.org.secure-login.com/

By the way: nowadays criminals also use https.

**4. Rule:** Having identified the who-area of the web address, check whether the who-area domain is related to the (apparent) sender and the contents of the message. If the sender or the subject does not match the content, do not click the link.

For example, in case you expect the link to lead you the website of the KIT:

✓ https://www.s.kit.edu/it-security

✗ https://www.s-o-k.de/secure

Criminals replace the expected who-area domain in the web address to deceive you, e.g.

✓ https://www.my-parcelservice.de/

✗ https://www.my-parcelservice.de.online-shopping.de/

✗ https://www.online-shopping.de/my-parcelservice.de

Criminals register who-area domains that are very similar to the correct who-area domain with only a few characters difference:

✓ https://www.farmers-market-total.de/

✗ https://www.farmers-rnarket-total.de/

✗ https://www.farrners-market-total.de/

✗ https://www.farmrers-market-total.de/

**5. Rule**: Having identified the who-area in the web address, but you find you still cannot validate it, collect further information, e.g. by searching for the address in a search machine.

✓ https://www.secuso.org/

✗ https://www.secuso-research.org/

**6. Rule:** If the sender and contents of a message appear plausible and the message has an attachment, check whether this attachment has a potentially (very) dangerous file format.

Potentially dangerous file formats are:

- File formats that can be executed directly (very dangerous):
  e.g. .exe, .bat, .com, .cmd, .scr, .pif
- Formats that may contain macros:
  e. g. Microsoft Office files, such as .doc, .docx, .ppt, .pptx, .xls, .xlsx
- File formats you do not know

**7. Rule:** If the file format is potentially (very) dangerous, open the attachment only if you expected precisely this attachment from the sender. If you are uncertain, collect further information. In no case use the contact details given in the message. For example, call the sender.

If you have opened Office programs and you are asked whether so-called macros are may be executed, think again about whether the message containing the respective file is fraudulent. Terminate the process for the time being.

**Further Information**

How to recognize fraudulent messages is explained in three videos:

**To the videos:**
https://s.kit.edu/it-security.fraudulent-messages.video

You can find further information on fraudulent messages and other cyber security topics here:

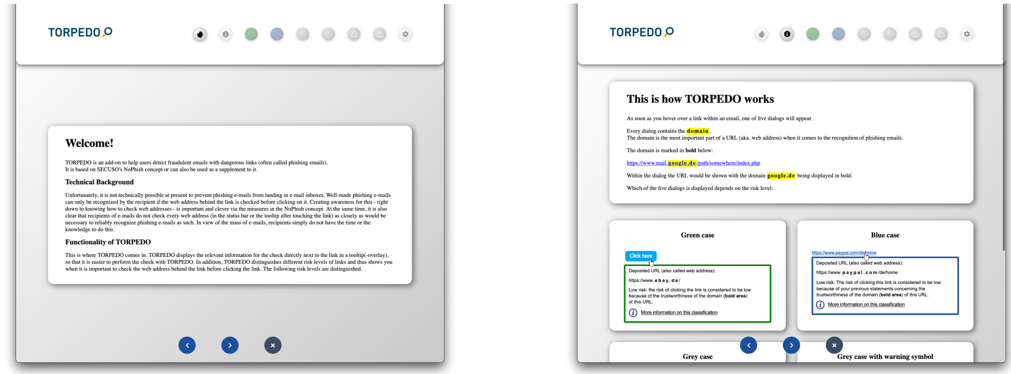
**Further Info:**
https://secuso.aifb.kit.edu/english/642.php

By the way: If you receive feedback that someone has received an email from you that you did not send at all, then you can also ask the BSI for advice.

**Figure A.3:** Second part of the back pages of the leaflet from the original measure.

## A.2 All pages of the tutorial

[2]



Welcome!

TORPEDO is an add-on to help users detect fraudulent emails with dangerous links (often called phishing emails).
It is based on SECUSO's NoPhish concept or can also be used as a supplement to it.

Technical Background

Unfortunately, it is not technically possible at present to prevent phishing e-mails from landing in e-mail inboxes. Well-made phishing e-mails can only be recognized by the recipient if the web address behind the link is checked before clicking on it. Creating awareness for this - right down to knowing how to check web addresses - is important and clever via the measures in the NoPhish concept. At the same time, it is also clear that recipients of e-mails do not check every web address (in the status bar or the tooltip after touching the link) as closely as would be necessary to reliably recognize phishing e-mails as such. In view of the mass of e-mails, recipients simply do not have the time or the knowledge to do this.

Functionality of TORPEDO

This is where TORPEDO comes in. TORPEDO displays the relevant information for the check directly next to the link in a tooltip(-overlay), so that it is easier to perform the check with TORPEDO. In addition, TORPEDO distinguishes different risk levels of links and thus shows you when it is important to check the web address behind the link before clicking the link. The following risk levels are distinguished.

This is how TORPEDO works

As soon as you hover over a link within an email, one of five dialogs will appear.

Every dialog contains the **domain**.
The domain is the most important part of a URL (aka. web address) when it comes to the recognition of phishing emails.

The domain is marked in **bold** below:

https://www.mail.**google.de**/path/somewhere/index.php

Within the dialog the URL would be shown with the domain **google.de** being displayed in bold.

Which of the five dialogs is displayed depends on the risk level:

Green case

Deposited URL (also called web address):

https://www.**ebay.de**/

Low risk: the risk of clicking the link is considered to be low because of the trustworthiness of the domain (**bold area**) of this URL.

More information on this classification

Blue case

https://www.paypal.com/de/home

Deposited URL (also called web address):

https://www.**paypal.com**/de/home

Low risk: The risk of clicking this link is considered to be low because of your previous statements concerning the trustworthiness of the domain (**bold area**) of this URL.

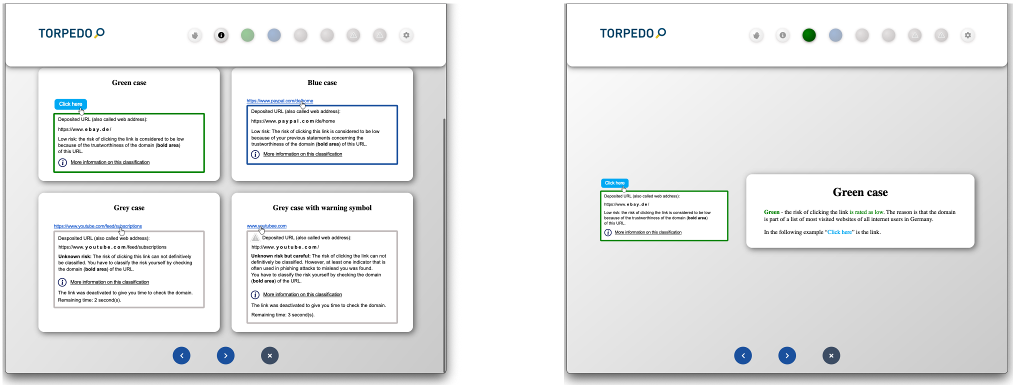
More information on this classification

Grey case

Grey case with warning symbol

**(a)** TORPEDO's tutorial page 1. **(b)** TORPEDO's tutorial top of page 2.

**Figure A.4:** Original version of the tutorial pages 1 and top of page 2.



Green case

Deposited URL (also called web address):

https://www.**ebay.de**/

Low risk: the risk of clicking the link is considered to be low because of the trustworthiness of the domain (**bold area**) of this URL.

More information on this classification

Blue case

https://www.paypal.com/de/home

Deposited URL (also called web address):

https://www.**paypal.com**/de/home

Low risk: The risk of clicking this link is considered to be low because of your previous statements concerning the trustworthiness of the domain (**bold area**) of this URL.

More information on this classification

Grey case

https://www.youtube.com/feed/subscriptions

Desposited URL (also called web address):

https://www.**youtube.com**/feed/subscriptions

**Unknown risk:** The risk of clicking this link can not definitively be classified. You have to classify the risk yourself by checking the domain (**bold area**) of the URL.

More information on this classification

The link was deactivated to give you time to check the domain.
Remaining time: 2 second(s).

Grey case with warning symbol

www.youtubee.com

Deposited URL (also called web address):

http://www.**youtube.com**/

**Unknown risk but careful:** The risk of clicking the link can not definitively be classified. However, at least one indicator that is often used in phishing attacks to mislead you was found.
You have to classify the risk yourself by checking the domain (**bold area**) of the URL.

More information on this classification

The link was deactivated to give you time to check the domain.

Remaining time: 3 second(s).

Deposited URL (also called web address):

https://www.**ebay.de**/

Low risk: the risk of clicking the link is considered to be low because of the trustworthiness of the domain (**bold area**) of this URL.

More information on this classification

Green case

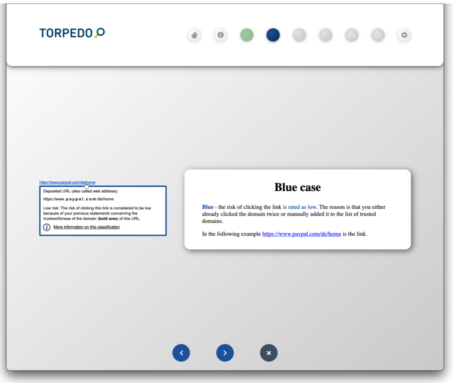
**Green** - the risk of clicking the link is rated as low. The reason is that the domain is part of a list of most visited websites of all internet users in Germany.

In the following example "Click here" is the link.

**(a)** TORPEDO's tutorial bottom of page 2. **(b)** TORPEDO's tutorial page 3.

**Figure A.5:** Original version of the tutorial bottom of page 2 and page 3.

[2] part of the current version available at `https://addons.mozilla.org/de/firefox/addon/torpedo-browser/?src=api`



https://www.paypal.com/de/home

Deposited URL (also called web address):

https://www. **p a y p a l . c o m** /de/home

Low risk: The risk of clicking this link is considered to be low because of your previous statements concerning the trustworthiness of the domain (**bold area**) of this URL.

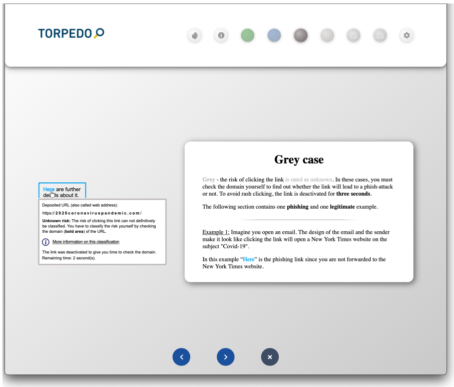
ⓘ More information on this classification

**Blue case**

**Blue** - the risk of clicking the link is rated as low. The reason is that you either already clicked the domain twice or manually added it to the list of trusted domains.

In the following example https://www.paypal.com/de/home is the link.

**(a)** TORPEDO's tutorial page 4.



Here are further details about it.

Deposited URL (also called web address):

https:// **2 0 2 0 c o r o n a v i r u s p a n d e m i c . c o m /**

**Unknown risk:** The risk of clicking this link can not definitively be classified. You have to classify the risk yourself by checking the domain (**bold area**) of the URL.

ⓘ More information on this classification

The link was deactivated to give you time to check the domain.
Remaining time: 2 second(s).

**Grey case**

Grey - the risk of clicking the link is rated as unknown. In these cases, you must check the domain yourself to find out whether the link will lead to a phish-attack or not. To avoid rash clicking, the link is deactivated for **three seconds**.

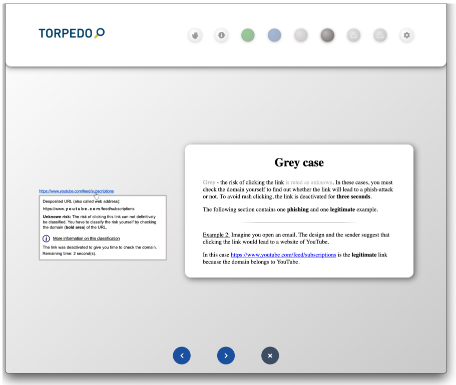
The following section contains one **phishing** and one **legitimate** example.

Example 1: Imagine you open an email. The design of the email and the sender make it look like clicking the link will open a New York Times website on the subject "Covid-19".

In this example "Here" is the phishing link since you are not forwarded to the New York Times website.

**(b)** TORPEDO's tutorial page 5.

**Figure A.6:** Original version of the tutorial pages 4 and 5.



https://www.youtube.com/feed/subscriptions

Desposited URL (also called web address):

https://www. **y o u t u b e . c o m** /feed/subscriptions

**Unknown risk:** The risk of clicking this link can not definitively be classified. You have to classify the risk yourself by checking the domain (**bold area**) of the URL.

ⓘ More information on this classification

The link was deactivated to give you time to check the domain.
Remaining time: 2 second(s).

**Grey case**

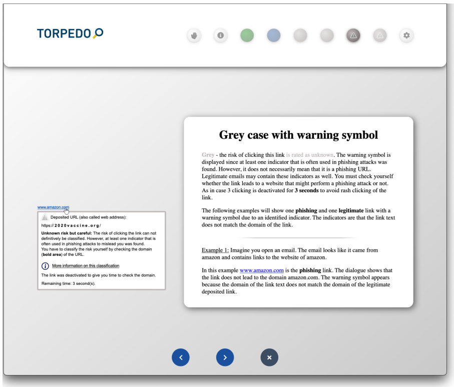
Grey - the risk of clicking the link is rated as unknown. In these cases, you must check the domain yourself to find out whether the link will lead to a phish-attack or not. To avoid rash clicking, the link is deactivated for **three seconds**.

The following section contains one **phishing** and one **legitimate** example.

Example 2: Imagine you open an email. The design and the sender suggest that clicking the link would lead to a website of YouTube.

In this case https://www.youtube.com/feed/subscriptions is the **legitimate** link because the domain belongs to YouTube.

**(a)** TORPEDO's tutorial page 6.



www.amazon.com

Deposited URL (also called web address):

https:// **2 0 2 0 v a c c i n e . o r g /**

**Unknown risk but careful:** The risk of clicking the link can not definitively be classified. However, at least one indicator that is often used in phishing attacks to mislead you was found.
You have to classify the risk yourself by checking the domain (**bold area**) of the URL.

ⓘ More information on this classification

The link was deactivated to give you time to check the domain.

Remaining time: 3 second(s).

**Grey case with warning symbol**

Grey - the risk of clicking this link is rated as unknown. The warning symbol is displayed since at least one indicator that is often used in phishing attacks was found. However, it does not necessarily mean that it is a phishing URL. Legitimate emails may contain these indicators as well. You must check yourself whether the link leads to a website that might perform a phishing attack or not. As in case 3 clicking is deactivated for **3 seconds** to avoid rash clicking of the link.

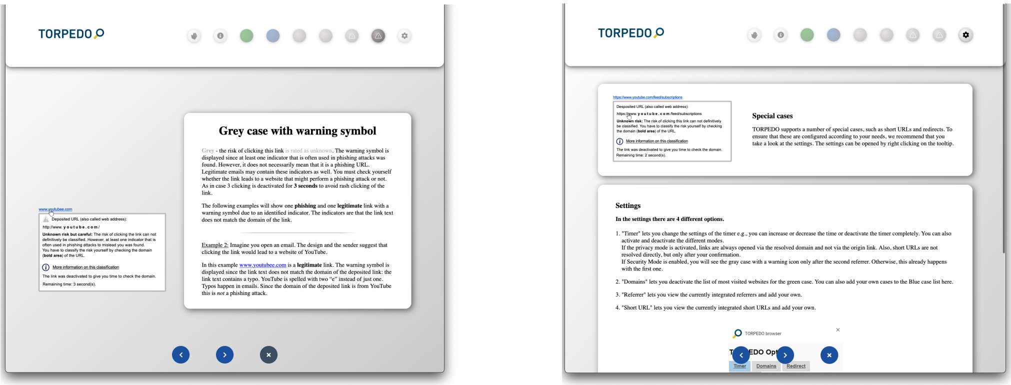
The following examples will show one **phishing** and one **legitimate** link with a warning symbol due to an identified indicator. The indicators are that the link text does not match the domain of the link.

Example 1: Imagine you open an email. The email looks like it came from amazon and contains links to the website of amazon.

In this example www.amazon.com is the **phishing** link. The dialogue shows that the link does not lead to the domain amazon.com. The warning symbol appears because the domain of the link text does not match the domain of the legitimate deposited link.

**(b)** TORPEDO's tutorial page 7.

**Figure A.7:** Original version of the tutorial pages 6 and 7.



## Grey case with warning symbol

Grey - the risk of clicking this link is rated as unknown. The warning symbol is displayed since at least one indicator that is often used in phishing attacks was found. However, it does not necessarily mean that it is a phishing URL. Legitimate emails may contain these indicators as well. You must check yourself whether the link leads to a website that might perform a phishing attack or not. As in case 3 clicking is deactivated for **3 seconds** to avoid rash clicking of the link.

The following examples will show one **phishing** and one **legitimate** link with a warning symbol due to an identified indicator. The indicators are that the link text does not match the domain of the link.

Example 2: Imagine you open an email. The design and the sender suggest that clicking the link would lead to a website of YouTube.

In this example www.youtubee.com is a **legitimate** link. The warning symbol is displayed since the link text does not match the domain of the deposited link: the link text contains a typo. YouTube is spelled with two "e" instead of just one. Typos happen in emails. Since the domain of the deposited link is from YouTube this is *not* a phishing attack.

www.youtubee.com

Deposited URL (also called web address):

http://www.**youtube.com**/

**Unknown risk but careful:** The risk of clicking the link can not definitively be classified. However, at least one indicator that is often used in phishing attacks to mislead you was found. You have to classify the risk yourself by checking the domain (**bold area**) of the URL.

More information on this classification

The link was deactivated to give you time to check the domain.

Remaining time: 3 second(s).

(**a**) TORPEDO's tutorial page 8.

https://www.youtube.com/feed/subscriptions

Desposited URL (also called web address):

https://www.**youtube.com**/feed/subscriptions

**Unknown risk:** The risk of clicking this link can not definitively be classified. You have to classify the risk yourself by checking the domain (**bold area**) of the URL.

More information on this classification

The link was deactivated to give you time to check the domain. Remaining time: 2 second(s).

### Special cases

TORPEDO supports a number of special cases, such as short URLs and redirects. To ensure that these are configured according to your needs, we recommend that you take a look at the settings. The settings can be opened by right clicking on the tooltip.

### Settings

**In the settings there are 4 different options.**

1. "Timer" lets you change the settings of the timer e.g., you can increase or decrease the time or deactivate the timer completely. You can also activate and deactivate the different modes.
   If the privacy mode is activated, links are always opened via the resolved domain and not via the origin link. Also, short URLs are not resolved directly, but only after your confirmation.
   If Security Mode is enabled, you will see the gray case with a warning icon only after the second referrer. Otherwise, this already happens with the first one.
2. "Domains" lets you deactivate the list of most visited websites for the green case. You can also add your own cases to the Blue case list here.
3. "Referrer" lets you view the currently integrated referrers and add your own.
4. "Short URL" lets you view the currently integrated short URLs and add your own.

(**b**) TORPEDO's tutorial page 9.

**Figure A.8:** Original version of the tutorial pages 8 and 9.

# List of Figures

2.1 Example for a warning page from Akhawe et al. [37]. . . . . . . . . . . . . . . . . . 8
2.2 Example for a banner warning from Marchal et al. [45]. . . . . . . . . . . . . . . . . 8
2.3 Example for a link-centric warning with a tooltip from Petelka et al. [5]. . . . . . . 9
3.1 Abstract representation of an email highlighting the status bar and the tooltip. . . . 17
3.2 Start sequence of the video showing the comic style and main character on the right side. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
3.3 Different forms of a link. Google logo censored due to missing copyrights. . . . . . 19
3.4 URL with highlighted 'who-area'. . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
3.5 Example of a page that concludes a chapter with badges of previously completed chapters. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 22
3.6 First part of the back pages of the leaflet from the original measure. . . . . . . . . . 25
3.7 Poster from the original measure. . . . . . . . . . . . . . . . . . . . . . . . . . . . 27
3.8 First page of the interactive example. . . . . . . . . . . . . . . . . . . . . . . . . . 28
3.9 TORPEDO's low risk level warning. . . . . . . . . . . . . . . . . . . . . . . . . . . 30
3.10 TORPEDO's unknown risk warning with indicator. Note, the "unknown risk with indicator" level does not necessarily means that it is a phishing link as shown here, a mismatch may also occur in benign emails. . . . . 31
3.11 TORPEDO's unknown risk warning. . . . . . . . . . . . . . . . . . . . . . . . . . . 32
3.12 Exemplary the first two pages of the original tutorial. . . . . . . . . . . . . . . . . . 33
3.13 Overview of the combination of the dimensions decision and signal in a 4x4 fields matrix. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35
4.1 Study design for the three time points (pre, post, and retention). . . . . . . . . . . . 45
4.2 Boxplot for sensitivity for all three points in time. The mean sensitivity for the sample is on the right side of the boxplots. "*" marking those post-hoc comparison that were significant and "n.s." those that were not significant. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 51
4.3 Study design for the three time points (pre, post and retention). . . . . . . . . . . . 57
4.4 Boxplot for sensitivity for all three points in time. The mean sensitivity is on the right side of the boxplots. "*" marking those post-hoc comparison that are significant ($p < .05$ and "n.s." those that are not significant. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66

4.5 Study design with all groups and comparisons. The dotted groups belong to research question 1 and the vertically striped groups to research question 2. The group without markings is the control group for the effect of the double measurement. . . . . . . . . . . . . . . . . . . . . . . . . . 77
4.6 Study Groups comparing significance to the pre-test. "*" marking those post-hoc comparison that are significant and "n.s." those that are not significant. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 84
5.1 Adapted link-centric warning (TORPEDO) scheme for the study. . . . . . . . . . . 110
5.2 Adapted version of the tutorial used for the study Part1 and Part2. . . . . . . . . . 112
5.3 Adapted version of the tutorial used for the study Part3 and Part4. . . . . . . . . . 113
A.1 First part of the front pages of the leaflet from the original measure. . . . . . . . . . 135
A.2 First part of the back pages of the leaflet from the original measure. . . . . . . . . . 136
A.3 Second part of the back pages of the leaflet from the original measure. . . . . . . . 137
A.4 Original version of the tutorial pages 1 and top of page 2. . . . . . . . . . . . . . . . 138
A.5 Original version of the tutorial bottom of page 2 and page 3. . . . . . . . . . . . . . . 138
A.6 Original version of the tutorial pages 4 and 5. . . . . . . . . . . . . . . . . . . . . . . . 139
A.7 Original version of the tutorial pages 6 and 7. . . . . . . . . . . . . . . . . . . . . . . . 139
A.8 Original version of the tutorial pages 8 and 9. . . . . . . . . . . . . . . . . . . . . . . . 140

# List of Tables

2.1 Overview of awareness retention studies with first author, retention period, result (either * for significant or n.s. for non significant) and special characteristic of the study. "Significance" means that after the measure, the tested period still shows a significant result compared to before. . . . . 5

2.2 Overview of security intervention studies with first author, kind of tool support, result (either * for significant or n.s. for non significant) and special characteristic of the study. "Significance" means that the tool support tested showed a significant result compared to either a control group or to a measurement before. . . . . 6

3.1 Overview of the different phishing tricks or manipulation strategies across different papers. An "x" signals that the mentioned literature included the phishing trick. . . . . 12

3.2 Overview of six phishing tricks from Section 3.1 included per measure and refresher measures. . . . . 18

3.3 Number of pages for each chapter both for content and exercise part. . . . . 21

4.1 Overview of presented phishing messages. "TT" stands for tooltip and "ST" stands for status bar. . . . . 48

4.2 Overview of presented legitimate messages. In the sixth column TT stands for tooltip and ST stands for status bar. . . . . 49

4.3 Overview of the average rates for all three points in time in % for the correct answers for phishing and legitimate examples as well as overall. Also adding the values for the signal detection theory. . . . . 50

4.4 Detection rate in % for phishing tricks for the two phases. . . . . 52

4.5 Detection rate in % for individual phishing URLs for the two phases. . . . . 53

4.6 Detection rate in % for individual legitimate URLs for the two phases. In contrast to the phishing examples, there is no phishing trick included in the legitimate examples, so the column is not present. . . . . 54

4.7 Overview of the phishing URLs for the e-learning study. "TT" stands for tooltip and "ST" stands for status bar. . . . . 61

4.8 Overview of the legitimate URLs for the e-learning study. "TT" stands for tooltip and "ST" stands for status bar. . . . . 62

4.9 Overview of the average rates for all three points in time in % for the correct answers for phishing and legitimate examples as well as overall. Also adding the values for the signal detection theory. . . . . . . . . . . . 64
4.10 Percentages of correct answers per phishing trick for all two phases. . . . . . . . . 67
4.11 Percentages of correct answers per phishing email with phishing trick information for all two phases. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 68
4.12 Percentages of correct answers per legitimate email for all two phases. In contrast to the phishing examples, there is no phishing trick included in the legitimate examples, so the column is not present. . . . . . . . . . . . . . . . 69
4.13 Percentages of correct answers per phishing email. “TT” stands for tooltip and “ST” stands for status bar. . . . . . . . . . . . . . . . . . . . . . . . . 80
4.14 Percentages of correct answers per legitimate email “TT” stands for tooltip and “ST” stands for status bar. . . . . . . . . . . . . . . . . . . . . . . . . 81
4.15 Overview of the average rates for all groups including the refresher groups (LT = long text, IE = interactive email example, V = video and P = short text) in % for the correct answers for phishing and legitimate examples as well as overall. Together with the values for the signal detection theory (SDT). N = total number of participants per group. . . . . . . . . 82
4.16 Percentage of correct answers for phishing tricks for all groups. Also with the data for the refresher groups (LT = long text, IE = interactive email example, V = video and P = short text). . . . . . . . . . . . . . . . . . . . . 86
4.17 Number of correct answers in percent for phishing examples for all groups. Also with the data for the refresher groups (LT = long text, IE = interactive email example, V = video and P = short text). . . . . . . . . . . . . 89
4.18 Number of correct answers in percent for legitimate examples for all groups. Also with the data for the refresher groups (LT = long text, IE = interactive email example, V = video and P = short text). In contrast to the phishing examples, there is no phishing trick included in the legitimate examples, so the column is not present. . . . . . . . . . . . . . . . 90
4.19 Sensitivity overview for all studies of Part I for pre, post and the respective retention phases. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 94
4.20 Sensitivity overview for all refresher groups from study 3 (LT = long text, IE = interactive email example, V = video and P = short text). . . . . . . . . . 94
5.1 Phishing Examples with TORPEDO Risk Levels. . . . . . . . . . . . . . . . . . . . 118
5.2 Legitimate Examples with TORPEDO Risk Levels. . . . . . . . . . . . . . . . . . . 119
5.3 Overview of the average rates for all eight groups in % for the correct answers for phishing and legitimate examples as well as overall. Also adding the values for the signal detection theory (SDT). . . . . . . . . . . . . . . . 120

5.4 Overview of the percentage of correct answers for phishing tricks. Grp1 = status bar, Grp2 = tooltip, Grp3 = video + status bar, Grp4 = video + tooltip, Grp5 = link-centric warning, Grp6 = video + link-centric warning, Grp7 = link-centric warning without tutorial, Grp8 = video + link-centric warning without tutorial. . . . . . . . . . . . . . . . . . . . . . 123

5.5 Overview of the percentage of correct answers for the phishing examples. Grp1 = status bar, Grp2 = tooltip, Grp3 = video + status bar, Grp4 = video + tooltip, Grp5 = link-centric warning, Grp6 = video + link-centric warning, Grp7 = link-centric warning without tutorial, Grp8 = video + link-centric warning without tutorial. . . . . . . . . . . . . . . . . . 124

5.6 Overview of the percentage of correct answers for the legitimate examples. Grp1 = status bar, Grp2 = tooltip, Grp3 = video + status bar, Grp4 = video + tooltip, Grp5 = link-centric warning, Grp6 = video + link-centric warning, Grp7 = link-centric warning without tutorial, Grp8 = video + link-centric warning without tutorial. In contrast to the phishing examples, there is no phishing trick included in the legitimate examples, so the column is not present. . . . . . . . . . . . . . . . . . . . 125

# Bibliography

[1] Omid Asudeh and Mathew Wright. Poster: Phishing website detection with a multiphase framework to find visual similarity. In *CCS 2016*, pages 1790–1792. ACM, 2016.

[2] Xiao Han, Nizar Kheir, and Davide Balzarotti. Phisheye: Live monitoring of sandboxed phishing kits. In *CCS 2016*, pages 1402–1413. ACM, 2016.

[3] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Cranor, Jason Hong, and Elizabeth Nunge. Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. *Symposium on Usable Privacy and Security (SOUPS)*, 2007.

[4] Stephan Neumann, Benjamin Reinheimer, and Melanie Volkamer. Don't Be Deceived: The Message Might Be Fake. *Trust, Privacy and Security in Digital Business (TrustBus)*, pages 199–214, 2017.

[5] Justin Petelka, Yixin Zou, and Florian Schaub. 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, page 1–15, New York, NY, USA, 2019. Association for Computing Machinery.

[6] Melanie Volkamer, Karen Renaud, Benjamin Reinheimer, Philipp Rack, Marco Ghiglieri, Peter Mayer, Alexandra Kunz, and Nina Gerber. Developing and Evaluating a Five Minute Phishing Awareness Video. *Trust, Privacy and Security in Digital Business (TrustBus)*, pages 119–134, 2018.

[7] Benjamin Berens, Mattia Mossano, and Melanie Volkamer. Phishing awareness and education – When to best remind? In *Symposium on Usable Security and Privacy (USEC)*, volume USEC of *Symposium on Usable Security and Privacy (USEC)*, 2022.

[8] Benjamin Reinheimer, Lukas Aldag, Peter Mayer, Mattia Mossano, Reyhan Duezguen, Bettina Lofthouse, Tatiana von Landesberger, and Melanie Volkamer. An investigation of phishing awareness and education over time: When and how to best remind users. In *Symposium on Usable Privacy and Security (SOUPS)*, pages 259–284, 2020.

[9] Benjamin Maximilian Berens, Florian Schaub, Mattia Mossano, and Melanie Volkamer. Better together: The interplay between a phishing awareness video and a link-centric phishing support tool. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.

[10] Stuart E. Schechter, Rachna Dhamija, Andy Ozment, and Ian Fischer. The Emperor's New Security Indicators. In *2007 IEEE Symposium on Security and Privacy (SP '07)*, pages 51–65, 2007.

[11] Ponnurangam Kumaraguru, Yong Rhee, Alessandro Acquisti, Lorrie Cranor, Jason Hong, and Elizabeth Nunge. Protecting people from phishing: the design and evaluation of an embedded training email system. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, Conference on Human factors in computing systems, pages 905–914, 2007. User Study User Study.

[12] Julie S. Downs, Mandy Holbrook, and Lorrie Faith Cranor. Behavioral response to phishing risk. *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit*, pages 37–44, 2007.

[13] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong. Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology (TOIT)*, 10(2):7, 2010.

[14] Rachna Dhamija, J. D. Tygar, and Marti Hearst. Why phishing works. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, page 581–590, New York, NY, USA, 2006. Association for Computing Machinery.

[15] Serge Egelman, Lorrie Faith Cranor, and Jason Hong. You've Been Warned: An Empirical Study of the Effectiveness of Web Browser Phishing Warnings. In *CHI*, pages 1065–1074. ACM, 2008.

[16] Norman Hänsch and Zinaida Benenson. Specifying IT security awareness. *2014 25th International Workshop on Database and Expert Systems Applications*, pages 326–330, 2014.

[17] Bartlomiej Hanus, John C. Windsor, and Yu Wu. Definition and Multidimensionality of Security Awareness. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 49(SI):103–133, 2018.

[18] John Correia and Deborah Compeau. Information Privacy Awareness (IPA): A Review of the Use, Definition and Measurement of IPA. *Proceedings of the 50th Hawaii International Conference on System Sciences (2017)*, 2017.

[19] Ronald W Rogers. Cognitive and physiological processes in fear appeals and attitude change: A revised theory of protection motivation. *Social psychology: A source book*, pages 153–176, 1983.

[20] Hermann Ebbinghaus. Memory: A Contribution to Experimental Psychology. *Annals of Neurosciences*, 20(4):155–156, 2013.

[21] Tianjian Zhang. Knowledge Expiration in Security Awareness Training. *Conference on Digital Forensics, Security and Law (ADFSL)*, 2018.

[22] C. Lanier Benkard. Learning and Forgetting: The Dynamics of Aircraft Production. *American Economic Review*, 90(4):1034–1054, 2000.

[23] Price B. Kerfoot, Yineng Fu, Harley Baker, Donna Connelly, Michael L. Ritchey, and Elizabeth M. Genega. Online Spaced Education Generates Transfer and Improves Long-Term Retention of Diagnostic Skills: A Randomized Controlled Trial. *Journal of the American College of Surgeons*, 211(3):331–337.e1, 2010.

[24] Eric D. Darr, Linda Argote, and Dennis Epple. The Acquisition, Transfer, and Depreciation of Knowledge in Service Organizations: Productivity in Franchises. *Management Science*, 41(11):1750–1762, 1995.

[25] Joel D. Schendel and Joseph D. Hagman. On Sustaining Procedural Skills Over Prolonged Retention Intervals. *Journal of Applied Psychology*, 67:605, 1982.

[26] Punam Anand and Brian Sternthal. Ease of Message Processing as a Moderator of Repetition Effects in Advertising. *Journal of Marketing Research*, 27(3):345–353, 1990.

[27] Frank N. Dempster. The Spacing Effect. *Journal of Research & Development in Education*, 1990.

[28] Chris Janiszewski, Hayden Noel, and Alan G. Sawyer. A Meta-analysis of the Spacing Effect in Verbal Learning: Implications for Research on Advertising Repetition and Consumer Memory. *Journal of Consumer Research*, 30(1):138–149, 2003.

[29] Matthew L. Jensen, Michael Dinger, Ryan T. Wright, and Jason Bennett Thatcher. Training to Mitigate Phishing Attacks Using Mindfulness Techniques. *Journal of Management Information Systems*, 34(2):597–626, 2017.

[30] Elmer Lastdrager, Iné Carvajal Gallardo, Pieter Hartel, and Marianne Junger. How effective is anti-phishing training for children? *Symposium on Usable Privacy and Security*, 2017.

[31] Ponnurangam Kumaraguru, Justin Cranshaw, and Alessandro Acquisti. School of phish: a real-world evaluation of anti-phishing training. *SOUPS*, 2009.

[32] Christopher Nguyen, Matthew Jensen, and Eric Day. Learning not to take the bait: a longitudinal examination of digital training methods and overlearning on phishing susceptibility. *European Journal of Information Systems*, 32(2):238–262, 2023.

[33] Gamze Canova, Melanie Volkamer, Clemens Bergmann, and Benjamin Reinheimer. NoPhish App Evaluation: Lab and Retention Study. In *USEC*. Internet Society, 2015.

[34] Benjamin M. Berens, Mattia Mossano, and Melanie Volkamer. Taking 5 minutes protects you for 5 months: Evaluating an anti-phishing awareness video. *Computers & Security*, 137:103620, 2024.

[35] David A. Nembhard and Mustafa V. Uzumeri. An individual-based description of learning within an organization. *IEEE Transactions on Engineering Management*, 47(3):370–378, 2000.

[36] Mohamad Y. Jaber, Hemant V. Kher, and Darwin J. Davis. Countering forgetting through training and deployment. *International Journal of Production Economics*, 85(1), 2003.

[37] Devdatta Akhawe and Adrienne Porter Felt. Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness. *USENIX security symposium*, 13, 2013.

[38] Serge Egelman, Lorrie Faith Cranor, and Jason Hong. You've Been Warned: An Empirical Study of the Effectiveness of Web Browser Phishing Warnings. In *CHI*, pages 1065–1074. ACM, 2008.

[39] Adrienne Porter Felt, Alex Ainslie, Robert W. Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettes, Helen Harris, and Jeff Grimes. Improving SSL Warnings: Comprehension and Adherence. *Proceedings of the 33rd*, pages 2893–2902, 2015.

[40] Joshua Sunshine, Serge Egelman, Hazim Almuhimedi, Neha Atri, and Lorrie Cranor. Crying Wolf: An Empirical Study of SSL Warning Effectiveness. *USENIX security symposium*, pages 399–416, 2009.

[41] José Carlos Brustoloni and Ricardo Villamarín-Salomón. Improving security decisions with polymorphic and audited dialogs. In *Proceedings of the 3rd Symposium on Usable Privacy and Security*, SOUPS '07, page 76–85, New York, NY, USA, 2007. Association for Computing Machinery.

[42] Nathalie Stembert, Arne Padmos, Mortaza S. Bargh, Sunil Choenni, and Frans Jansen. A Study of Preventing Email (Spear) Phishing by Enabling Human Intelligence. *2015 European Intelligence and Security Informatics Conference*, pages 113–120, 2015.

[43] Bonnie Brinton Anderson, C. Brock Kirwan, Jeffrey L. Jenkins, David Eargle, Seth Howard, and Anthony Vance. How polymorphic warnings reduce habituation in the brain: Insights from an fmri study. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 2883–2892, New York, NY, USA, 2015. Association for Computing Machinery.

[44] Min Wu, Robert C. Miller, and Simson L. Garfinkel. Do security toolbars actually prevent phishing attacks? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, page 601–610, New York, NY, USA, 2006. Association for Computing Machinery.

[45] Samuel Marchal, Giovanni Armano, Tommi Grondahl, Kalle Saari, Nidhi Singh, and N. Asokan. Off-the-Hook: An Efficient and Usable Client-Side Phishing Prevention Application. *IEEE Transactions on Computers*, 66(10), 2017.

[46] Mario Silic, Dianne Cyr, Andrea Back, and Adrian Holzer. Effects of Color Appeal, Perceived Risk and Culture on User's Decision in Presence of Warning Banner Message. *Proceedings of the 50th Hawaii International Conference on System Sciences (2017)*, 2017.

[47] Adrienne Porter Felt, Robert W Reeder, Alex Ainslie, Helen Harris, Max Walker, Christopher Thompson, Mustafa Embre Acer, Elisabeth Morant, and Sunny Consolvo. Rethinking connection security indicators. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, pages 1–14, 2016.

[48] Yoshifumi Kitamura, Aaron Quigley, Katherine Isbister, Takeo Igarashi, Pernille Bjørn, Steven Drucker, Kholoud Althobaiti, Nicole Meng, and Kami Vaniea. I Don't Need an Expert! Making URL Phishing Features Human Comprehensible. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2021.

[49] Rachna Dhamija and J. D. Tygar. The battle against phishing: Dynamic Security Skins. In *Symposium on Usable Privacy and Security*, pages 77–88, New York, NY, USA, 2005. ACM.

[50] Peter Likarish, Donald E. Dunbar, Juan Pablo Hourcade, and Eunjin Jung. Bayeshield: conversational anti-phishing user interface. In *SOUPS*, volume 9, pages 1–1, 2009.

[51] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Cranor, Saranga Komanduri, Pedro Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Yang Wang, and Shomir Wilson. Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. *ACM Computing Surveys (CSUR)*, 50(3):44, 2017.

[52] Daniel Kahneman and Amos Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263, 1979.

[53] Kholoud Althobaiti, Kami Vaniea, and Serena Zheng. Faheem: Explaining URLs to people using a slack bot. In *Symposium on Digital Behaviour Intervention for Cyber Security*, page 8, 2018.

[54] Duncan P. Brumby, Christian P. Janssen, and Gloria Mark. How Do Interruptions Affect Productivity? In *Rethinking Productivity in Software Engineering*, pages 85–107. Springer Nature, 2019.

[55] Mary Czerwinski, Arnie Lund, Desney Tan, Gloria Mark, Daniela Gudith, and Ulrich Klocke. The cost of interrupted work. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 107–110, 2008.

[56] Eric Lin, Saul Greenberg, Eileah Trotter, David Ma, and John Aycock. Does domain highlighting help people identify phishing sites? In *Proceedings of CHI 2011 Human Factors in Computing Systems*, pages 2075–2084. ACM, 2011.

[57] Eint Sandi Aung, Chaw Thet Zan, and Hayato YAMANA. URL Based Phishing Detection. *International Journal of Recent Technology and Engineering (IJRTE)*, 9(1):1872–1875, 2020.

[58] Robert Reeder, Iulia Ion, and Sunny Consolvo. 152 Simple Steps to Stay Safe Online: Security Advice for Non-tech-savvy Users. *IEEE Security & Privacy*, PP(99):1–1, 2017.

[59] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Cranor, and Jason Hong. Lessons From a Real World Evaluation of Anti-Phishing Training. *2008 eCrime Researchers Summit*, pages 1–12, 2008.

[60] Deanna D. Caputo, Shari Lawrence Pfleeger, Jesse D. Freeman, and M. Eric Johnson. Going spear phishing: Exploring embedded training and awareness. *IEEE Security & Privacy*, 2014.

[61] Simon Stockhardt, Benjamin Reinheimer, Melanie Volkamer, Peter Mayer, Alexandra Kunz, Philipp Rack, and Daniel Lehmann. Teaching Phishing-Security: Which Way is Best? *International Conference on ICT Systems Security and Privacy Protection (IFIP SEC)*, 2016.

[62] Steve Sheng, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Faith Cranor, and Julie Downs. Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the 2010 CHI Conference on Human Factors in Computing Systems*, Conference on Human Factors in Computing Systems, 2010. User Study Training Material.

[63] Keith Rayner, Alexander Pollatsek, Jane Ashby, and Charles Clifton Jr. *Psychology of reading*. Psychology Press, 2012.

[64] Stanislas Dehaene. *Reading in the brain: The new science of how we read.* Penguin, 2010.

[65] Markus Jakobsson, Alex Tsow, Ankur Shah, Eli Blevis, and Youn-Kyung Lim. What instills trust? A qualitative study of phishing. In *Financial Crypto*, pages 356–361. Springer, 2007.

[66] Xun Dong, John A. Clark, and Jeremy Jacob. Modelling user-phishing interaction. In *Human System Interactions*, pages 627–632. IEEE, 2008.

[67] Julie S. Downs, Mandy B. Holbrook, and Lorrie Faith Cranor. Decision strategies and susceptibility to phishing. In *SOUPS*, pages 79–90, Pittsburgh, Pennsylvania, USA, 2006. ACM.

[68] Rufai Ahmad and Sotirios Terzis. Understanding phishing in mobile instant messaging: A study into user behaviour toward shared links. In Nathan Clarke and Steven Furnell, editors, *Human Aspects of Information Security and Assurance*, pages 197–206, Cham, 2022. Springer International Publishing.

[69] Michaela Kauer, Thomas Pfeiffer, Melanie Volkamer, Heike Theuerling, and Ralph Bruder. It is not about the design — it is about the content! Making warnings more efficient by communicating risks appropriately. In *Sicherheit*, volume 195. GI, 2012.

[70] Tom N. Jagatic, Nathaniel A. Johnson, Markus Jakobsson, and Filippo Menczer. Social phishing. *Communications of the ACM*, 50(10):94–100, 2007.

[71] Leyla Bilge, Thorsten Strufe, Davide Balzarotti, and Engin Kirda. All your contacts are belong to us: automated identity theft attacks on social networks. In *Proceedings of the 18th International Conference on World Wide Web*, WWW ’09, page 551–560, New York, NY, USA, 2009. Association for Computing Machinery.

[72] Simon Stockhardt, Benjamin Reinheimer, and Melanie Volkamer. Über die Wirksamkeit von Anti-Phishing-Training. *21st International Workshop on Intelligent and Personalized Human-Computer Interaction*, pages 647–655, 2015.

[73] ILIAS open source e-Learning e.V. The Open Source Learning Management System. https://www.ilias.de/, 2023. Accessed 10 October 2023.

[74] Thunderbird AddOns. AddOn - TORPEDO - TOoltip-poweRed Phishing Email DetectiOn. `https://addons.thunderbird.net/de/thunderbird/addon/torpedo-phishing-detection/`, 2023. Accessed 23 October 2023.

[75] Melanie Volkamer, Karen Renaud, Benjamin Reinheimer, and Alexandra Kunz. User experiences of TORPEDO: TOoltip-poweRed phishing email DetectiOn. *Computers & Security*, 71:100–113, 2017.

[76] Melanie Volkamer, Karen Renaud, and Benjamin Reinheimer. Torpedo: Tooltip-powered phishing email detection. In *IFIP SEC*, pages 161–175. Springer, 2016.

[77] Andrew J. Elliot, Markus A. Maier, Martin J. Binser, Ron Friedman, and Reinhard Pekrun. The Effect of Red on Avoidance Behavior in Achievement Contexts. *Personality and Social Psychology Bulletin*, 35(3):365–375, 2009.

[78] Harold Stanislaw and Natasha Todorov. Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1):137–149, 1999.

[79] Christopher B. Mayhorn and Patrick G. Nyeste. Training users to counteract phishing. *Work (Reading, Mass.)*, 41 Suppl 1:3549–52, 2012.

[80] María M. Moreno-Fernández, Fernando Blanco, Pablo Garaizar, and Helena Matute. Fishing for phishers. Improving Internet users' sensitivity to visual deception cues to prevent electronic fraud. *Computers in Human Behavior*, 69:421–436, 2017.

[81] Iain Embrey and Kim Kaivanto. Many Phish in the C: A Coexisting-Choice-Criteria Model of Security Behavior. *arxiv*, 2018.

[82] Casey Canfield, Alex Davis, Baruch Fischhoff, Alain Forget, Sarah Pearman, and Jeremy Thomas. Replication: Challenges in using data logs to validate phishing detection ability metrics. *Symposium on Usable Privacy and Security (SOUPS)*, 2017.

[83] Casey Inez Canfield and Baruch Fischhoff. Setting Priorities in Behavioral Interventions: An Application to Reducing Phishing Risk. *Risk Analysis*, 38:826–838, 2018.

[84] Jaclyn Martin, Chad Dubé, and Michael D. Coovert. Signal Detection Theory (SDT) Is Effective for Modeling User Behavior Toward Phishing and Spear-Phishing Attacks. *Human Factors: The Journal of Human Factors and Ergonomics Society*, 60:1179–1191, 2018.

[85] Marcus A Butavicius, Kathryn Parsons, Malcolm R Pattinson, Agata McCormac, Dragana Calic, and Meredith Lillie. Understanding Susceptibility to Phishing Emails: Assessing the Impact of Individual Differences and Culture. *International Symposium on Human Aspects of Information Security & Assurance (HAISA)*, 2017.

[86] Casey Canfield, Baruch Fischhoff, and Alex Davis. Quantifying Phishing Susceptibility for Detection and Behavior Decisions. *Human Factors: The Journal of Human Factors and Ergonomics Society*, 58:1158–1172, 2016.

[87] Casey Canfield, Baruch Fischhoff, and Alex Davis. Using Signal Detection Theory to Measure Phishing Detection Ability and Behavior. In *SOUPS*, 2015.

[88] Jaclyn Martin. *Something Looks Phishy Here: Applications of Signal Detection Theory to Cyber-Security Behaviors in the Workplace*. PhD thesis, University of South Florida, 2017.

[89] Marcus Butavicius, Kathryn Parsons, Malcolm Pattinson, and Agata McCormac. Breaching the human firewall: Social engineering in phishing and spear-phishing emails. *Australasian Conference on Information Systems*, 2016.

[90] Dominique Makowski. The psycho Package: an Efficient and Publishing-Oriented Workflow for Psychological Science. *The Journal of Open Source Software*, 3(22):470, 2018.

[91] Sosci Survey. Kontaktdaten getrennt erheben. `https://www.soscisurvey.de/help/doku.php/de:create:questions:contact#:~:text=Um%20die%20Anonymität%20der%20Teilnehmer,Stelle%20in%20den%20Fragebogen%20ein`, 2023. Accessed 18 September 2023.

[92] Jacob Cohen. *Statistical power analysis for the behavioral sciences*. Academic press, 2013.

[93] Ayako A Hasegawa, Naomi Yamashita, Mitsuaki Akiyama, and Tatsuya Mori. Why they ignore english emails: The challenges of non-native speakers in identifying phishing emails. In *Seventeenth Symposium on Usable Privacy and Security ({SOUPS} 2021)*, pages 319–338, 2021.

[94] Sosci Survey. Personal Data. `https://www.soscisurvey.de/help/doku.php/en:general:dsgvo`, 2023. Accessed 18 September 2023.

[95] Andy Field. *Discovering statistics using IBM SPSS statistics*. sage, 2013.

[96] Thomas D. Wickens and Geoffrey Keppel. *Design and analysis: A researcher's handbook*. Pearson Prentice-Hall Upper Saddle River, NJ, 2004.

[97] Paul D. Ellis. *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge university press, 2010.

[98] Steven J. Murdoch and Martina A. Sasse. Should you phish your own employees? https://www.benthamsgaze.org/2017/08/22/should-you-phish-your-own-employees/, 2017. Accessed 22 October 2023.

[99] Rick Wash and Molly Cooper. Who Provides Phishing Training? In *Proceedings of CHI 2018 Human Factors in Computing Systems*, 2018.

[100] Peter C. Brown, Henry L. Roediger III, and Mark A. McDaniel. *Make it stick: The science of successful learning*. Harvard University Press, 2014.

[101] James G. March and David Easton. The power of power. *Classics of organization theory*, pages 261–273, 1966.

[102] Ben D. Sawyer and Peter A. Hancock. Hacking the Human: The Prevalence Paradox in Cybersecurity. *Human Factors: The Journal of Human Factors and Ergonomics Society*, 60(5):597–609, 2018.

[103] Kuldeep Singh, Palvi Aggarwal, Prashanth Rajivan, and Cleotilde Gonzalez. Training to Detect Phishing Emails: Effects of the Frequency of Experienced Phishing Emails. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 63(1):453–457, 2019.

[104] Nicola Döring and Jürgen Bortz. Forschungsmethoden und evaluation. *Wiesbaden: Springerverlag*, 2016.

[105] Daniela Oliveira, Harold Rocha, Huizi Yang, Donovan Ellis, Sandeep Dommaraju, Melis Muradoglu, Devon Weir, Adam Soliman, Tian Lin, and Natalie Ebner. Dissecting spear phishing emails for older vs young adults: On the interplay of weapons of influence and life domains in predicting susceptibility to phishing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Conference on Human Factors in Computing Systems (CHI), 2017.

[106] James Nicholson, Lynne Coventry, and Pam Briggs. Can we fight social engineering attacks by social means? Assessing social salience as a means to improve phish detection. *Symposium on Usable Privacy and Security (SOUPS)*, 2017.

[107] Brij B. Gupta, Aakanksha Tewari, Ankit Kumar Jain, and Dharma P. Agrawal. Fighting against phishing attacks: state of the art and future challenges. *Neural Computing and Applications*, 28(12), 2017. No Study.

[108] Zainab Alkhalil, Chaminda Hewage, Liqaa Nawaf, and Imtiaz Khan. Phishing Attacks: A Recent Comprehensive Study and a New Anatomy. *Frontiers in Computer Science*, 3:563060, 2021.

[109] Sara Albakry, Kami Vaniea, and Maria K. Wolters. What is this URL's destination? empirical evaluation of users' URL reading. In *Conference on Human Factors in Computing Systems (CHI)*, pages 1–12. ACM, 2020.

[110] Litmus. Email Client Market Share. `https://www.litmus.com/email-client-market-share/`, 2023. Accessed 23 October 2023.

[111] APWG. Global Phishing Survey. `https://docs.apwg.org/reports/apwg_trends_report_q4_2016.pdf`, 2016. Accessed 22 October 2023.

[112] Wombat Security Technologies. State of the Phish: Effectively Reducing Phishing and Malware Infections. `https://www.proofpoint.com/sites/default/files/wombatsecurity/WombatThreatSim-StateofPhish2016_final_web.pdf`, 2016. Accessed 22 October 2023.

[113] Vaibhav Garg, Jean Camp, Lesa Mae, and Katherine Connelly. Designing risk communication for older adults. In *Symposium on Usable Privacy and Security (SOUPS)*, 2011.

[114] Lennon YC Chang and Nicholas Coppel. Building cyber security awareness in a developing country: lessons from myanmar. *Computers & Security*, 97:101959, 2020.

[115] Zikai Wen, Zhiqiu Lin, Rowena Chen, and Erik Andersen. What.Hack: Engaging Anti-Phishing Training Through a Role-playing Phishing Simulation Game. In *Conference on Human Factors in Computing Systems (CHI)*, page 108, 2019.

[116] Nalin Asanka Gamagedara Arachchilage and Steve Love. Security awareness of computer users: A phishing threat avoidance perspective. *Computers in Human Behavior*, pages 304–312, 2014.

[117] Nalin Asanka Gamagedara Arachchilage, Steve Love, and Konstantin Beznosov. Phishing threat avoidance behaviour: An empirical investigation. *Computers in Human Behavior*, 60:185–197, 2016.

[118] Abdullah Alnajim and Malcolm Munro. An Anti-Phishing Approach that Uses Training Intervention for Phishing Websites Detection . In *6th International Conference on Information Technology: New Generations*, pages 405–410. IEEE, 2009.

[119] Yue Zhang, Jason I Hong, and Lorrie F Cranor. CANTINA: A Content-Based Approach to Detecting Phishing Web Sites. In *16th International World Wide Web Conference*, pages 639–648, 2007.

[120] Rebecca Smith. How a U.S. utility Got Hacked. `https://www.wsj.com/articles/how-a-u-s-utility-got-hacked-1483120856`, 2016. Accessed 23 October 2023.

[121] Leah Zhang-Kennedy, Elias Fares, Sonia Chiasson, and Robert Biddle. Geo-Phisher: The Design and Evaluation of Information Visualizations About Internet Phishing Trends. *2016 APWG Symposium on Electronic Crime Research (eCrime)*, pages 1–12, 2016.

[122] Carlo Sugatan and Florian Schaub. Interactive Stories for Security Education: A Case Study on Password Managers. *Symposium on Usable Privacy and Security*, 2020.

[123] Nazmi Al-Shalabi. Keeping Students Engaged: A Prerequisite for Learning. *Mediterranean Journal of Social Sciences*, 6(5):576–580, 2015.

[124] Wei Bai, Michael Pearson, Patrick Gage Kelley, and Michelle L. Mazurek. Improving Non-Experts' Understanding of End-to-End Encryption: An Exploratory Study. *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 00:210–219, 2020.

[125] Sanchari Das, Shrirang Mare, and L. Jean Camp. Smart Storytelling: Video and Text Risk Communication to Increase MFA Acceptability. *2020 IEEE 6th International Conference on Collaboration and Internet Computing (CIC)*, 00:153–160, 2020.

[126] Melanie Volkamer, Martina Angela Sasse, and Franziska Boehm. Analysing Simulated Phishing Campaigns for Staff. *Lecture Notes in Computer Science*, pages 312–328, 2020.

[127] Melanie Volkamer, Martina A. Sasse, and Franziska Boehm. Phishing-Kampagnen zur Steigerung der Mitarbeiter-Awareness. *Datenschutz und Datensicherheit - DuD*, 44(8):518–521, 2020.

[128] dejure.org Rechtsinformationssysteme GmbH. Grundsätze für die Verarbeitung personenbezogener Daten. https://dejure.org/gesetze/DSGVO/5.html, 2023. Accessed 10 October 2023.

[129] dejure.org Rechtsinformationssysteme GmbH. Sicherheit der Verarbeitung. https://dejure.org/gesetze/DSGVO/32.html, 2016. Accessed 10 October 2023.

[130] dejure.org Rechtsinformationssysteme GmbH. Aufgaben des Datenschutzbeauftragten. https://dejure.org/gesetze/DSGVO/39.html, 2016. Accessed 10 October 2023.

[131] International Organization for Standardization. ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection — Information security management systems — Requirements. `https://www.iso.org/obp/ui/en/#iso:std:82875:en`, 2023. Accessed 10 October 2023.

[132] Daniele Lain, Kari Kostiainen, and Srdjan Čapkun. Phishing in Organizations: Findings from a Large-Scale and Long-Term Study. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 842–859. IEEE, 2022.

[133] Kai Florian Tschakert and Sudsanguan Ngamsuriyaroj. Effectiveness of and user preferences for security awareness training methodologies. *Heliyon*, 5(6):e02010, 2019. User Study (Pre-Post) Training vs. video-, game-, and text-based.