Microsoft Office Macro Warnings: A Design Comedy of Errors with Tragic Security Consequences

Marco Gutfleisch
Marco.Gutfleisch@rub.de
Ruhr University Bochum
Bochum, North Rhine-Westphalia, Germany

Maximilian Peiffer
Maximilian.Peiffer@rub.de
Ruhr University Bochum
Bochum, North Rhine-Westphalia, Germany

Selim Erk
Selim.Erk@rub.de
Ruhr University Bochum
Bochum, North Rhine-Westphalia, Germany

Martina Angela Sasse
Martina.Sasse@rub.de
Ruhr University Bochum
Bochum, North Rhine-Westphalia, Germany

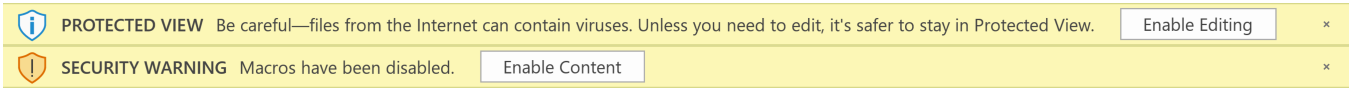

Figure 1: Warning messages in Microsoft Office 2019: Protected view and macro warning.

ABSTRACT

The security threat emanating from macro viruses is currently on the rise. Macros are deactivated by default, but when opening a Microsoft Office document with embedded macros, users are presented with a warning message and a one-click option to activate the macro. The aim of the study was to investigate how users interact with this design, to what extent they are aware of the implications of their choices, and how much they know about macros at all. We designed a mixed-methods experiment - consisting of a set of benchmark tasks, knowledge questions, and interviews, which we conducted remotely. To avoid priming participants, the study was advertised as a performance test of a new Outlook Plugin. 36 participants were presented with a naturalistic workflow of emails, some of which contained attachments with macros. We captured how participants interacted with warning messages, and whether they enabled macros. In a subsequent interview, we explored their perception of what had happened, and why they had chosen to enable macros. We found out that 63.9 % of the participants unnecessarily enabled at least one macro when seeing the messages, and that most did not have an accurate mental model of how macros work or the risks associated with opening them. We discuss what elements lead to the enabling of macros and examine them from different perspectives.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in interaction design**; • **Security and privacy** → **Usability in security and privacy**; **Phishing**; **Malware and its mitigation**.

EuroUSEC '21, October 11–12, 2021, Karlsruhe, Germany

ACM ISBN 978-1-4503-8423-0/21/10.
https://doi.org/10.1145/3481357.3481512

KEYWORDS

macros, mental models, risk perception, warnings

ACM Reference Format:
Marco Gutfleisch, Maximilian Peiffer, Selim Erk, and Martina Angela Sasse. 2021. Microsoft Office Macro Warnings: A Design Comedy of Errors with Tragic Security Consequences. In *European Symposium on Usable Security 2021 (EuroUSEC '21), October 11–12, 2021, Karlsruhe, Germany*. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3481357.3481512

1 INTRODUCTION

For more than 20 years, it has been possible to automate processes in Microsoft Office products with scripts - commonly referred to as macros. Macros are popular because they standardise, reduce workload and increase productivity. But they can also be an attack vector, allowing viruses and Trojans to enter and spread through company systems or infect private computers [8]. The first scripts containing malicious code in Microsoft Office documents were found in the nineties, and shortly after the term 'macro virus' was coined. The first major security incident involving macros occurred in 1999 when the Melissa macro virus spread rapidly and affected more than 1 million computers and 50,000 servers of almost 8,000 companies in North America. The cost was estimated at nearly half a billion USD [16].

Since 2014, Emotet has been spread via macros; the United States Department of Homeland Security (DHS) and the German Federal Office for Information Security (BSI) classify it as the most dangerous malware in the world [5, 6, 39]. In 2021, the infrastructure for Emotet was destroyed after seven years of research effort by several government organisations from all over the world [12]. But the attack vector macro remains, for other malware to exploit. In 2007, Microsoft introduced a security barrier against macro viruses to their OfficeSuite: macros are blocked by default, and warning messages with a one-click option to enable them are displayed [2]. The final security decision is left to the user. It is currently unknown to what extent the macro warnings in Microsoft Office are helpful in communicating risk to users or enable them to make

appropriate decisions on whether to enable macros or not. Design and security practitioners have commented informally that this particular warning might be misinterpreted by users [9, 35], but there was no published study on how users respond to the macro warnings.

In order to assess whether the macro warnings in Microsoft Office 2019 led to appropriate security decisions, we want to answer the following questions:

(1) How do users interact with the warning?
(2) Do macro warning messages effectively communicate the security risk involved?
(3) Do users understand how macros work, and the potential risks associated with enabling them?

Answering the questions above should allow us to answer our overall research question:

Do Microsoft Office macro warnings help users understand and assess the risks posed by Office macros?

We found that the macro warnings trigger insecure choices, without users realising. This is partly because they do not recognise the warning message as such. A detailed examination of our results suggests that the preceding warning message establishes a dangerous routine of clicking a box ('click habit'), which users then execute when asked whether they want to enable macros. Building on experience with improving SSL warnings [13], the appearance, text and required user action should be changed. For companies that have IT and security departments, the job of vetting macros should be done there, not by the users. Finally, whilst we are mindful that user education is a last, not first resort, we suggest that providing some very basic instruction to users who work with macros would have a significant security benefit.

2 RELATED WORK

Here we consider the existing literature on user choices and warnings in the context of security. One of the basic principles is that users are focused on completing production (primary) tasks - such as answering a set of emails; security tasks are enabling (secondary) tasks that should be designed to allow efficient completion [36]. Security tasks that disrupt production tasks create friction; whilst small amounts of friction are tolerated - especially when users understand the risks that are being mitigated - too much friction leads to the ignoring or bypassing of security tasks [3, 18]. In daily life - work or home - 80-95 percent of behaviour is automatic: actions we carry out frequently become automatic routines [15] - it is what makes us efficient and productive [7]. These routines are triggered in long-term memory when users encounter familiar cues [32], and are executed 'without thinking' - i.e. the user does not contemplate alternatives or make a deliberate decision [11]. IT users today have hundreds of such routines embedded in their memory; 'unlearning' a routine once it is embedded requires a significant and sustained effort over a period of time [24]. In the following, we examine the history of macros and their associated security measures, before examining the UI design issues associated with them. Since the usability of the security mechanism has not been studied, we consider related research on security warnings and phishing awareness.

Macro warnings. In Office 2000, Microsoft introduced the ability to allow only digitally signed macros from trusted sources to run and to automatically disable unsigned macros. As a result, most users could not activate embedded macros. Only when the security settings were changed, unsigned macros could be manually executed. With the start of Office 2007, Microsoft set the option 'Disable all macros with notification' as a default setting [2], thus downgrading the original idea of allowing only digitally signed macros as a default setting, and it once again opened the way for macro viruses to be enabled easily and accidentally. In 2016, Microsoft added a second warning message, which appears before the actual macro warning if files from the internet and other potentially unsafe locations are opened. The user then has to click on 'enable editing' in order to make changes in the document [29]. Both warnings are shown in figure 1. Additionally, Microsoft has introduced the functionality to restrict the execution of macros globally in the corporate edition [28]. Most people today who use Microsoft Office use it for many hours a day - meaning most actions they carry out are automatic. An analytical walk-through using a method such as Reason would predict a tendency for users to 'yes-click' both warnings [34]. Müller et al. demonstrate that macros can be activated via social engineering attacks, and once activated, there is nothing that limits their function - they are as a result one of the most devastating attack vectors [30]. The work of Gajek demonstrates the far-reaching consequences: attackers use a range of obfuscation techniques to create countless unique variants of malicious code that cannot be detected by scanning techniques, nor clearly classified as dangerous, or not [14].

Usability considerations. User attention is focused on their primary task - when it comes to macros, the task is set by the content of the email the users are working on. Only two of the three control elements (the small cross on the far right and the button labelled 'enable content') provide an affordance for 'doing something' to reach the next step, by clicking on the element. In the MS Office environment, users have to click such buttons all the time - meaning it is a highly learnt routine, embedded in the user's memory. Such routines are executed automatically when the cue that triggers it is presented - meaning there is a change in cognitive activity [17]. Security experts who say people should 'stop and think' before moving to the next step to be secure ignore the massive productivity cost that this would entail [18, 19].

Users can also click on the text "SECURITY WARNING" to receive further information about the underlying security risk. However, users have been trained to dismiss security warnings. Akhawe et al. found that users were exposed to large numbers of SSL security warnings that were mostly false alarms [1], thus causing alarm fatigue. The experience of warnings they cannot understand is frustrating at best [42], but also leads to misinterpretations that they are not about significant risk - a conviction that deepens over time because users ignore the warning, and find that nothing 'bad' happens [25]. Thus, many users perceive security warnings as hurdles that block their road to primary task completion, which has led users to develop a routine of ignoring or 'swatting' security warnings [4, 38], so it is unlikely that most users would pursue this option unless specifically primed to watch for security risks. Passive security indicators become 'blind spots' after repeated exposure, meaning

that a change in status is not noticed. Schechter et al. reported that removal of SSL indicators (signifying a bone fide website) on a bank website was not noticed, and users went ahead and entered their passwords [37]. The website requiring login credentials provided the cue that triggered the deeply embedded routine. Wu et al. tested the effectiveness of security toolbars to prevent phishing attacks [41]. They found that users did not pay attention to these indicators, and 34 percent of participants entered their credentials even when the toolbar indicated the website was unsafe. Active browser warnings have been more effective in laboratory studies. Egelman et al. compared the effect of active and passive warning indicators to prevent phishing: 79 percent of participants closed the malicious warnings when confronting an active warning, whereas only 13 percent of participants presented with passive indicators did [10].

3 METHODOLOGY

Because this is the first study that explores users' security behaviour and understanding of Microsoft Office (2019) macros, we decided to combine qualitative and quantitative methods within an experiment. To avoid priming the participants with security and macros, they were recruited for a study "to measure the performance of a new Microsoft Outlook Plugin" regarding its speed of processing emails. The participants were told that they were part of the control (benchmarking) group which completed the same tasks as the experimental group without the new plugin. We created a realistic work scenario, where participants were asked to assume the role of a personal assistant processing a set of emails. Participants were given a briefing of the fictitious company - very close to the type of starter pack newcomers in companies are given. When the participants had processed all emails, they had to answer a short questionnaire. At the end, the experimenter conducted a short interview with each participant. After this, we briefed participants about the true aim of the study and offered to explain the security implications of macros. Since we only recruited German native speakers, we kept all study materials as well as the communication in their native language. The study materials and quotes presented within this work were subsequently translated by us.

3.1 Study Procedure

In this chapter, the consecutive phases of the study shown in figure 2 are chronologically explained in detail.

3.1.1 Recruiting. We recruited participants for the study through our personal network, but none worked at the university or the company where one of the authors was based. Potential participants had to complete a screening questionnaire, to ensure they were over 18 years old, and familiar with the Windows operating system as well and the Microsoft Office Suite. Participants did not receive payment, they were asked to 'help out a friend' of one of their relatives or friends who was conducting a research project. They were also asked to rate their computer skills, since we wanted around 50 percent of our sample to not be IT experts, and the other half to have significant IT experience and skills. At the end of the study, we informed the participants that the study was about macros in Microsoft Office, and offered to explain the risks associated with enabling them and answer any questions they might have. Most participants wanted to know and afterwards expressed thanks for the explanations and guidance on secure choices.

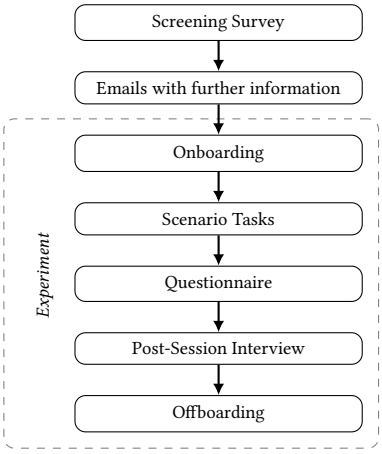

Figure 2: Overview study procedure

3.1.2 Onboarding & briefing. After participants filled out the screening questionnaire, we send them the consent form and more information about the purpose of the study. After they agreed to participate in the study, a day and time were scheduled. Each participant was sent additional information with the scenario and tasks five days before the day of the study. The email contained multiple documents: *Scenario Introduction, Company Guidelines*. Within the scenario introduction participants were explained that they should take on the role of the executive assistant Erika Müller within a chemical company. Ms. Müller normally receives her tasks from her supervisor in the form of emails. Her job is to process all relevant and time-critical emails and associated tasks for her supervisor as efficiently as possible. To balance this instruction, we also informed participants that the company had recently suffered a security breach, and invested in a major security awareness campaign to prevent this from happening again. After explaining the scenario, we re-iterated the purpose of the experiment and the steps they would work through, and that they could abandon the experiment at any time. In order to keep the scenario as close as possible to what it might look like in the real world, we sent the participants various company guidelines (*Building Security - Access control system*, *IT-Guidelines*, *Classification and handling of information*). The layout and content of the documents were very much based on what an onboarding package for a real company might look like. We advised our participants to read those documents carefully in advance of the experiment. On the day of the experiment, the participant joined the online meeting room, with the experimenter giving step-by-step instructions on how to establish the remote connection to the experiment computer. We guided participants to switch to full-screen mode to reduce the number of confounding variables. Also, the setup allowed us to create uniform experimental conditions on the remote machine. After the participants successfully connected to the remote machine, they were asked if they had any questions or concerns before commencing the tasks. Next, we

Table 1: Overview of emails provided in the study.

Email	Type	Sender	Macro	Description
E1	Informative	Work Colleague	○	Information that the finance department collected some money as a birthday gift for a colleague.
E2	Task	Superior	○	Birthday gift money should be calculated correctly and sent to the boss.
E3	Spam	Netflix	○	Information about price increase of the Netflix service.
E4	Task	Superior	●	Contained an Excel document of a list of companies' addresses and the task to fill out a missing address.
E5	Spear Phishing	Adversary	●	Supposed email from the board that a concern needed to be urgently addressed.
E6	Phishing	Adversary	●	A message that the supposed HR department had released a file, which is attached.
E7	Task	Superior	●	The task is to complete an entry in an Excel document. The required information can be found in the attached email.
E8	Informative	Superior	○	The superior informed that a payment was made.
E9	Phishing	Adversary	●	Email with an attached invoice from a well-known telecommunication provider.
E10	Phishing	Adversary	○	The email contains information about a supposedly winning prize.
E11	Task	Superior	●	The student wages of working students are to be adjusted in a document for the supervisor.
E12	Task	Superior	●	A bid with suitable names should be added to an offer, because the supervisor cannot open the document.
E13	Phishing	Adversary	●	(Extension to E5) Supposed email from the board that asks to return an Excel document as soon as possible. The document contains an instruction to activate macros.
E14	Spam	Gambling company	○	An invitation to participate in a promising lottery.

started the recording and guided the participants through the experiment. We tried to formalise the introduction as much as possible to ensure similar starting conditions for all participants.

3.1.3 Scenario tasks. Participants started by opening Microsoft Outlook and working through the 10 unread emails in their inboxes. During the first two emails, the investigator stayed available for queries. The order in which the emails were processed was left up to the subjects. After participants processed the second, third, and fourth emails, the trial provider sent emails E11-E14, respectively. Overall, the emails can be classified into three groups: *Informative*, *Tasks from the superior*, *Phishing* or *Spam Emails*. Emails from the first group do not require any action from the participant. These are purely informative and serve mainly to make the scenario more realistic and to help the participant put himself or herself in the position of the fictional character. Table 1 illustrates the different emails and a short description of their purpose.

3.1.4 Questionnaire. After participants had completed the last task, the investigator guided the participants to open the questionnaire on the remote computer. The questionnaire contained 32 questions and included questions about macros in general and about demographic information. The remaining questions asked about the alerts displayed by Microsoft Office, the participant's awareness of security risks in email, as well as two questions each about the validity of the scenario and the company policies they were given in advance. To further obfuscate the study purpose, we added some questions related to the advertised purpose of the study. We used a four-point Likert scale [27] for most questions, with the additional option to indicate "I don't know". The complete questionnaire can be found in the appendix A.1.

3.1.5 Post-session interview. After completing the questionnaire, the investigator introduced the interview session to the participant. We first asked participants whether they had trouble with processing the emails and whether they had problems solving the questionnaire. We then asked them several questions about the previously displayed warning messages. Next, participants were asked to tell us what they think macros are and what they are used for, which we explained to them directly after this. Subsequently, the investigator asked why participants ignored messages that contained advertisements, and what the participant thought when opening an email (*E5* or *E13*) that contained a description of how to enable macros. Finally, we revealed the actual purpose of the study, the security risks related to office macros, as well as the reason why it was important to obfuscate the purpose of the experiment. None of our participants expressed annoyance or anger at having been deceived. The interviews were fully transcribed by the research team. Within section A.2 the complete interview guide is shown.

3.2 Piloting

Within our research group, we piloted the study procedure (except for the interview part) with three participants who met our screening criteria. We removed one task from the scenario due to time constraints, so participants had to process 14 emails in the study. After that, we tested the whole study including the interview with one additional participant after testing just the interview guideline internally within our research group. The data of the pilot participants are not included in the results.

3.3 Analysis

At the beginning, a deductive code system was created based on the interview guide and extended with the experiences we made during the interview. Subsequently, the interviews were coded by one author, and the system was extended and adapted by inductive codes. Thus, we followed the concept of thematic analysis of Kuckartz and Rädiker [26, 33]. As recommended by the authors, we conducted multiple iterations. We used the qualitative data analysis software MAXQDA [40], which is designed to support the chosen qualitative content analysis approach. Moreover, it enabled us to work with both qualitative and quantitative data in one tool. Our final codebook is shown in section A.3.

3.4 Ethics and Data Privacy

Since our institution did not have an institutional review board or required ethics approval for this type of study, we adhered to the national and EU privacy laws. Our consent form was compliant with the European General Data Protection Regulation and covered all information usually required for US IRB approval. In addition, we repeatedly emphasised that participation is voluntary and that

participants could stop at any time, and without giving reasons. At the end, when we had explained the actual purpose of the study, none of the participants complained about the misdirections. Most participants thanked the researcher conducting the sessions for explaining the warning messages and the risks associated with macros.

3.5 Limitations

Our study has a number of limitations: Recruiting participants through our personal network connections could have introduced a systematic bias, and could have led participants to infer that the study might have been about security. Participants were recruited via a convenience sample and skewed towards young and educated, and came from the region where our university is based - this limits the generalizability of our results. The counts we report from in our qualitative analysis and the results from our questionnaire are used to convey weight, and not taken as quantitative results. Laboratory experiments may not reflect behaviour in real life; however, 33 (91.7 %) participants answered (agree or strongly agree) that they thought the study tasks reflected those they would encounter in a company context. 34 (94.4 %) said that their business emails reflected what they were presented in the study (agree or strongly agree). We cannot be sure that participants read all the provided guidelines; therefore, some may have started the session better informed than others - even though this would be the same in a real onboarding process. The answers in the interviews, as well as in the questionnaire itself, may have been influenced by the contents of the experiment given to the participants, as well as by the investigator itself. However, we tried to design the scenarios and content as they could occur in a real business environment. Although we tried to conceal the actual purpose of the study, some participants may have inferred before the end of the study what it was about; however, the results (overwhelmingly insecure choices and answers in the interviews) suggest that this was not the case.

4 RESULTS

In this section, we present the results gathered from the benchmark tasks, the questionnaire, and the interviews. In the first part, we present the analysis of the quantitative results in line with our research questions, whereas, in the qualitative part, we follow the structure of the themes that emerged from our codebook (see section A.3).

4.1 Demographics

We interviewed 36 participants. Table 2 gives an overview of participant demographics, as well as their level of IT experience. All participants were from Germany and most of them were highly educated. 27 participants identified as male (75 %), and 9 as female (25 %). Most (18; 50 %) were between 26 and 31 years old, 10 (27.8 %) were younger and 8 (22.2 %) were older. Only three participants (8.3 %) owned five or more digital devices that they were using to access their emails. 18 (50 %) owned three or four and 15 (41.7 %) one or two. Four participants (11.1 %) described their IT knowledge as none or poor, 15 (41.7 %) as fair, 16 (44.4 %) as good or very good, and one (2.8 %) as excellent. Thus, we almost achieved our goal of 50 % non-IT experts and 50 % with significant IT experience and skills. On average, the participants needed 16.7 minutes (± 3.8, min 10, max 24, n=36) to complete the questionnaire and 36.7 minutes (± 10.5, min 22, max 68, n=34 because two recordings were damaged) to complete the experiment.

Table 2: Participants' demographics and experience (n = 36).

Gender	Male	27	Female	9
Age [years]	Min.	20	Max.	38
	Median	28	Mean	28
	Std.	4.1		
Education	High school	8	College	2
	Graduate school	2	Apprenticeship	2
	Bachelor's degree	16	Master's degree	5
	Doctorate / PhD	1		
Electrical devices	one or two	15	three or four	18
	five or more	3		
IT knowledge	none	1	poor	3
	fair	15	good	11
	very good	5	excellent	1

4.2 Quantitative Results

In this section, we will present the quantitative data of the questionnaire and the quantifiable data of the interview. We will follow the structure of the questions introduced in section 1. The exact wording of the questions can be found in the questionnaire in chapter A.1.

Click rates. The participants received 14 emails with attachments, 5 of which were phishing emails - 4/5 contained a macro, the 5th was advertising without malicious content. In total, macros were activated 93 times (44.7 % of all macros) from 23 (63.9 %) participants. Phishing emails, however, were often recognised as such, and hence the included macros were only activated 39 times (25.7 %) from 17 (47.2 %) participants. We did not find any significant correlation between activating/ignoring/cancelling macros and the self-reported IT knowledge, a correct mental model of macros in the interview, or the number of correct answers regarding macros in the questionnaire. Only two participants (5.6 %) discarded the macro warnings in five emails (1.7 %) by clicking the X button. Most of the participants (29; 80,6 %) just ignored the warning, but only in 25.3 % (73) opened attachments. This shows that it was not clear to the participants how to deal with macro warnings, as macros never had to be activated to fulfill the task. Looking at the phishing emails, no participant discarded the warnings and only one (2.8 %) ignored one phishing email (0.7 %).

4.2.1 Interaction with warning messages. The participants were asked if they had noticed the warning messages and if they thought that they had to enable macros in order to work with the document. Moreover, we asked about their risk perception regarding macros.

Noticing of the warning message. When the participants were asked if they had noticed the warning messages in Microsoft Office when opening the attachments (Q2.1), 31 (86.1 %) replied that they noticed both messages. Three participants (8.3 %) noticed only the first and one (2.8 %) only the second message. There was only one participant (2.8 %) who answered that none of the messages were noticed. 33 participants (91.7 %) had already seen macro warning

messages in the past, two (5.6 %) had not, and one participant (2.8 %) was unsure.

Perceived need to activate macros. The participants were asked whether they had to click the buttons on both warning messages (activating editing and activating macros) to see the content of the document (Q2.6). 22 participants (61.1 %) answered that this statement was not true and that the content could be seen without activating editing and/or macros. 13 participants (36.1 %), however, stated that they had to activate both editing and macros in order to view the full content. One participant (2.8 %) expressed uncertainty about this question.

Relation between risk perception and clicking behaviour. When selecting the answers to the statement "Email attachments are usually harmless." (Q1.4) and the statement "After opening an Office document I had to click on the buttons of both warning messages in the upper yellow bar to see the content of the Office document" (Q2.6), it became obvious that only one participant (2,8 %) incorrectly agreed to both statements. This shows that although most participants either think that email attachments are generally not harmful, or that macros/editing have to be activated to view the content of a document, most of them are aware of at least one risk.

4.2.2 Perception of communicated risk. The participants were asked how they perceived the warning and if they understood the underlying risks.

Perceived necessity of warnings. When the participants were asked if they found the macro warning unnecessary (Q1.8), eleven (30.6 %) agreed, and one (2.8 %) did not know. 24 participants (66.7 %) disagreed and found the warning useful.

Understanding of the warnings. 28 participants (77.8 %) stated that they understood the meaning of the warning message (Q2.9). Seven (19.4 %), however, did not and one (2.8 %) did not know.

Relation between understanding and perceived necessity. When investigating the relationship between the perceived necessity of the warning messages (Q1.8) and the understanding of the warning and the underlying risk (Q2.9), the data shows that three participants (8.3 %) found the warnings unnecessary but did not understand them, and 19 participants (52.8 %) found the warnings necessary and understood them. The other participants either found the warnings unnecessary and understood them (8; 22.2 %), necessary and did not understand them (4; 11.1 %), or did not answer either one of the questions (2; 5.6 %).

4.2.3 Understanding of macros. The participants were asked several questions about the functionality and use of macros in the questionnaire. Additionally, in the interviews, we asked them to describe in their own words how macros work.

Understanding of functionality - Questionnaire. Most participants' answers on office macros work were incorrect. Figure 3 shows the eight questions in the questionnaire regarding the understanding of macros. The graph shows all answers that were either incorrect (dark grey) or the participants did not know the answer to (light grey). The chart shows that more than half of the questions were answered incorrectly. This indicates that they underestimate the risk arising from macros.

Understanding of functionality - Interview. In the interviews, we asked participants to explain how they thought macros worked. We evaluated each part of each statement if it was correct, incorrect, or if they said that they did not know. Eleven participants (30.6 %) made only correct statements about macros, and three (8.3 %) only incorrect statements. The other 22 (61.1 %) participants said that they did not know what macros were and how they were used. Of these 22 participants, 14 (38.9 % of all, 63.6 % of unknowing participants) tried to explain macros despite not being sure. These 'guesses' were 35.7 % correct, 50 % incorrect and in 14.3 % both correct and incorrect answers were given. When comparing the five most relevant questions in the questionnaire concerning the mental model of macros (Q1.5, Q1.7, Q2.3, Q3.1, Q3.6) and the answers the participants gave in the interview regarding their mental model of macros, we found a significant but only moderate correlation ($r = 0.598, p < 0.01$).

4.3 Qualitative Results

In the following, the statements and opinions of the participants are presented in more detail. The structure, as well as the content of the individual chapters, are based on the codes from the codebook (table 3). We focus on qualitative results directly related to the interaction of the warning messages and their perception.

4.3.1 Reasons for not enabling macros. Participants were asked why they had closed the macro warnings by clicking on the small cross, or why they had ignored the warning messages.

Reasons for ignoring. Eight participants stated that they had not seen a need to enable macros, because the tasks could be solved without activating them: *"Well, for the processing of the task it was not necessary to activate the macros, therefore I didn't do it. [...]"* — [P8]. One of the eight, however, did activate macros at least once during the study. All other participants did not activate any of the unneeded macros.

Unfamiliar. Two participants stated that they did not know what this warning was about, and therefore they would rather leave it deactivated: *"I think because I'm not very familiar with the subject matter."* — [P19]. P16 was scared of breaking something when clicking on the warning message: *"So, because I didn't know what it was. So, I thought, if I don't click, then I can't break it (laughs)"* — [P27].

Security reasons. Three participants cited the security risk associated with macros as a reason for ignoring the warnings: *"Yes, well, I think that there is a possibility that malware can be spread via text, via Office documents, but I was inevitably dependent on editing, therefore I practically clicked away or ignored this warning."* — [P27].

4.3.2 Reasons for enabling macros. We also asked participants why they had activated macros; the answers showed a variety of reasons, most of them related to the interaction design of the warning message.

It was necessary. About a quarter of all participants stated that they had no choice but to click (and remove) the warning messages to be able to carry on with the task. P15 stated that he had thought that it was necessary to continue to work with the document: *"So I just clicked to get on with it.[...]"* — [P15]. Participant P10 told us that he

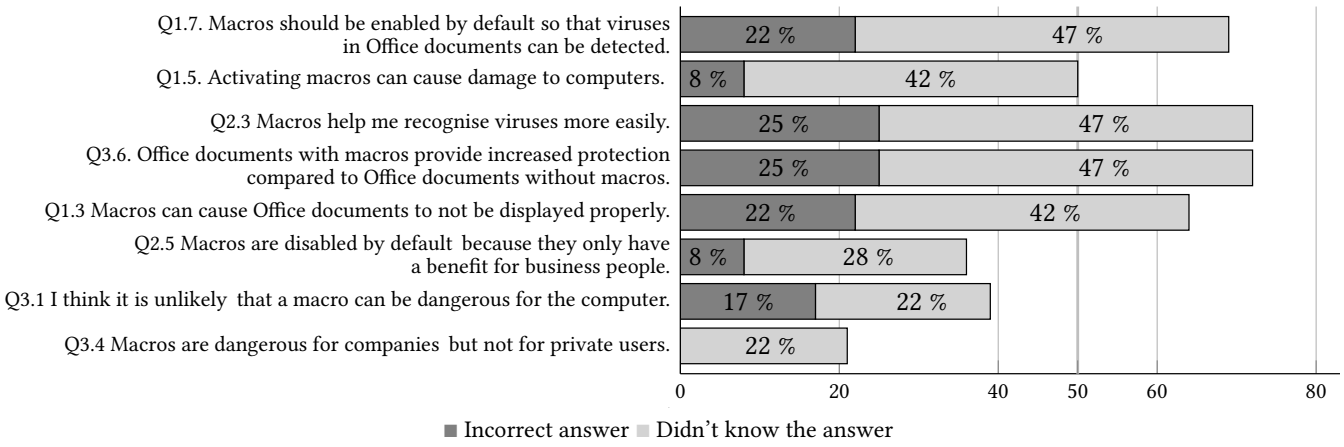

Figure 3: Results of questions on macros (n = 36)

had had the assumption that writing would not have been possible without activating macros: *"I thought I could not write then."* — [P10].

P26 stated initially recognising the possibility of danger, but after carefully reading the warning message concluded (incorrectly) that she had to click on it to continue: *"The first thought is, of course, ok, there can be a danger now, somehow, that someone wants to access my system or my computer, and when I read through it, I then realised, yes, all right, if I do not activate the things above, then I do not get on with the email or with the attachment, and then I finally decided to activate them."* — [P26].

One participant thought that the emails were trustworthy because he assumed that the IT department would filter out malicious emails: *"Yes, I actually assumed [...] that the IT somewhere filters out such things after all."* — [P21]. Also, the other six participants based their decision on the fact that they were already convinced that the email was trustworthy when they opened it, and that the content was as well. Consequently, they were not interested in any warnings.

Habits. Most participants expressed annoyance with the warning messages, and found a plausible explanation why they had to click. However, eleven participants honestly reflected that they had clicked instinctively, or out of habit. Participant P10 described to us that he had seen the warning message before and that he knew that everything worked out when he enabled it. Also, participant P26 described that he was used to activating macros in his everyday life: *"but as I said, since it is a bit common in everyday life that you activate macros, I just adopted it that way. Whether that is beneficial on the business level or not, I can't evaluate."* — [P26].

Prompt to activate. Four participants stated clearly that they understood the warnings as some kind of prompt to activate macros. One participant described it in a way that he had even felt somehow forced to click the message, although everything had been displayed correctly: *"you are just somehow forced in a way to do that, to see the content, although the content is already displayed."* — [P39].

Influenced from first warning message. P21 expressed his annoyance that he could not understand why the warning message was yellow and looked the same as the previous warning message. That led him to finally enable macros: *"[...] So, it does not appear as a warning. It's more like a request. Because it is yellow when you activate editing, I don't know, was that also in yellow? You just can't tell the difference. A warning should be highlighted in red, not in yellow, especially not if I request something else beforehand that I can use to edit something and send out a warning afterwards in the same colour. Then the difference between the two is simply not there."* — [P21].

Six participants conveyed anger and confusion about the preceding warning message. One of the most frequently mentioned reasons, apart from the external appearance, was that the content was misleading and not clearly distinguishable. Some did not understand the difference between 'enable editing' and 'enable content': *"Well, I don't know. What is the difference between enabling editing and enabling content?"* — [P31].

4.3.3 Sources for help. During the experiment, we observed that nobody clicked on the warning messages to obtain more information. We also asked participants how they would proceed to get information about the warning message. More than two-thirds said that they would look somewhere on the internet; thirteen participants would either ask a colleague or someone from the IT department. Only two participants mentioned that they would look for the original documentation from Microsoft Office within the program. And even in these cases, it was not clear from the statements that they were aware that they could easily reach the information from the warning messages themselves.

4.3.4 Perception of communicated risk. We also asked participants whether they associated any risk with clicking on the warning message. 26 Participants stated that they had not seen any risk related to this warning message, or that they had not understood what the warning message was about. One participant explained that he had thought about something like viruses when first noticing the warning message. But he did not understand the context, and then clicked on activate after all: *"So yes, I just kept thinking, as I said, whether there could be something behind it. Now, as a risk, I was thinking of viruses somehow with content activation. But I didn't see such a danger. [...]"* — [P26]. P20 reported that he had not known exactly what happened when he clicked on 'enable content': *"No, not really. Not at first but afterwards you thought, what if I activate it right away? Will something happen (laughs)?"* — [P20]. One other participant reported

that he had ignored the warning message because he claimed that he had checked the respecting email for its trustworthiness. He had already made this decision earlier: *"To be honest, I didn't because I had already made the preselection in the first place when I sent the email. So, for me, this security risk had actually already been ruled out."* — [P14]. Participant P15 described that he had not assessed the warning as dangerous because of its colour: *"Since it has a yellow background, I didn't see any risk, if something would pop up in red danger or something, but I thought nothing could happen, click on it, go ahead."* — [P15]. One other participant also argued that the size had been too small and that he, therefore, had not noticed the warning.

5 DISCUSSION

The results from the questionnaire (Q1.7, Q1.5, Q2.3, Q3.1, Q3.6) showed that 89 percent of our participants had at least one material misconception about the security risks associated with macros. The same conclusion emerges from the qualitative results, as 26 participants told us that they did not associate any risk with the warning messages. In the experimental part, 64 percent of participants activated macros unnecessarily at least once, and 47 percent activated macros in dangerous situations (phishing emails E5, E6, E9, E13). 54 percent of the participants had an incorrect or incomplete mental model of macros. There were two main things that made us conclude that the warning messages were anything but helpful in terms of security:

(1) We found that the majority of our sample had an incorrect threat model as well as an incomplete mental model for what macros are used for and how they work.
(2) We identified a number of design errors that led to a misinterpretation of the warning messages and consequential insecure behaviour.

We derived individual factors from the results and analysed them in conjunction with each other. Based on our analysis, we conclude that the design of the warning messages mislead participants and that this is likely to be one of the main reasons why users enable macros so often. Figure 4 illustrates the influencing factors and their relationships. Those factors and their interrelationships are presented in the following chapters.

5.1 Incorrect or Incomplete Mental Model

A large part of our participants did not understand how macros work, and what they are used for. Some participants believed that enabling macros was necessary to display content. One participant believed that otherwise, he would not have been able to write. Also, many participants even admitted to us directly that they did not know how macros work - so it is not surprising they did not know that enabling them might enable an attack. What users have learnt in their daily experience with Microsoft Office is that they have to click on elements in boxes to be able to proceed with their tasks.

5.2 Incorrect Threat Model

It would be reasonable to assume that a correct mental model of macros would also lead to a correct threat model. However, we had participants who understood what macros were and what they were used for but did not correctly identify the risks associated with macros. Only the participants who answered all the security questions correctly also had a correct idea of how macros worked and what they were used for. The interpretation of the warning messages strongly influences the security model, and even if people are aware of the risks related to macros, they still have to interact correctly with the warning messages. Stress, habits, and misread cues in the design can lead to clicking on the warning anyway. We were able to observe this in our data, albeit only in some cases.

5.3 The 'Click Habit'

One part of the problem we also identified we describe as a 'click habit', established by UI design over the past two decades. Users have been trained to always click 'yes', and like all frequently executed routines, this automatic becomes a habit with dangerous consequences in security contexts such as this one. Porter Felt et al. showed that the click habit users had developed as a result of unnecessary SSL warnings can be circumvented, and user attention captured with the right design choices [13]. But in the design of macro warnings, several factors combine to lead participants to click on 'enable macros'. Through interaction with the previous warning message (see above) the user is triggered to click on the second one as well. The level of awareness, alertness, and attention that is required to notice that the second warning requires a different response would require users to be 'looking for it' i.e. be in System 2 mode [22]. When working on a secondary task such as security, and in system 1 mode, a design that triggers an insecure response entraps users into making a predictable mistake, especially when under stress or time pressure. To make matters worse, both in our experiment and in the real world, users get the macro warning message even if the file actually comes from a trusted source. Habits are formed whenever we repeat an action in a certain context. Requiring the same action sequences for safe acts as well as unsafe ones is causing users to stay in System 1 mode and not notice the difference - enabling an unsigned macro should be a rare event that is carried out in System 2 mode. Our participants confirmed that they 'just click so it will work out somehow.' But not only the interaction order of the two warnings and the false-positive warnings play into the formation of the bad habit. The similarity to other warning messages and what one user described as an 'inviting button' to activate macros also encourage bad behaviour. Some participants even described that they were really tempted by the warning message to click the button.

5.4 A Comedy of Design Errors, with Tragic Security Consequences

Participants' statements provided examples of how the design of the warning messages pushed them to make insecure choices. First, the colour yellow (used for the warning) was not perceived as threatening or something they should pay attention to. The small size of the warning message compounded this further, as did the fact that the 'enable content' button seemed to be the only option available to 'do something' that would allow them to proceed with their task. The text of the warning message suggested to the participants who did not have an accurate mental model of macros and no threat model that it was okay to proceed in this way. Participants did not understand the difference from the previous warning message. Their attention was focused on the description of the button labels.

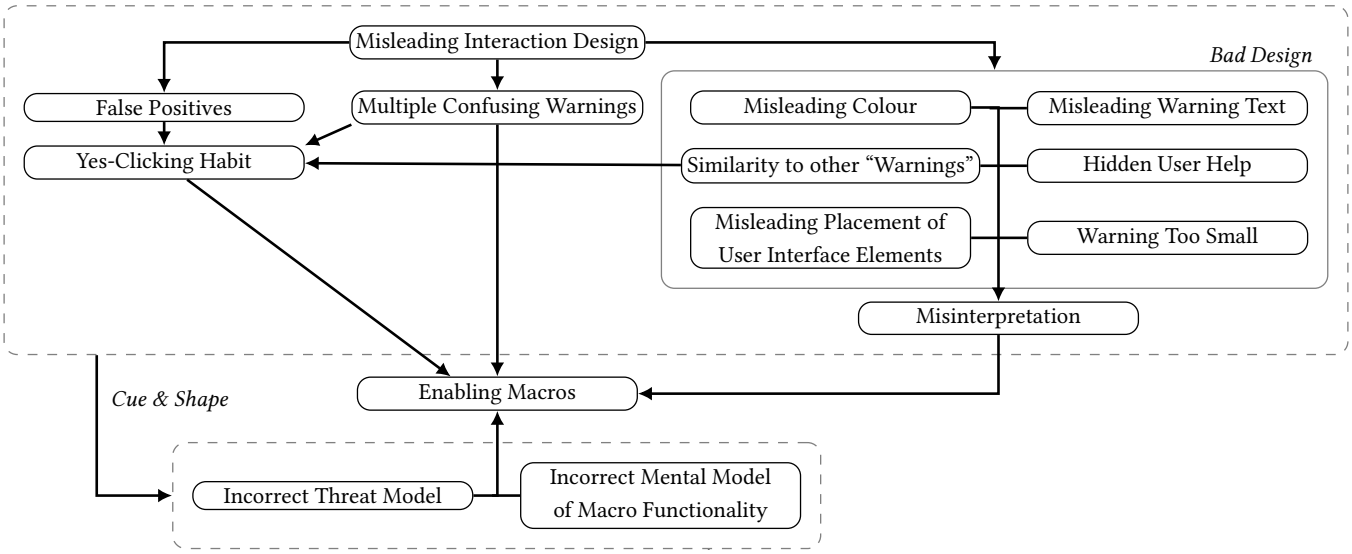

Figure 4: Overview of causes for users to enable macros

Johnson-Laird's mental models theory describes the mechanism of linguistic cueing: when instructions are too complex or confusing, people stop parsing them and just respond to cues or words they recognise as relevant in the context [20, 21]. The phrase 'enable content' might have had the effect, together with the other factors described in this section, to click the button that contained the word 'content', which participants wanted to interact with. The warning message itself is too complicated to understand (which is shown in the many misconceptions about macros), so participants just followed the cue of the button in order to work with the content. Participants found it confusing that they had to activate the first warning message, but not the second - the design led them to click through them (see section 5.3), since the warning messages look almost identical (see figure 1). Research in the past has shown that users do not read text on repeated exposure - the brain is trying to be efficient and moves on when it recognises something that is not new for it [4] - and that the 'click habit' (see section 5.3) that is activated in the previous step will trigger a click at the next one in the users - except for those who are very alert to the risks associated with macros. That they have to accept the first warning to proceed and ignore or deny the second, triggers a 'security error foretold', with potentially serious consequences.

6 CONCLUSION

The aim of our study was to follow up on indications that a significant number of users enable Microsoft Office macros without intending to, and/or realising the associated security risks. We conducted a remote study with 36 participants that was advertised as a usability test, in which they worked through a set of tasks processing emails (some with macros, and some without), complete a questionnaire, and then an interview about the choices they had made in the benchmarks tasks. We found that 64 percent of our participants enabled at least one macro, and 89 percent had at least one material security misconception about macros. The interviews revealed that the design of the user interaction triggers the 'click habit' to enable it and that the design and the text of the warning message do not alert users to the potential security risk. In addition, most participants did not know how macros work, nor the implications for security. Basically, we saw that users just clicked 'enable content' because they are used to do so and not because they are aware of the risk and needed to activate macros.

6.1 The Attacker's Perspective

By now, it becomes clear that the design is a gift to attackers seeking to entrap users into enabling macros - they just have to choose which barn door to go through. In the simplest case, the attacker does not need to do more than send an email with an attachment containing a macro. A quarter of our participants reported that they had simply pressed 'enable content' without giving it much thought. Given that our sample consisted of young, highly educated, and rather tech-savvy participants, we have to assume that in the general population, the percentage would be higher.

A second option for the attacker is to match the content of an Office document to one of the misconceptions we presented within our work. For example, we presented a document that requested the user to enable macros in order to repair the broken document. If the users do not know how macros work, they might follow the instructions presented by the attacker.

6.2 Recommendations for Industry

In many organisations and products, the usability of security mechanisms and tools are not considered - security instructions are issued and users are supposed to do as they are told, 'because security is important'. Over the past two decades, there has been evidence that users bypass unworkable security or create their own 'shadow security' practices [23]. Despite security agencies such as the NCSC in the UK highlighting that companies need to work with their employees to check that security is doable [31], companies do not

engage with their employees to talk about their security practices, their understanding, and what could be done better. Many phishing campaigns, in addition to evaluating click rates, rely on a questionnaire in order to check if the users understand how phishing works. We showed that there is a high discrepancy between the mental model that was gathered by the questionnaire and the model that the participants gave when asked to describe it in their own words during the interview. This leads to the conclusion that questionnaires are not adequate to check the understanding (of mental models) that drive their behaviour, the daily practices they engage in.

6.2.1 Engage with your users. The fact that many users may not know what macros are, that they are not aware of the risks associated with enabling them, and the possible consequences for the organisation, could have been discovered in the context of user requirement elicitation. The usable security principle is that we should fix the system, not the user - and our results certainly suggest that the UI and macro warning need to be fixed (see Section 5.4). But this is not a short-term solution; organisations that allow the use of macros should, in the short term, review the use of macros within the organisation, engage with those employees that write and use them, and improve their awareness and skills to do so in a secure fashion. This is currently not part of standard awareness packages and training that most organisations provide for their staff.

6.2.2 Disable or sign macros: company-wide, and beyond. In 2016 Microsoft introduced a feature within their corporate edition to disable macros company-wide. And for more than 20 years, it has been possible to sign macros [28] to avoid the execution of untrusted code. But not every company may be able to deactivate macros completely or convince external customers and partners to sign their macros. It would require a broader collaboration and willingness to invest resources to introduce signed macros across supply chains and organizations that interact on a regular basis - but this would not only improve security, but save user-time and attention that is currently required when interacting with legitimate, but unsigned macros.

6.3 Recommendations for Research

Within these sections we would like to share the insights we gained through our study, highlight existing research challenges, and provide ideas on how to further explore the field of Office macros.

6.3.1 Experiment design. The combination of quantitative and qualitative methods allowed us to observe the behaviour of the users as well as to investigate its causes. The approach of setting the focus of the participants on a specific topic in the interview by letting them work on a specific task beforehand worked well for us. In retrospect, we would say that if we had only conducted the experiment, the questionnaire, or the interviews individually, we would have missed some important insights and connections. Consequently, our analysis would have been much less broadly based. We noticed this concretely, for example, when we tried to understand the mental models users have of macros. We first expected to find a stronger relationship between the mental model in the questionnaire and in the interview. However, the results show that just by looking at the questionnaire we could not have inferred that the general mental model was correct. Even though the setup of the remote study demanded slightly more effort than that of a classical laboratory study, it worked surprisingly well. For our study, we were able to provide the virtual lab computer to the participant through a virtualisation on a local machine using a remote desktop application, as well as a short guided introduction. For our experiment, we set up our technical setup locally using virtualisation and then guided the participant onto the test environment. The guidance to connect to the remote computer took a bit of the experiment time away, but it helped our non-technical participants to start the experiment relaxed. It also made recruitment easier, as technical hurdles can often be a deterrent for non-technical participants. For future studies, we would advise using a cloud service, as they perform much better, are easier to set up, and more comfortable for the subjects.

6.3.2 Call for more usable security. Our results highlight the importance of conducting usability tests of security features, even on established products widely used in business. Many businesses may assume that their suppliers do this - after all, a company like Microsoft has usability labs and researchers who publish on this topic. However, suppliers may also make assumptions - namely that their customers vet and sign macros, and train their employees to understand the security implications of particular actions. Thus, problems such as the macro warnings 'fall between the cracks' in terms of investigation and mitigation, even when practitioners notice and report them informally. We need a more collaborative effort between suppliers, business organisations, and academic researchers to test and improve the usability of important security features before the attackers start to exploit them in earnest. Current developments in ransomware attacks illustrate the consequences of not doing so.

6.3.3 Follow-up work. We suspect that understanding macros is indeed also a global problem. However, our participants were outstandingly high educated and within a tiny age range. Even though our sample was sufficient for our analysis, we suggest exploring the users' understanding regarding macros and the related threat model with a greater, more heterogeneous sample. Even if the designs of the warning messages change significantly in the future, there are alternatives that do not present users with the choice of a risk assessment in the first place. Already in the early days of macro warnings, it was possible to sign and verify macros. It is unclear whether and how well users and companies can handle this. Likewise, there are some software solutions that scan targeted attachments for malicious macros and then warn the user. The resulting interaction also offers a lot of potential for analysis as well as a high risk of not being understood by the user.

ACKNOWLEDGMENTS

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2092 CaSa – 390781972. The authors would like to thank the anonymous reviewers for their valuable feedback and for helping us to improve this paper. Furthermore, we want to thank all our participants for taking the time and helping us with our research.

REFERENCES

[1] Devdatta Akhawe and Adrienne Porter Felt. 2013. Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness. In *22nd USENIX Security Symposium*, Sam King (Ed.). USENIX Association, Berkeley, Calif., 257–272. https://www.usenix.org/conference/usenixsecurity13/technical-sessions/presentation/akhawe

[2] Alison Balter. 2007. *Microsoft Office Access 2007 Security (Digital Short Cut)*. Pearson Education, London, UK.

[3] Adam Beautement, M. Angela Sasse, and Mike Wonham. 2008. The compliance budget: Managing security behaviour in organisations. In *Proceedings of the 2008 Workshop on New Security Paradigms*, Angelos Keromytis, Anil Somayaji, Christian W. Probst, and Matt Bishop (Eds.). Association for Computing Machinery, New York, 47. https://doi.org/10.1145/1595676.1595684

[4] Bonnie Anderson, Tony Vance, Brock Kirwan, David Eargle, and and Seth Howard. 2014. Users Aren't (Necessarily) Lazy: Using NeuroIS to Explain Habituation to Security Warnings.

[5] Bundesamt für Sicherheit in der Informationstechnik. 2020. Die Lage der IT-Sicherheit in Deutschland 2019.

[6] Bundesamt für Sicherheit in der Informationstechnik. 2021. Die Lage der IT-Sicherheit in Deutschland 2020.

[7] Florence A Clark. 2000. The concepts of habit and routine: A preliminary theoretical synthesis. *The Occupational Therapy Journal of Research* 20, 1_suppl (2000), 123S–137S.

[8] Jonathan Dechaux, Eric Filiol, and Jean-Paul Fizaine. 2010. Office documents: New weapons of cyberwarfare. http://2015.hack.lu/archive/2010/Filiol-Office-Documents-New-Weapons-of-Cyberwarfare-paper.pdf

[9] Will Dormann. 2016. Who Needs to Exploit Vulnerabilities When You Have Macros? https://insights.sei.cmu.edu/blog/who-needs-to-exploit-vulnerabilities-when-you-have-macros/

[10] Serge Egelman, Lorrie Faith Cranor, and Jason Hong. 2008. You've been warned: An empirical study of the effectiveness of web browser phishing warnings. https://doi.org/10.1145/1357054.1357219

[11] Karen D Ersche, Tsen-Vei Lim, Laetitia HE Ward, Trevor W Robbins, and Jan Stochl. 2017. Creature of Habit: A self-report measure of habitual routines and automatic tendencies in everyday life. *Personality and Individual Differences* 116 (2017), 73–85.

[12] Europol. 27.01.2021. World's most dangerous malware EMOTET disrupted through global action. https://www.europol.europa.eu/newsroom/news/world%E2%80%99s-most-dangerous-malware-emotet-disrupted-through-global-action

[13] Adrienne Porter Felt, Alex Ainslie, Robert W. Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettes, Helen Harris, and Jeff Grimes. 2015. Improving SSL Warnings. In *CHI 2015 crossings*, Jinwoo Kim (Ed.). ACM, New York, NY, 2893–2902. https://doi.org/10.1145/2702123.2702442

[14] Jacob Gajek. 2017. Macro malware: dissecting a malicious Word document. *Network Security* 2017, 5 (2017), 8–13. https://doi.org/10.1016/S1353-4858(17)30049-1

[15] Ronald Gallimore and Edward M Lopez. 2002. Everyday routines, human agency, and ecocultural context: Construction and maintenance of individual habits. *OTJR: Occupation, Participation and Health* 22, 1_suppl (2002), 70S–77S.

[16] Leo Garber. 1999. Melissa Virus Creates a New Type of Threat. *Computer* 32, 6 (1999), 16–19. https://doi.org/10.1109/MC.1999.769438

[17] Ann M Graybiel. 2008. Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.* 31 (2008), 359–387.

[18] Cormac Herley. 2009. So long, and no thanks for the externalities: The rational rejection of security advice by users. In *Proceedings of the 2009 workshop on New security paradigms workshop*, Anil Somayaji (Ed.). ACM, New York, NY, 133. https://doi.org/10.1145/1719030.1719050

[19] Cormac Herley. 2014. More Is Not the Answer. *IEEE Security & Privacy* 12, 1 (2014), 14–19. https://doi.org/10.1109/MSP.2013.134

[20] Philip N. Johnson-Laird. 2010. Mental models and human reasoning. *Proceedings of the National Academy of Sciences of the United States of America* 107, 43 (2010), 18243–18250. https://doi.org/10.1073/pnas.1012933107

[21] P. N. Johnson-Laird and Ruth M. J. Byrne. 2002. Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review* 109, 4 (2002), 646–678. https://doi.org/10.1037//0033-295X.109.4.646

[22] Daniel Kahneman. 2012. *Thinking, fast and slow.* Penguin Books, London.

[23] Iacovos Kirlappos, Simon Parkin, and M. Angela Sasse. 2014. Learning from "Shadow Security": Why Understanding Non-Compliant Behaviors Provides the Basis for Effective Security. https://www.ndss-symposium.org/wp-content/uploads/2017/09/01_4-paper.pdf

[24] Annette Kluge and Norbert Gronau. 2018. Intentional Forgetting in Organizations: The Importance of Eliminating Retrieval Cues for Implementing New Routines. *Frontiers in psychology* 9 (2018), 51. https://doi.org/10.3389/fpsyg.2018.00051

[25] Kat Krol, Matthew Moroz, and M. Angela Sasse. 2012. Don't work. Can't work? Why it's time to rethink security warnings. In *CRiSIS 2012*. IEEE, [Piscataway, N.J.], 1–8. https://doi.org/10.1109/CRISIS.2012.6378951

[26] Udo Kuckartz. 2014. *Qualitative text analysis: A guide to methods, practice & using software.* SAGE, Los Angeles and London and New Delhi and Singapore and Washington, DC.

[27] Rensis Likert. 1932. A technique for the measurement of attitudes. *Archives of psychology* 22, 140 (1932), 55.

[28] Microsoft. 2016. New feature in Office 2016 can block macros and help prevent infection. https://www.microsoft.com/security/blog/2016/03/22/new-feature-in-office-2016-can-block-macros-and-help-prevent-infection/

[29] Microsoft. 2021. What is Protected View? https://support.microsoft.com/en-us/topic/what-is-protected-view-d6f09ac7-e6b9-4495-8e43-2bbcdbcb6653

[30] Jens Müller, Fabian Ising, Christian Mainka, Vladislav Mladenov, Sebastian Schinzel, and Jörg Schwenk. 2020. Office Document Security and Privacy. https://www.usenix.org/conference/woot20/presentation/muller

[31] National Cyber Security Center. 2017. People: The Strongest Link. https://www.ncsc.gov.uk/speech/people--the-strongest-link

[32] David T Neal, Wendy Wood, Jennifer S Labrecque, and Phillippa Lally. 2012. How do habits guide behavior? Perceived and actual triggers of habits in daily life. *Journal of Experimental Social Psychology* 48, 2 (2012), 492–498.

[33] Stefan Rädiker and Udo Kuckartz. 2020. *Focused Analysis of Qualitative Interviews with MAXQDA: Step by Step* (1 ed.). MAXQDA Press, Berlin. https://doi.org/10.36192/978-3-948768072

[34] James Reason. 1990. *Human error.* Cambridge Univ. Press, Cambridge.

[35] Florian Roth. 2021. Microsoft Office Warnings: A Communication Disaster. https://twitter.com/cyb3rops/status/1402878336611782656

[36] M. A. Sasse, S. Brostoff, and D. Weirich. 2001. Transforming the 'Weakest L ink': A human/computer interaction approach to usable and effective security. *BT Technology Journal* 19, 3 (2001), 122–131. https://doi.org/10.1023/A:1011902718709

[37] Stuart E. Schechter, Rachna Dhamija, Andy Ozment, and Ian Fischer. 2007. The Emperor's New Security Indicators. In *Proceedings of IEEE S&P 2007, the 28th IEEE Symposium on Security and Privacy*. IEEE, Piscataway, NJ, 51–65. https://doi.org/10.1109/SP.2007.35

[38] Joshua Sunshine, Serge Egelman, Hazim Almuhimedi, Neha Atri, and Lorrie Faith Cranor. 2009. Crying Wolf: An Empirical Study of SSL Warning Effectiveness.. In *USENIX security symposium*. USENIX Association, Montreal, Canada, 399–416.

[39] United States Department of Homeland Security, Cybersecurity and Infrastructure Security Agency. 23.01.2020. Emotet Malware (Alert TA18-201A). https://us-cert.cisa.gov/ncas/alerts/TA18-201A

[40] VERBI Software. 2020. MAXQDA 2020. https://www.maxqda.com/

[41] Min Wu, Robert C. Miller, and Simson L. Garfinkel (Eds.). 2006. *Do Security Toolbars Actually Prevent Phishing Attacks?*

[42] Zarul Fitri Zaaba and Teo Keng Boon. 2015. Examination on Usability Issues of Security Warning Dialogs. *Age* 18, 25 (2015), 26–35.

A APPENDICES

A.1 Questionnaire

Q1.1. Please rate the level of difficulty of the study.

- ○ Very Simple
- ○ Simple
- ○ Neutral
- ○ Difficult
- ○ Very difficult

Q1.2. I have seen the yellow warning message(s) in the top bar when opening Office documents in the past.

- ○ Yes
- ○ No
- ○ I don't know

Q1.3. Macros can cause Office documents to not be displayed properly.

- ○ Wrong
- ○ Rather wrong
- ○ Rather correct
- ○ Correct
- ○ I don't know

Q1.4. Email attachments are usually harmless.

- ○ Wrong
- ○ Rather wrong
- ○ Rather correct
- ○ Correct

○ I don't know

Q1.5. Activating macros can cause damage to computers.
○ Wrong
○ Rather wrong
○ Rather correct
○ Correct
○ I don't know

Q1.6. I use the following email programs on my computer:
○ Apple Mail
○ IBM Notes
○ Microsoft Outlook
○ eM Client
○ Mozilla Thunderbird
○ Windows Mail
○ Other *(Please don't insert smartphone apps)* : _____

Q1.7. Macros should be enabled by default so that viruses in Office documents can be detected.
○ Wrong
○ Rather wrong
○ Rather correct
○ Correct
○ I don't know

Q1.8. I find the yellow office warning messages in the top bar to be unnecessary.
○ Totally disagree
○ Rather disagree
○ Rather agree
○ Totally agree
○ I don't know

Q1.9. The study was very close to a real business scenario.
○ Totally disagree
○ Rather disagree
○ Rather agree
○ Totally agree
○ I don't know

Q2.1. Did you notice the yellow Office warning messages in the top bar?
○ I only noticed the first warning message, but not the second
○ I only noticed the second warning, but not the first
○ I noticed both warning messages
○ I did not notice both warning messages
○ Other

Q2.2. *Note:* When opening some Office documents, a tutorial was displayed, describing how the user can repair/open the corresponding Office document. Did the instructions help you to repair/open the Office document?
○ I did not see any instructions
○ I followed the instructions and was able to repair/open the Office document
○ I followed the instructions but could not repair/open the Office document
○ I saw the instructions but did not follow them
○ Other: _____

Q2.3. Macros help me recognise viruses more easily.
○ Wrong
○ Rather wrong
○ Rather correct
○ Correct
○ I don't know

Q2.4. The computer may take damage if files with the following extension are opened (matrix question):
[$.pdf$, $.exe$, $.zip$, $.mp3$, $.ppt$, $.html$, $.lnk$, $.jpg/png/bmp$, $.doc/docm/docx$, $.txt$, $xls/xlsm/xlsx$]
○ Wrong
○ Rather wrong
○ Rather correct
○ Correct
○ I don't know

Q2.5. Macros are disabled by default because they only have a benefit for business people.
○ Wrong
○ Rather wrong
○ Rather correct
○ Correct
○ I don't know

Q2.6. After opening an Office document I had to click on the buttons of both warning messages in the upper yellow bar to see the content of the Office document.
○ Wrong
○ Rather wrong
○ Rather correct
○ Correct
○ I don't know

Q2.7. Please specify your favorite operating system.
○ Windows
○ Linux
○ macOS
○ Other: _____

Q2.8. Please provide as accurate an assessment as possible (matrix question).
Please indicate how many Office documents you open daily.
Please indicate how often you send emails with Office documents as attachments each week.
Please indicate how many emails you read daily.
Please indicate how many emails you write daily.
○ 0
○ 1-5
○ 6-10
○ 11-15
○ More than 15

Q2.9. I have understood what the yellow Office warning messages in the top bar mean.
○ Totally disagree
○ Rather disagree
○ Rather agree
○ Totally agree
○ I don't know

Q3.1. I think it is unlikely that a macro can be dangerous for the computer.
○ Totally disagree
○ Rather disagree
○ Rather agreee
○ Totally agre
○ I don't know

Q3.2. You received the company guidelines prior to the study. Did you follow these guidelines when processing the emails?

○ Yes
○ No
○ Partially
○ I don't know
Please indicate why you did/didn't follow the company guidelines:

Q3.3. How would you recognise a fake email?
*Please provide bullet points only:*_____

Q3.4. Macros are dangerous for companies, but not for private users.

○ Wrong
○ Rather wrong
○ Rather correct
○ Correct
○ I don't know

Q3.5. I can certainly imagine that employees in real companies are confronted with business emails like those shown in this study.

○ Totally disagree
○ Rather disagree
○ Rather agree
○ Totally agree
○ I don't know

Q3.6. Office documents with macros provide increased protection compared to Office documents without macros.

○ Wrong
○ Rather wrong
○ Rather correct
○ Correct
○ I don't know

Q3.7. Your inbox also contained emails that had no connection to the company. How did you deal with these emails?

○ I deleted the emails and marked them as spam
○ I ignored or skipped the emails
○ I forwarded the emails
○ I replied to the emails
○ Other: _____

Q3.8. Should a study like this be repeated to find out more about the efficient handling of business emails?

○ Yes
○ No

Q4.1. How many electrical devices do you own with which you have access to your emails?

○ 0
○ 1-2
○ 3-4
○ 5 and more

Q4.2. How would you rate your IT-knowledge?

○ No knowledge
○ Barely any knowledge
○ Little knowledge
○ Good knowledge
○ Very good knowledge
○ Excellent knowledge

Q4.3. What is the highest level of school you have completed or the highest degree you have received?

○ Still in school education
○ No graduation
○ High school Diploma
○ College
○ Bachelor's degree
○ Master's degree
○ Diploma
○ Doctorate degree/PhD
○ apprenticeship (1 years)
○ apprenticeship (2 years)
○ apprenticeship (3 years)
○ Prefer not to answer
○ Other: _____

Q4.4. Please enter your age in years: _____

Q4.5. Please indicate which gender you feel you belong to.

○ Prefer not to answer
○ Male
○ Female
○ Non-binary
○ Other: _____

Q4.6. Here you have the possibility to send us questions or suggestions.
If you do not want to tell us anything, you can leave this field empty.

A.2 Interview guide

Question category *Introduction*

(1) Did you have any difficulties with the emails?
If yes: What difficulties did you have with the emails?
(2) Did you have any difficulties with the questionnaire?
If yes: What difficulties did you have with the questionnaire?

Question category *Warning message*

(1) What was your first thought when you saw the second office warning message?
(2) What do you think the warning message was about?
(3) Why does the yellow warning message appear in the top bar when opening some Office documents?
(4) What was the reason you clicked on the warning message?
(5) Did you read the text of the warning message?
(6) Did you have a choice at that moment?
(7) Did you associate a risk with the warning message?
(8) Are you familiar with the warning message in a professional context?
(9) What would be your first idea to get more information about the warning message?

Question category *Macros*

(1) What is a macro?

Question category *Emails*

(1) Why did you ignore or delete the ad-like emails?
(2) What was your first thought when you saw the instructions to repair files in some emails?

A.3 Codebook

Table 3: Codebook

Code	Description	Example Quote
Style and design	Statements about the style or the design of the warning messages	*But you think it's trustworthy because it somehow looks like an Office product and had the normal design or also the col [...] such a [...] simply so plausibly stood there, in the place where it otherwise also stands [...] (P16)*
Reasons for activating macros	Reported reasons for enabling macros in the experiment	*I assumed that I would not have been able to edit it then. (P20)*
Reasons for not activating macros	Reported reasons for not enabling macros in the experiment	*Yes, ok, I knew I didn't need macros for this [...] (P16)*
Perception of risk	–	
Risk & careful behaviour	Participants stated directly that they had acted carefully or that they had perceived a risk related to the warning message	*[...] when I am at work and open documents, generally somehow some risk. When I open documents. Whether there is a security warning or not, yes, a little (P20)*
No sense of risk	Participants described that they had not seen any risk related to the warning message	*From the gut feeling, rather no. [...] (P22)*
Perception of choice	–	
Safe choice	Participants described one of the three alternatives: clicking the 'x', ignoring the warning, or closing the document	*Not to click. (P38)*
No choice at all	Statements that implicate that participants had not seen an alternative to enabling macros	*Because if I hadn't clicked on it, then I wouldn't have been able to edit it. (P10)*
Unsure	Participants described directly that they had been unsure which choices they had had	*I'm just being honest, I don't think I even know. (P30)*
Ways of information retrieval	–	
Online search	Any online search method participants described to look for more information about the warning	*I would google, honestly. (P7)*
Asking for help	Participants described that they would ask someone for help	*And I would perhaps also write to our IT department and see what they have to say about it. (P8)*
Official documentation	Participants stated that they would look for help in the official documentation from Microsoft	*[...] There is also the help, there you can look through Office itself. (P21)*
Macros	–	
Correct mental model	Correct elements of answers on how macros work or what they are used for	*[...] that macros are some kind of series of commands [...] So, you can, for example, program it in Excel [...] (P8)*
Incorrect mental model	Incorrect elements of answers on how macros work or what they are used for	*A macro, that's probably some kind of anti-virus program. [...] (P10)*
Unaware	Participants stated directly that they did not know how macros worked or what they were used for	*Yes, well, I did not know what macros were [...] (P5)*