Exploring the Effects of Cybersecurity Awareness and Decision-Making Under Risk

Jan Hörnemann[1][0009−0001−8503−9654], Oskar Braun[1][0009−0001−8503−9654],
Daniel Theis[2][0009−0009−2764−1936] Norbert Pohlmann[2][0009−0007−6221−7327],
Tobias Urban[2][0000−0003−0908−0038], and Matteo
Große-Kampmann[1,3][0000−0001−9127−969X]

[1] AWARE7 GmbH firstname@aware7.de
[2] Institute for Internet Security, Westphalian University of Applied Sciences
lastname@internet-sicherheit.de
[3] Rhine-Waal University of Applied Sciences
matteo.grosse-kampmann@hochschule-rhein-waal.de

Abstract. This paper challenges the conventional assumption in cybersecurity that users act as rational actors. Despite numerous technical solutions, awareness campaigns, and organizational strategies aimed at bolstering cybersecurity, these often overlook the prevalence of non-rational user behavior. Our study, involving a survey of 208 participants, empirically demonstrates this aspect. We found that a significant portion of users (55.3%) would accept a substantial risk (35%) to click on a potentially malicious link or attachment. This propensity increases to 61% when users are led to believe there is a 65% chance of facing no adverse consequences. To address this irrationality, we explored the efficacy of nudging mechanisms within email systems. Our qualitative user study revealed that incorporating a simple colored nudge in the email inbox can notably enhance the ability of users to discern malicious emails, improving decision-making accuracy by an average of 10%.

Keywords: Economics of Cybersecurity · User Behavior · Behavioral Economics.

1 Introduction

Threats to internet-facing users and systems are manifold. Ransomware, spam, fraud, and malware delivery are just a few to be named. The delivery vector email is a threat to users and organizations especially. Human users are often emphasized and framed as *the last line of defense*, yet little is known about the economics of decision-making in cybersecurity. Malicious actors use different delivery vectors for various kinds of illicit activities, ranging from stealing data and compromising single machines to whole networks and compromising the privacy of victims. The victims do not recognize that they are victims of fraud because the attack is deceptive. Thus, building awareness among users is essential when protecting modern information systems. This paper aims to understand how decision-making under risk is done and how awareness measures help users improve these decisions. We do this using an online survey, in which 208 participants took part. Furthermore, we wanted to understand how users perceive the warning marker. Therefore, we conducted a qualitative user study asking 31 participants to determine if a mail is malicious or legit. We used a simple color nudge during our experiment to evaluate the effectiveness of this approach and whether this changed the participant's detection capabilities. Our results show that participants perform better regarding email classification if they are nudged in their inbox and that a misclassified email, whether false negative or false positive, is correctly classified.

In summary, we make the following contributions:

- To the best of our knowledge, this is the first work to empirically analyze behavioral economics in cybersecurity with a focus on decision-making under risk and stress.
- We show that 'stress' affects the chances that users might click on malicious emails and that awareness measures help users perceive themselves or their organization as more secure.
- We empirically show that more than half of our surveyed users would take a risk in clicking a potentially malicious link or attachment, and if they are framed to believe there is a chance that nothing will happen, the share of risk-takers rises.
- We conducted an experiment with a follow-up survey that shows that placing nudges in email inboxes helps users decide if an email is malicious or legitimate. It also seems to raise their confidence in their detection capabilities.

2 Background

Before we describe our approaches to determine the effects of cybersecurity awareness and decision-making under risk, we briefly provide the background information necessary to follow our methods.

2.1 Human User

Several papers describe the users of information processing systems as the weakest link [38], and human-centered cybersecurity is getting more and more attention nowadays. Cybersecurity decision-making is similar to other kinds of decisions, but cybersecurity decisions have distinctly other features. Security and Risk in themselves are intangible concepts, especially in the cyber domain invisible to users. As Schneier states: "*Security is both a feeling and a reality. And they are not the same*" [39]. For example, the presence of a TLS warning is often not enough to stop users from visiting a website anyway [2]. *Human Behavior and its Economics.* Behavioral Economics is the combination of psychology and economics. It considers human limitations and complications and determines what happens if these humans make decisions within a market [29]. It is important to note, that these models extend the predominant equilibrium and rational choice models [18]. Conventional decision-making, however, describes the trade-off between expected return and risk by combining risk and return calculations [27]. This result translates to the following: A decision maker in cybersecurity (user) will invest in cybersecurity if it yields a positive return under rational risks, chances, and returns. While this has various implications for the actual cost management of security [13] and investments into security [5, 7] it also has implications for cybersecurity decisions that users make. As Pfleeger and Caputo point out, security is "*rarely the primary task of those who use information infrastructure*" [33]. Kahnemann and Tversky defined prospect theory in 1991 [42]. The central argument of the prospect theory is, that humans do not have the underlying objective probabilities by which they measure gains and losses. They weigh the gains and losses (or their value) with a nonlinear transformation of these probabilities. The general assumption of loss aversion theory, as described by Tversky et al. [42], is that losses and/or disadvantages have a greater impact on preference than potential gains and advantages.

2.2 Cybersecurity Awareness

Awareness, in general, can be defined as "*knowledge that something exists, or understanding of a situation or subject at present based on information or experience*" according to the Cambridge Dictionary [6]. Cybersecurity awareness can thus be seen as the level of understanding, knowledge, and timely appreciation of cybersecurity aspects by an individual or a group. By researching websites of cybersecurity awareness providers and conducting a literature review, we manually identified the following four measures to raise awareness.

- **Live Hacking** Live Hacking is an event format. One to two hackers perform various pre-planned attacks on stage to raise awareness of threats on the Internet. The duration ranges from 30 to 90 minutes.
- **Phishing Simulation Campaign:** A phishing simulation campaign is a cybersecurity activity characterized by the fact that participants are not necessarily aware of the activity. Selected employees are sent phishing emails at irregular intervals. Afterward, the type of phishing emails the employees could recognize, best or worst, is evaluated.
- **Seminar / course / workshop:** This measure involves the active development of learning content. For this, a group size of approx. 15–30 participants are advised, as all participants should actively partake. Often, this measure lasts one or more days, with the possibility of receiving a certificate or similar.
- **eLearning:** By eLearning, we mean a digital learning platform with short videos that explain various topics are available. Short tests and quizzes at the end of the different lessons test the participants' knowledge: In and consolidate the other contents.

3 Methodology

In this section, we introduce the two user studies that we conducted in our work. We conducted an online survey (see Section 3.1) and an experiment with a follow-up survey (see Section 3.2) to obtain results about behavioral changes in cybersecurity awareness and decisions under risks. We wanted to elicit factors that contribute to cybersecurity perception and behavior accordingly. Our studies are based on the *Protection Motivation Theory* (PMT) [26]. PMT clarifies the cognitive processes that emphasize protective behavior in the event of threats. There are two approaches when facing a threat: (1) Focusing on the threat itself and (2) mitigation options (threat appraisal and coping appraisal). Based on the outcome of this assessment, humans adapt their behavior. Our survey focuses on the threat appraisal and the user study (see Section 3.2) on the coping appraisal.

3.1 Online User Survey

Our quantitative online user study aimed to understand how awareness measures and stress might affect risk tolerance and whether this decision is manipulated if framed otherwise. For most questions, we used 5-point Likert scales, and for the remainder, we used single and multiple-choice or open-ended questions. We provided 5-point Likert scales (ranging from "5 – Strongly Agree" to "1 – Strongly Disagree") [24] and an "I prefer not to answer" answer option. For further reference, we uploaded the survey online[4]. We conducted a pre-study ($n = 99$) in Germany. The goal of the pre-study was to understand which awareness measures users know. We used the study's results to compile a list of the four best-known awareness-raising measures (see Section 2.2). The participants in this sample can be seen as representatives of the general population. Based on the pre-study results, we dropped questions that were too detailed, e.g., regarding specific awareness measures that were not known to the participants in the pre-study. Further, we clarified the wording of the questions. Moreover, we dropped questions in which the participants did not provide meaningful insights, i.e., the respondents all selected "I do not know" or "I am not sure". Based on the feedback from the pre-study, we structured our final survey in three blocks (`I`–`III`). Block I focuses on "General information about awareness-raising measures" and the participants' knowledge of different awareness measures. We surveyed how well-known these measures are after introducing four different awareness-raising measures (see Section 2.2). Afterward, we asked the participants whether they already participated in such a measure. If not, they were forwarded to block `IIa` and otherwise to block `IIb`. Block `IIa` aims to determine why participants did not take part in an awareness training and how the measures would need to be changed so that the participants would take part. Furthermore, we asked participants to provide their stress level when they use the Internet privately or professionally and to estimate the level of cybersecurity their employer has. Block `IIb` tries to determine how participants perceive the effectiveness of the

[4] `https://anonymous.4open.science/r/scisec2024-1FF5/appendix.md`

awareness-raising measures they participated in. After they answered this, they were asked the same questions as block `IIa`. Block `III` elicits each participant's assessments of cybersecurity. Finally, demographic data of the participants is collected. To analyze the participants' willingness to take risks without any previous experience, we asked them two hypothetical questions concerning their decision-making under risk. Half of all participants (randomized by the survey tool) received the response options as a single-choice question worded to emphasize risk. The other half receives response options that emphasize the chance of not being compromised. The two questions on risk-taking differ in terms of time because the first question, which each participant receives, would have immediate consequences. The second question, however, would have the consequences occur after a certain period. The different time frames allow us to evaluate whether the timing of the possible consequences influences the participant's risk-taking. The following two scenarios were used:

Scenario 1: Imagine the following situation: You receive mail that looks like it is from your bank. There is a link in it. You click on the link and see a website that looks familiar. You are prompted to enter your username and password. Your primary goal is to check your account balance. What do you do now?

Answers for group 1: a) Do not enter username and password and do not check balance; b) Enter username and password and take a 35% risk of data being stolen immediately.

Answers group 2: a) Do not enter username and password and do not check balance; b) Enter the username and password and take a 65% chance that the data will not be stolen.

Scenario 2: Imagine the following situation: You work in a human resources department and receive a direct email with an application attached from a potential employee. If you do not open it, you risk a candidate who applied to the company dropping out and facing consequences because you failed as a recruiter. Your productivity is measured mainly by the number of candidates who reach the second step of the HR process.

Answers for group 1: a) Do not open the attachment and do not evaluate the applicant; b) Open the attachment and take a 25% chance that the attachment will endanger your job at a later time.

Answers for group 2: a) Do not open the attachment and do not evaluate the applicant; b) Open the attachment and take a 75% chance that the attachment will not endanger your job at a later time. We recruited participants using Amazon's *Mechanical Turk* (MTurk). We only accepted participants who are at least 18 years old with more than 100 tasks completed and high task completion rates ($\geq 75\%$) from around the globe. We asked for their consent to participate in our survey and disclosed our names, affiliations and all sponsors. We used Google Forms to conduct the survey and an instance of Google Workspace where the data location is set to EU. The workers received $1.40 for completing the study, and it took them, on average, 7 minutes (median: 6 minutes) to complete the survey. We saved all answers pseudonymously using MTurk's random unique string to pay the workers. Afterward, we deleted the string to increase the participants' level of anonymity.

3.2 Experiment with follow-up survey

To get a deeper understanding of how stress affects behavior and decision-making in situations that directly influence cybersecurity, an experiment with a follow-up survey was performed. We especially wanted to understand the impact of stress on security-relevant activities like categorizing whether an email is malicious or not. This scenario was the focus of Scenario 2 (cf. Block II, see Section 3.1) Therefore, this experiment explores how subjects identify potentially harmful emails when under stress. Our objective was to assess users' accuracy in detecting such emails without any prior indication of their malicious nature and how stress conditions influence this accuracy.

Study Design. The participants were all located in Germany. The study presented them with an email inbox simulation resembling real-world interfaces (Thunderbird). It was conducted online using a video tool (Zoom), with researchers guiding participants through the tasks. Four unique email datasets and a corresponding questionnaire were utilized for this purpose. The questionnaire, maintained by the researcher guiding the task, ensured a consistent interview process. *Introduction and Methodology.* Initially, to establish a baseline understanding for all participants, we briefed them on identifying malicious emails, referencing the German Federal Office for Security in Information Technology's "3 second check" [15], which includes checking the sender, subject, and attachments. These markers to detect phishing in an email are also recommended by the UK's National Cybersecurity Council [30], the French Agence nationale de la sécurité des systèmes d'information [1], and the French Commission Nationale de l'Informatique et des Libertés [10]. To simulate a stressful environment, participants were given only five seconds to classify each email in the datasets as malicious or not. This time constraint was based on findings by Li et al. [23] and supported by the work of Van der Heijden and Jalali [20,43] on stress and email management. *Survey Process.* The process began with a Likert scale evaluation of participants' confidence in their ability to detect malicious emails, both with and without technological assistance. Following this, participants were asked to classify emails from each dataset as malicious or not, using a binary "Yes/No" system. *Datasets.* The four datasets, each containing ten emails, were designed to assess different scenarios.

- **Dataset 1:** A standard inbox with unmarked emails.
- **Dataset 2:** Emails marked according to a color scheme, differentiating between malicious and legitimate emails.
- **Dataset 3:** With false positives (legitimate emails misclassified as phishing).
- **Dataset 4:** With false negatives (phishing emails misclassified as legitimate).

Afterward, participants also evaluated the utility of the color scheme and their trust in protection against malicious emails. *Participant Recruitment and Data*

Handling. We recruited participants via social networks associated with the authors and their institutions. Eligible participants were over 18 and consented to participate in the study. The study was unpaid, with an average completion time of 15 minutes. Data was stored pseudonymously to maintain confidentiality.
Data Analysis. Data analysis was conducted using the Matthews Correlation Coefficient (MCC), as recommended by Powers [34] and Chicco [8]. The MCC was selected due to its effectiveness in handling imbalanced datasets and its capacity to simultaneously minimize false positives and negatives while maximizing true positives and negatives. Throughout the paper, we use the *Analysis of variance* (ANOVA) test with a 95% confidence interval ($\alpha = 0.5$) to find statistically significant differences between the measures of independent groups.

4 Results

In this section, we provide an overview of the results. First, we present the result of the online user study (see Section 4.1), and then we discuss the findings of the quantitative interview study (see Section 4.2).

4.1 Online Survey Results

In March 2021, 208 participants participated in our survey, which we recruited via Amazon's Mechanical Turk. The following describes the main results.

Table 1. Demographic overview of the online user study.

Region	Participants	
North America	90	43%
Asia/Oceania	68	33%
South America	27	13%
Europe	14	7%
Africa	4	2%
Others	3	1%
Not specified	2	1%

Age	Participants	
Under 25	15	7%
25–35	139	67%
36–46	32	15%
47–57	11	5%
Over 57	11	5%

Demographics Table 1 shows the demographics of the survey participants. The analysis of the localization of the 208 participants shows a strong predominance of North America and Asia/Oceania. Male participants are over-represented in our study, which is common if participants are recruited via MTurk. Many studies have examined gender distribution. One highly regarded study, which analyzed 24 European countries over 18 years, showed that there are approximately 105–107 male newborns for every 100 female newborns [9]. Also, common for

MTurk studies, the age distribution shifts towards 25–35-year-olds (66.8%). Our recruitment approach has primarily reached participants who are strongly confronted with IT. More than three-fourths (82%) of the participants state that they are firmly or even very intensely involved with IT professionally. Therefore, the results of this survey are shifted to the professional sector.

Block I In this block, we want to know how familiar our participants are with different cybersecurity measures. We ask what measures they know and how often they participated in any of those events. For all four queried event formats (i.e., live hacking show, course / seminar / workshop, phishing campaign, eLearning), an average of 63% indicated that they had a rough or even a firm idea of the format. The detailed numbers are as follows: Live Hacking (mean: 3.57; median: 4; SD: 0.84), Phishing Campaign (mean: 3.67; median: 4; SD: 0.84), Seminar / Course / Workshop (mean: 3.75; median: 4; SD: 0.92), and eLearning (mean: 3.92; median: 4; SD: 0.96). eLearning, in particular, is familiar to participants, with 73% (151) indicating that they have at least a rough idea of this awareness-raising measure. The most unknown measure in this survey is live hacking. 116 participants (56%) answered with the statement that they had at least a rough idea. Although various awareness events are very well known, only 14% (28 out of 208) have participated in more than two of such events. Only 38 participants (18.3%) have not yet taken part in any awareness-raising event, mainly because they did not receive an offer. This high participation rate shows that cybersecurity is gaining popularity worldwide. By far, most participants (142; 68%) stated that they had taken part in one or two such measures.

Block IIa This section addresses the participants who indicated in the first section that they did not attend an awareness event. *Perception of security in the workspace.* Participants rate the cybersecurity of their own company as secure (71%), although some participants rate it as extremely poor (11%) or extremely good (11%) (mean: 3.29; median: 3; SD: 1.14). *Awareness of participants whether they clicked on malicious mail.* The high level of interest and self-assessment about knowledge in the area of cybersecurity is precise, as almost every participant who indicated in the first section that they did not attend an awareness event is sure whether he/she has already clicked on a malicious or fraudulent email (see Figure 3; 1: “No, definitely not; 5: “Yes, fully”; mean: 3.29; median: 2; SD: 1.45). The standard deviation of almost 1.5 shows that most participants tend strongly in one direction, whether fraudulent emails have already been clicked.

Stress when using the Internet. The stress rating shows that stress perception is equally distributed across the 38 participants' private and professional Internet use. It is noticeable that the Internet's professional use hardly influences the participants' stress levels. *Decision-making under risk.* The participants' willingness to take risks shows a clear difference in how the question or scenario is formulated. As explained in Section 3.1, all participants were asked the same question. The difference is that half of them are framed on the opportunity and the other half on the risk. In total, 38.2% of the participants would accept the presented risk and open the unknown email attachment. However, this is split between the two groups with the differently worded questions. The first group received the questions focusing on the risk, while the second group received the questions focusing on the chance that no bad result would occur. This formulation affects the users as 72% would not take the risk, and thus only 28% would accept the risk. The other half of the participants received a similar question, but it was formulated to focus on the chance that no damage would occur. As a result of this formal change, more than half of the participants, 53%, would now accept the risk. This change is significant ($t = 2.774$, $p < 0.01$). Thus, rephrasing the question to focus on the chance of no harm rather than the risk has a significant effect: Participants are more inhibited from taking risks. If we want to persuade other people not to take a particular risk, the risk should be addressed, not the chance that nothing will happen.

Block IIb This section addresses only those participants who indicated in the first survey section that they had attended at least one awareness event. In total, this section has 170 participants. *Perception of instructional and entertaining aspects of awareness measures.* In particular, eLearning was rated as "very instructive" by 44% of the participants. The detailed numbers are as follows: Live Hacking (mean: 4.02; median: 4; SD: 0.96), Phishing Campaign (mean: 4.04; median: 4; SD: 0.95), Seminar / Course / Workshop (mean: 4.11; median: 4; SD: 0.91), and eLearning (mean: 4.27; median: 4; SD: 0.86). The assessments of the individual measures can also be seen in Figure 1.

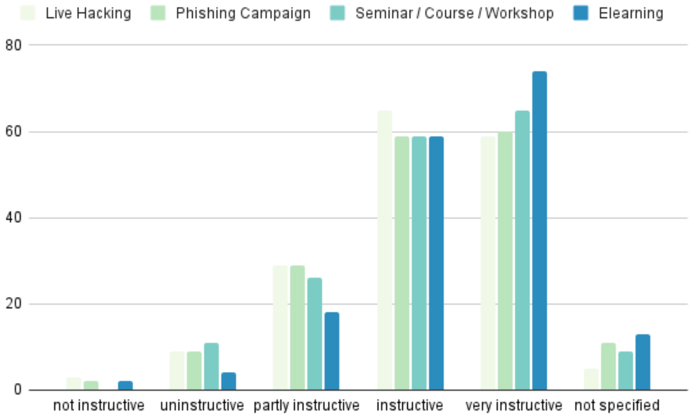

Fig. 1. Perception of instructiveness of the various cybersecurity awareness measures

Perception of cybersecurity in the workplace. The participants assess the cybersecurity of their own company as secure (71%), whereas no participant assesses it as extremely poor (1: "very poor", 5: "very good"; mean: 3.94; median: 4; SD: 0.81). Let us compare the views of participants who have already taken part in at least one cybersecurity measure with the participants' without cybersecurity measures. It is noticeable that their interest and the company's cybersecurity are assessed as better/higher. The participants who have already participated in a cybersecurity event rate their interest and the company's security better or higher. This observation shows the positive perception of such events among the participants. The estimated company security difference is statistically significant ($t = 4.13$, $p < 0.00001$). Likewise, the estimate about one's interest in cybersecurity is statistically significant ($t = -2.77$, $p < 0.01$).

Stress when using the Internet. Figure 2 shows the stress level in a private and professional context. It is noticeable that the participants are slightly stressed and that this assessment hardly changes between the professional and private contexts. This difference is *not* statistically significant ($t = 0.042$, $p = 0.966$).

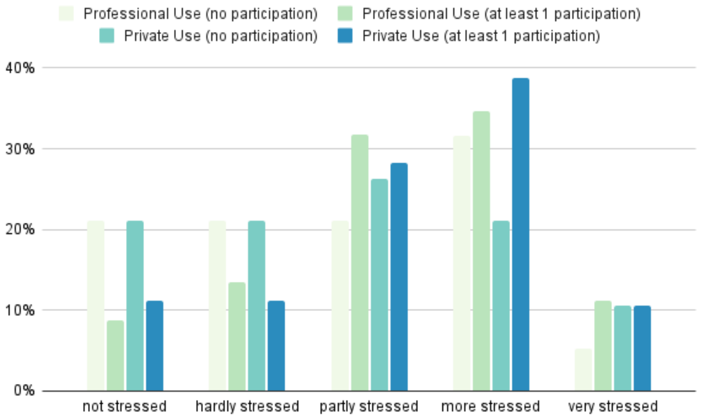

Fig. 2. Perception of stress in the private and professional context

The stress perception between the participants without participation in a cybersecurity measure and those of the participants with at least one participation shows that the most significant proportion is rather stressed. This perception is also seen in Figure 2 and is consistent with a study from European Neuropsychopharmacology, from June 2020 [44]. We found a correlation between "stressed" participants and the assessment of clicking on fraudulent emails. According to their judgment, participants who are more stressed in their private or professional Internet use are significantly more likely to click on fraudulent emails ($t = 3.39$, $p < 0.001$); see Figure 3.

Decision-making under risk. The participants' willingness to take risks shows a clear difference in how the question or scenario is formulated. As explained in Section 3, the participants were asked the same question, but the answers were

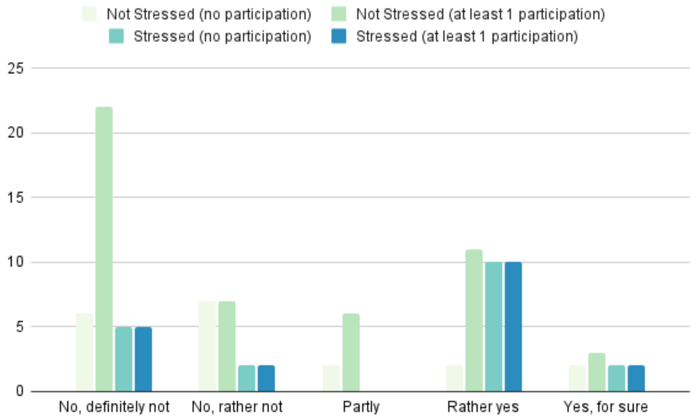

Fig. 3. Participant assessment whether they clicked on fraudulent emails. Differentiating stressed and non-stressed participants.

framed differently: either on opportunity or risk. In our survey, 59.1% of the participants would accept the presented risk and open the unknown email attachment. The difference between the participants who did not participate in any event is vast. Only 38.2% of these participants would take this risk and are thus much more cautious than the experienced participants. However, 59.1% is divided between the two groups with different phrasing. The first group received the questions focusing on risk. This rewording also affects these participants because now only around 55.6% would take the risk. The other half of the participants were asked a similar question with the difference that it focused less on the risk and was phrased so that it addressed the high chance that no damage would occur. As a result of this formal change, shown in Table 2, significantly more than half of the participants (62.4%) would accept the risk.

Combination of Blocks IIa and IIb Here, we summarize the answers from all participants, regardless of which subgroup they were in. *Connection between stress and perception of security.* The stress level of the participants has an impact on the perception of the cybersecurity posture. The difference between private and professional use is not statistically significant ($t = 0.042$, $p = 0.966$). The participants who indicated they were “stressed” (i.e., rated their stress as 4 or 5) about using the Internet in either a personal or professional context rated their company’s cybersecurity as follows: very bad: 0%; bad: 2.4%; average: 28.8%; good: 45.3%; very good: 23.5%. Figure 4 shows that participants who have already taken part in a cybersecurity event rate the company’s security as well as their interest better or higher. The difference in the estimated company security is statistically significant ($t = 4.13$, $p < 0.0001$). Likewise, the estimate about

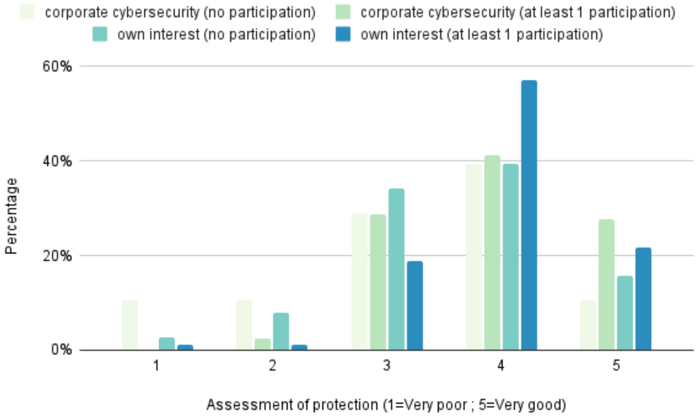

Fig. 4. Perception of company's cybersecurity

the participants' interest in IT security is statistically significant ($t = -2.77$, $p < 0.01$).

User risk perception. In summary, 55.3% of all participants would accept the risk. This refers to both groups of participants, with more critical questions about the risk and with less critical questions. If we consider only the group with phrasing focusing on the risk, half of the participants (49.25%) would take the risk presented. This difference from all participants' risk-taking is insignificant ($t = 1.42$, $p = 0.15$). Rephrasing the question focusing on the high chance that no harm will occur, led to 61.1% willing to take this risk (see Table 2). However, there is no added risk compared to the question that we framed otherwise. This difference is insignificant compared to all participants' risk-taking ($t = 1.31$, $p = 0.191$). As a result, the phrasing focusing on the opportunity or the risk affects the participants' willingness to take risks. If, for example, we as an employer would like our employees to show the lowest possible willingness to take risks, we should choose formulations focusing on risk since this formulation has shown that the participants would take a lower risk.

Table 2. Differences between answers that focus on risks or chances

	Take Risk	**Avoid Risk**
Focus on Risk	49,25%	50,75%
Focus on Chance	61,1%	38,9%
Total	55,3%	44,7%

The different periods in which the possible consequences can occur have no statistically relevant effect on the participants' willingness to take risks. We want to assess if employees recognize this issue and are unwilling to take risks. We can distinguish between four user types depending on whether they participated in an awareness measure and whether they are willing to take a risk. For this comparison, the ANOVA test is suitable: With an f-value of 5.94 and a p-value of < 0.001, we can assume that the tested groups show different behavior on average. Accordingly, we can prove with this survey that the participants' behavior in terms of risk-taking depends on the phrasing of the scenario and the participants' previous experiences.

4.2 Qualitative User Study Results

Here we summarize the study results we introduced in Section 3.2. It was conducted in July 2020 with 31 participants using in-person interviews.

Demographics 61% of the participants are male, and the majority (58%) are between 25 and 37 years old. Table 3 shows the participants' demographics and the industry in which the participants work. The majority of participants (38.71%) work in an IT-related domain.

Table 3. Participant demographics of the interview study.

		Participants	
Gender	Male	19	61.3%
	Female	12	38.7%
	non-binary	0	0%
Age	18–37	18	58.1%
	37–49	7	22.6%
	49–65	6	19.4%

Field of work	**Participants**	
IT	12	38.7%
Education	7	22.6%
Consulting	4	12.9%
Healthcare	4	12.9%
Trade	2	6.5%
Politics	1	3.2%
Construction	1	3.2%

Users' ability to detect malicious emails. Half of the participants (51.6%) rank themselves at least somewhat confident (4 or 5) in their ability to detect malicious emails. The mean is 3.3 for all participants with SD = 1.12.

Users' trust in technical measures. The minority of participants (29.0%) rank technical measures to protect them from malicious emails as mostly sufficient. However, none of the participants believed that technical measures completely protect them from malicious emails. The majority of participants (71.0%) feel neutral or are skeptical about whether they should trust technical measures with their protection. The mean is below 3 with 2.87 and SD = 0.92.

Unmarked dataset. Participants binary classified each email in the dataset based on whether they believed it was malicious. The average MCC is 0.650 (min: -0.327; max: 1; SD: 0.275), and five participants correctly identified every malicious email as malicious and every legitimate email as legitimate, scoring

MCC = 1. An MCC of -1 would indicate that the participant did not flag any malicious email as malicious and flagged all legitimate emails as malicious. An MCC of 0 would indicate that the results resemble a completely random guess. Only one participant scored a negative MCC. However, 30 participants scored MCC > 0 on the first dataset, meaning they took at least *educated guesses*. Five participants (16.1%) misjudged malicious emails as benevolent. Lastly, 26 participants (83.9%) misjudged at least one legitimate email as malicious.

Correctly Marked Dataset. The average MCC is 0.761 (min: 0.218; max: 1; SD: 0.231), and 12 participants (38.7%) correctly identified every malicious email as malicious and every legitimate email as legitimate, scoring MCC = 1. The mean MCC is the highest among the four datasets. No participant scored MCC ≤ 0. Only 2 participants (6%) misjudged a maliciously marked email as benevolent. And 19 participants (61%) misjudged a legitimate email as malicious although it was flagged as legitimate. The correctly classified mean is 86.46% (SD = 0.12).

Dataset with one false positive. The average MCC is 0.714 (min: 0,102; max: 1; SD: 0.244) and 9 participants (29.0%) correctly identified every malicious email as malicious and every legitimate email as legitimate scoring MCC = 1. No participant scored ≤ 0. Three participants (9.7%) misjudged a maliciously marked email as benevolent. 21 participants (67.7%) misjudged one of the legitimate emails as malicious, even though it was flagged as legitimate. We see correctness < 0.7 for the overall classification correctness in this specific dataset.

One email in our dataset was only identified correctly by 61.3% of the participants as legitimate, although it is tagged as legitimate. One approach to explaining why the email is classified as malicious by so many participants could be that the email is very generic. There is no introductory note, and it starts directly with linked images and bold typefaces. Email number 8 is the false positive item that was marked as malicious by the authors. The vast majority (96.8%) of participants detected the email as legitimate, although it was colored as a malicious one.

Dataset with one false negative. This dataset contains the worst possible error with an email indicating "legit" even though it is *malicious*. The average MCC is 0.744 (min: 0.357; max: 1; SD: 0.205), and eleven participants (35.5%) correctly identified every malicious email as malicious and every legitimate email as legitimate, scoring MCC = 1. No participant scored MCC ≤ 0. Four participants (12.9%) misjudged a maliciously marked email as benevolent. Eighteen participants (58.1%) misjudged one of the legitimate emails as malicious, although it was flagged as legitimate. Email number 5 is the false negative item the authors placed wrongly marked in the dataset. The vast majority (93.5%) of the participants detected that email as malicious even though it was considered legitimate. This might be due to the raised awareness of participants during the whole experiment. Dataset 4 contains the most misclassified email of the whole experiment. It is legitimate and marked as legitimate. However, only a bit more than half of the respondents (58.1%) classified the email correctly as legit. Another email in our dataset was identified by only 58.1% of the participants correctly as legitimate, although it is tagged as legitimate. The email uses a generic address in the

From field, but there is a *reply-to* field that might look suspicious. Further, the email is very generic, and there is no personalized introduction. These factors probably lead to the comparable low correctness rate of only 58.1%, which is significantly lower than the overall correctness rate of 85.4%. The average MCCs for all four datasets are summarized in Table 4. We list the rates of how many participants correctly identified the wrongly marked emails in the datasets with one false positive or negative, respectively.

Table 4. Average MCCs for the different datasets.

Dataset	**MCC**	**False marking recognized**
Unmarked	0.650	—
Correctly marked	0.761	—
One false positive	0.714	96.8%
One false negative	0.744	93.5%

Perception of color marking as helpful. Most users (71.0%) rank the color marking as at least *somewhat helpful* with a 4 or 5 on the Likert scale when asked if they believe that marking emails with a color in the inbox helps detect malicious emails. The mean is 3.74 with SD = 0.93 for the 31 participants. Only one respondent evaluates the color marking as "not helpful". *Perception of color marking to increase confidence in own detection capabilities.* Most participants (65%) rank the color marking at least *somewhat helpful* (4 or 5) when asked if they believe that it increases their confidence in their detection capabilities (mean: 3.52; SD: 1.06). Only 4 respondents (13%) believe that the color marking would be counterproductive for their detection capabilities.

5 Related Work

Farahmand analyzes the decision weights of underwriters and corporate managers for cybersecurity [12]. He concludes that they overweigh low probability cybersecurity events and underweigh high probability cybersecurity events. He furthermore shows that the value function changes if an organization experiences a breach. Fineberg states that current cyber strategies are still operated as if the actors in cyberspace are acting rationally [14]. He extends the work of Herley, who showed that users reject security advice rationally [17]. Lahcen et al. also point out that research in behavioral economics intersecting with cybersecurity needs to be done, especially with the emerging importance of humans as an integral part of cybersecurity strategies [25]. Bada et al. examined cybersecurity awareness campaigns and found that they mostly fail because the simple transfer of knowledge is not enough. Positive cybersecurity behaviors need to be enforced so that thinking becomes a habit and part of organizational culture [4]. Current research primarily focuses on the poor design of security systems and policies, but not on the behavioral aspect and individual decision-making [19, 31]. Jalali et al. find in a cybersecurity game experiment that decision-making is profoundly entrenched, and management decision alone is not helping in decision-making compared to inexperienced players [21]. Qu et al. investigate another approach by applying prospect theory to security decisions and show that those in a "disadvantage" situation are more likely to be persuaded to make better security decisions [35]. These findings are supported by the work of Amador et al. who investigated password selection processes and also found that intervention guided by prospect theory is causing 25% of users to improve their password strength [3]. To face this challenge from a research perspective, the field of *nudging* is more researched. Peer et al. show that personalized nudges increase the effect of nudging in choosing a solid password [32]. Zimmermann and Renaud show that the poor understanding of nudges in cybersecurity is currently hindering effective nudging. The *hybrid nudge*, consisting of a nudge and information provision, is a practical decision helper in some context [47]. Furthermore, warnings for malware and phishing exist in various contexts, such as browser toolbars, pop-up screens from firewalls, or browsers [11, 16]. Warnings try to prevent users from entering sensitive information or executing attachments. Warnings should be understandable, authoritative, primarily accurate, and not just passive warnings that can be easily clicked away or ignored [28, 41]. Qu et al. analyze that the interaction of framing and timing is important when nudging users towards an improvement in their cybersecurity decisions [36].

6 Discussion

Economic phenomena generally involve distributing scarce resources in combination with human behavior. In cybersecurity, this scarce resource is *attention*. Getting this attention is all about communication, and our survey reveals that even slightly different wording can lead to different outcomes. When communicating about cyber risks and building awareness, it is essential not to monger fear among users because this is likely to backfire [46]. This is another indication that it is not just about informing the user but also focusing on details of effectively influencing users to make decisions [37] that favor their cybersecurity. However, our research empirically shows that communicating risks directly hinders potential chances that arise from digitalization. Besides, the risk appetite of the participants is vast. Some participants took all the risks presented, but other participants rejected all risks. Through the differently formulated scenarios in which the willingness to take risks was analyzed, it could be determined that the formulation of the situation exerts an effect on the willingness to take risks. The risk in the scenarios where the wording focused on the risk arising was taken significantly less often than in the scenarios where the wording focused on the chance of no harm arising. This finding is analog to the results of Tversky et al. [42], where they describe the fact that risks have a more significant effect on decisions than focusing on gains. This behavior shows that framing and economic principles are essential factors when dealing with cybersecurity risks.

According to our results, using the correct wording without exuding fear can lead to an effective behavioral change. The majority of the participants stated being stressed in their private and professional Internet use. This feeling of stress affects their company's IT security because the stressed participants assess the IT security of their own company as secure. However, our survey did not reveal any statistically significant connection between the respondents' feelings of stress and their willingness to take risks, although this connection seemed natural. In our survey, stress affects clicking on fraudulent or harmful emails. These measured effects suggest that, in reality, there is an effect between stress perception and risk-taking, as risk-taking is often a trade-off between risk and benefit, so stressed people may be inclined to take the risk to avoid building up further stress. Our results from the experiment with the follow-up survey in Section 4.2 indicate that placing a nudge before opening a potentially malicious mail is helping users detect potentially malicious emails. To some extent, the participants were stressed because they had a time limit of five seconds for their decision, which is a realistic setting as discussed in Section 3.2. Section 4 shows that the average MCC and the total number of participants who score correctly when detecting malicious emails in a binary classification scheme are significantly improving for nudged emails. This study indicates that a color marking is a nudge. Just one participant answered perfectly and detected every malicious mail as malicious and every legitimate mail as legitimate and has MCC = 1. The average MCC for all participants is 0.697 (SD = 0.17), and no participant scores a negative MCC. The average MCC of the nudged dataset 2 is 17% better than the MCC in the non-nudged dataset 1. Even the average MCC over all datasets is 10% higher than in the first dataset 1, which also shows that the color marking helps users to make profound decisions if nudged. Another important factor is the increased confidence in the user's detection capabilities. Confident humans are more confident in detecting deceit and can discriminate between accurate and inaccurate lie detection [40, 45].

7 Limitations

We decided to provide the questionnaire only in English. Therefore, there might be a language bias because users did not receive the questionnaire in their native language. Confirming or rejecting this correlation is essential from the perspective of cybersecurity, as it can prevent risk-taking and, thus, possible damage in the long term. While we conducted the user study, we were given the feedback that the red and green marked mails are challenging to differentiate for people with red-green colorblindness. This needs to be addressed in the future. Furthermore, we put users under one specific stress (time), which is just one of many ways to exert stress on humans [22]. This study does not consider participants' characteristics (i.e., personality traits) or contextual factors (e.g., daytime/nighttime). A more diverse participant pool should be studied if other research groups try to improve our approach.

7.1 Ethical Consideration

For this study, we gathered responses from various participants. Our research institution does not require approval for this type of study, nor does it provide an Institutional Review Board (IRB). Nevertheless, we took strict ethical considerations into account. We never collected any PII like *name* or other information that would make the identification of a single respondent possible.

8 Conclusion

It is apparent from our survey that the majority of participants are optimistic about the awareness measures currently used in the industry. This positive attitude relates to both the perception of instructiveness and the entertaining nature of the various measures. We studied behavior change when nudging users towards an absolute security-relevant decision and hypothetical risk-taking. In our user study, we see that nudging provides an effective way to influence users' decisions and, thus, cope with the situation better. Even when users experience stress, we show empirically that nudging has a positive effect on the decision-making of our participants. On the other hand, our survey shows that the threat appraisal is different, depending on the situation's framing. To our knowledge, this is the first work to focus on empirically showing this behavioral economic effect in cybersecurity. Focusing on communicating the risks of data losses is an effective way to raise awareness, yet it is crucial not to foment fear, but to generate changes in behavior.

Acknowledgments. The authors gratefully acknowledge funding from the *Federal Ministry of Education and Research* (16KIS1648 "DigiFit", 16KIS1628K & 16KIS1629 "UbiTrans").

References

1. Agence nationale de la sécurité des systèmes d'information: Signalement d'un contenu inadéquat (2023), `https://cyber.gouv.fr/signalement-dun-contenu-inadequat`, accessed: 2023-11-24
2. Akhawe, D., Felt, A.P.: Alice in warningland: A large-scale field study of browser security warning effectiveness. In: 22nd {USENIX} Security Symposium ({USENIX} Security 13) (2013)
3. Amador, J., Ma, Y., Hasama, S., Lumba, E., Lee, G., Birrell, E.: Prospects for improving password selection. In: Nineteenth Symposium on Usable Privacy and Security (SOUPS 2023). pp. 263–282 (2023)
4. Bada, M., Sasse, A.M., Nurse, J.R.C.: Cyber security awareness campaigns: Why do they fail to change behaviour? (2019)
5. Butler, S.A.: Security attribute evaluation method: a cost-benefit approach. In: Proceedings of the 24th international conference on Software engineering (2002)
6. Cambridge University Press: Cambridge Dictionary, `https://dictionary.cambridge.org/dictionary/english/awareness`, accessed: 2021-03-21

7. Chai, S., Kim, M., Rao, H.R.: Firms' information security investment decisions: Stock market evidence of investors' behavior. Decision Support Systems (2011)
8. Chicco, D., Jurman, G.: The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. BMC genomics (2020)
9. Coale, A.J.: Excess female mortality and the balance of the sexes in the population: an estimate of the number of" missing females". The Population and Development Review pp. 517–523 (1991)
10. Commission Nationale de l'Informatique et des Libertés: Phishing : détecter un message malveillant (2017), `https://www.cnil.fr/fr/phishing-detecter-un-message-malveillant`, accessed: 2023-11-24
11. Egelman, S., Cranor, L.F., Hong, J.: You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2008)
12. Farahmand, F.: Applying behavior economics to improve cyber security behaviors. Tech. rep., Georgia Institute of Technology Atlanta United States (2018)
13. Fielder, A., Panaousis, E., Malacaria, P., Hankin, C., Smeraldi, F.: Decision support approaches for cyber security investment. Decision support systems (2016)
14. Fineberg, V.: Bec: Applying behavioral economics to harden cyberspace. Journal of Cybersecurity and Information Systems (2014)
15. German Federal Office for Information Security: Drei Sekunden für mehr E-Mail-Sicherheit (2020), `https://www.bsi-fuer-buerger.de/BSIFB/DE/Empfehlungen/Menschenverstand/E-Mail/E-Mail_node.html`, accessed: 2020-09-02
16. Gupta, B.B., Tewari, A., Jain, A.K., Agrawal, D.P.: Fighting against phishing attacks: state of the art and future challenges. Neural Computing and Applications (2017)
17. Herley, C.: 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. pp. 133–144 (2009)
18. Ho, T.H., Lim, N., Camerer, C.F.: Modeling the psychology of consumer and firm behavior with behavioral economics. Journal of marketing Research (2006)
19. Iuga, C., Nurse, J.R., Erola, A.: Baiting the hook: factors impacting susceptibility to phishing attacks. Human-centric Computing and Information Sciences (2016)
20. Jalali, M.S., Bruckes, M., Westmattelmann, D., Schewe, G.: Why employees (still) click on phishing links: investigation in hospitals. Journal of medical Internet research (2020)
21. Jalali, M.S., Siegel, M., Madnick, S.: Decision-making and biases in cybersecurity capability development: Evidence from a simulation game experiment. The Journal of Strategic Information Systems (2019)
22. Keay, K.A., Bandler, R.: Parallel circuits mediating distinct emotional coping reactions to different types of stress. Neuroscience & Biobehavioral Reviews (2001)
23. Li, X., Lee, C.J., Shokouhi, M., Dumais, S.: Characterizing reading time on enterprise emails. arXiv preprint arXiv:2001.00802 (2020)
24. Likert, R.: A technique for the measurement of attitudes. Archives of psychology (1932)
25. Maalem Lahcen, R.A., Caulkins, B., Mohapatra, R., Kumar, M.: Review and insight on the behavioral aspects of cybersecurity. Cybersecurity (2020)
26. Maddux, J.E., Rogers, R.W.: Protection motivation and self-efficacy: A revised theory of fear appeals and attitude change. Journal of experimental social psychology (1983)

27. March, J.G., Shapira, Z.: Managerial perspectives on risk and risk taking. Management science (1987)
28. Modic, D., Anderson, R.: Reading this may harm your computer: The psychology of malware warnings. Computers in Human Behavior (2014)
29. Mullainathan, S., Thaler, R.H.: Behavioral economics. Tech. rep., National Bureau of Economic Research (2000)
30. National Cyber Security Centre: Quick Guide: Phishing (2022), https://www.ncsc.gov.ie/pdfs/NCSC_Quick_Guide_Phishing.pdf, accessed: 2023-11-24
31. Nurse, J.R., Creese, S., Goldsmith, M., Lamberts, K.: Guidelines for usable cybersecurity: Past and present. In: 2011 third international workshop on cyberspace safety and security (CSS). IEEE (2011)
32. Peer, E., Egelman, S., Harbach, M., Malkin, N., Mathur, A., Frik, A.: Nudge me right: Personalizing online security nudges to people's decision-making styles. Computers in Human Behavior (2020)
33. Pfleeger, S.L., Caputo, D.D.: Leveraging behavioral science to mitigate cyber security risk. Computers & security (2012)
34. Powers, D.M.: Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. Journal of Machine Learning Technologies (2011)
35. Qu, L., Wang, C., Xiao, R., Hou, J., Shi, W., Liang, B.: Towards better security decisions: applying prospect theory to cybersecurity. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. pp. 1–6 (2019)
36. Qu, L., Xiao, R., Shi, W.: Interactions of framing and timing in nudging online game security. Computers & Security **124**, 102962 (2023)
37. Rogers, R.W.: Attitude change and information integration in fear appeals. Psychological reports (1985)
38. Sasse, M.A., Brostoff, S., Weirich, D.: Transforming the 'weakest link'—a human/computer interaction approach to usable and effective security. BT technology journal (2001)
39. Schneier, B.: The psychology of security. In: International conference on cryptology in Africa. Springer (2008)
40. Smith, A.M., Leach, A.M.: Confidence can be used to discriminate between accurate and inaccurate lie decisions. Perspectives on Psychological Science (2019)
41. Stembert, N., Padmos, A., Bargh, M.S., Choenni, S., Jansen, F.: A study of preventing email (spear) phishing by enabling human intelligence. In: 2015 European Intelligence and Security Informatics Conference. IEEE (2015)
42. Tversky, A., Kahneman, D.: Loss aversion in riskless choice: A reference-dependent model. Quarterly Journal of Economics **106**, 1039–1061 (1991)
43. Van Der Heijden, A., Allodi, L.: Cognitive triaging of phishing attacks. In: 28th {USENIX} Security Symposium ({USENIX} Security 19) (2019)
44. Vinkers, C.H., van Amelsvoort, T., Bisson, J.I., Branchi, I., Cryan, J.F., Domschke, K., Howes, O.D., Manchia, M., Pinto, L., de Quervain, D., et al.: Stress resilience during the coronavirus pandemic. European Neuropsychopharmacology (2020)
45. Vrij, A., Baxter, M.: Accuracy and confidence in detecting truths andlies in elaborations and denials: Truth bias, lie bias and individual differences. Expert evidence (1999)
46. Witte, K.: Fear control and danger control: A test of the extended parallel process model (eppm). Communications Monographs (1994)
47. Zimmermann, V., Renaud, K.: The nudge puzzle: matching nudge interventions to cybersecurity decisions. ACM Transactions on Computer-Human Interaction (TOCHI) (2021)