'Protect and Fight Back': A Case Study on User Motivations to Report Phishing Emails

Pavlo Burda
Eindhoven University of Technology
Eindhoven, The Netherlands
p.burda@mailbox.org

Luca Allodi
Eindhoven University of Technology
Eindhoven, The Netherlands
l.allodi@tue.nl

Alexander Serebrenik
Eindhoven University of Technology
Eindhoven, The Netherlands
a.serebrenik@tue.nl

Nicola Zannone
Eindhoven University of Technology
Eindhoven, The Netherlands
n.zannone@tue.nl

ABSTRACT

Phishing reporting is emerging as a key defense mechanism against phishing attacks. Whereas large enough organizations have specific policies in place for phishing reporting, user uptake is still limited, and a clear picture of what motivates users to report and which types of emails is still to be drawn. Yet, this is critical to devising better policies and procedures and stimulating awareness and a cyber-security culture within organizations. In this work, we sample and interview $n = 49$ employees from the pool of phishing reporters at a medium-sized European technical university. We sample interviewees based on how sophisticated the emails they report are over contextual and technical dimensions and cluster reporters in terms of their (emerging) reporting behavior. We conduct semi-structured interviews up to thematic saturation and derive 13 main themes driving reporting motivations. We discuss the identified themes in the broader theoretical context, as well as the practical implications of our findings.

CCS CONCEPTS

• **Security and privacy → Social engineering attacks**; • **Human-centered computing → Empirical studies in HCI**; • **Social and professional topics → Phishing**.

KEYWORDS

Phishing, Reporting

ACM Reference Format:
Pavlo Burda, Luca Allodi, Alexander Serebrenik, and Nicola Zannone. 2024. 'Protect and Fight Back': A Case Study on User Motivations to Report Phishing Emails. In *The 2024 European Symposium on Usable Security (EuroUSEC 2024), September 30-October 1, 2024, Karlstad, Sweden.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3688459.3688473

EuroUSEC 2024, September 30-October 1, 2024, Karlstad, Sweden

ACM ISBN 979-8-4007-1796-3/24/09
https://doi.org/10.1145/3688459.3688473

1 INTRODUCTION

Phishing attacks are a major threat to organizations and private citizens alike. Automated phishing detection and filtering are measures commonly in place in most organizations, yet phishing emails regularly pass those filters and end up in users' inboxes. At that point, the user is the last line of defense against an attack that, if successful, may pose risks for the whole organization. Cyber security awareness campaigns, training, and phishing simulations are generally aimed at improving users' ability to *detect* phishing attacks [2]; on the other hand, a single successful attack may give way to an attacker to successfully breach through, for example through subsequent lateral movement attacks [23]. It is, therefore, key that the organization is in the position to take swift action upon the arrival of a new attack. *Reporting* is the main mechanism organizations rely on and one that received the attention of several recent research contributions [8, 16, 26, 28–30, 32]. User reporting is a mechanism that allows users to report suspicious emails to central analysis units (sometimes hosted within the organization, if large enough, or outsourced to service providers) that can then take action on that information. This may include updating automated detection filters, blocking associated domains, and/or checking (e.g., through the employment of a monitoring infrastructure such as a Security Operation Center) whether reported emails led to other users clicking on suspicious links or opening malicious attachments. On the other hand, phishing reporting is currently a relatively under-studied topic. Crucially, users are the main driver behind a successful defense mechanism, at least in part relying on reporting [8, 12].

Reported important factors driving users' decision to report phishing emails include the technical confidence they have in using the mechanism (*self-efficacy*) [29, 32], the characteristics of the attacks [8, 16], and knowledge of organizational policies on phishing reporting [32]. Interestingly, the role of the organization appears to emerge frequently in the literature as associated with a user's likelihood to report. For example, keeping the reporter informed on the outcomes of their reporting seems to motivate users to report [26, 30, 57]. Similarly, previous research showed the role of personality traits associated with so-called *positive cyber security behaviors*, within the broader context of *Organizational Citizenship Behaviors* (OCB), in affecting a user's propensity to report phishing [32]. Indeed, reporting assumes the traits of discretionary

behavior that individuals engage with to the benefit of the organization as a whole without specific obligations or rewards to do so. In this context, it becomes critical to understand what *motivates* individuals in reporting phishing emails in an organizational setting. Doing so will allow us to devise better policies, instruments, and processes to identify and act on reported phishing and create safer IT environments for users to operate in (including those not inclined to reporting) [9]. Importantly, the link between reporting behavior and motivations to report has not yet been fully explored, leaving a large gap in the characterization of security behavior and open questions on how to best nudge or incentivize users towards positive cybersecurity behaviours [9, 32]. To address this gap, in this work, we answer the following research question:

What are the factors driving employee decisions to report suspicious phishing emails?

To answer this question, we collaborated with the Operations Security Team of our institution, a medium-large technical university in Europe, to analyze 8369 emails reported by employees over 766 days. We first analyzed the reported emails to identify the overall emerging reporting behavior of individual reporters in terms of reporting frequency and type of reported emails. For the latter, we devised heuristics to evaluate at scale to what degree a reported email is contextually and technically sophisticated with respect to the organization's environment. For example, an email spoofing our institution's domain would be considered technically sophisticated. Similarly, one mimicking internal communication styles would be considered contextually sophisticated. By evaluating the characteristics of reported emails over these dimensions, we clustered reporters based on their *emergent behavior* in reporting (e.g., users that tend to report highly sophisticated emails on both dimensions are more likely to be clustered together than not). We then employed these clusters as 'strata' to sample employees and interviewed them to gauge what motivates them in reporting. We iteratively coded interview transcripts to identify key emerging themes and continued sampling from said clusters until we reached 'thematic saturation', i.e., no new themes emerged from the latest two interviews in that cluster. We reached thematic saturation after $n = 49$ interviews with as many employees.

Our contribution is multi-faced. We identify a number of themes that motivate employees to report. The interplay between these themes is complex and ranges from the desire to protect the organization and help less security-conscious colleagues to the desire to fight back and neutralize attackers. These can be used to shape organizational policies on reporting, as well as awareness programs focusing on outcomes that align well with users' motivations. Awareness, doubt, and (technical) self-efficacy play an important role in determining whether an employee will report a suspicious email. Interestingly, 'doubt' can be a motivating factor pushing employees to report 'just in case' the email might be malicious. We discuss our findings in the broader theoretical context of Protection Motivation Theory (PMT), identifying several parallels between identified themes and PMT, thus suggesting that it may represent a meaningful framework for evaluating phishing reporting mechanisms.

Outline. The paper is structured as follows. Section 2 discusses relevant background. Section 3 presents our methodology and data. Section 4 and Section 5 respectively present and discuss results. Section 6 concludes the paper.

2 BACKGROUND AND RELATED WORK

2.1 Phishing reporting

Phishing reporting is part of cyber-security response strategies at organizations whereby employees who detect phishing attempts may notify the relevant (IT) department of an ongoing campaign. The remediation procedure can intercept the attack, for example, by blocking traffic to rogue domains or alerting users who have not yet fallen for the attack. An efficient phishing reporting process is thus fundamental to enable a timely response as a significant portion of targeted employees is victimized in the first few hours from attack delivery [8]. In response to the growing threat of more sophisticated phishing attacks, such as spear phishing, research has increasingly focused on phishing reporting and factors influencing it [7, 8, 16, 26, 28–30, 32, 49]. Previous work explored how organizations can improve reporting, for example, by examining incentives to report [26], by identifying 'naturally immune' individuals [8], and by testing reporting effectiveness in the field [30]. Among factors influencing reporting, much of the research explored individual factors, such as perceptions, beliefs, and attitudes towards the organization [29, 32], and contextual factors, such as user interface, job role, and situation [16, 30, 49]. However, what reasons and motivations drive individuals to report phishing attacks in organizations remains unclear. For instance, investigators in [8] and [16] interviewed respectively 12 and 14 reporters of a spear phishing campaign. The findings reveal that employees may be unable to generalize the rationale for reporting a suspicious email, stating various reasons for reporting, such as being aware of the sophistication of an attack or feeling responsibility, and *not* reporting due to ill-perceived liability or lacking efficacy towards phishing. In another study with nine participants [7], the authors suggest that, on top of feeling responsible for colleagues, some employees may decide to only report phishing emails that are more sophisticated or more *believable* than 'generic' phishing. On the other hand, evidence suggests that more believable emails are less likely to be reported as more believable emails are, in principle, less detectable [28]. Overall, phishing reporting depends on a variety of motivations, attitudes, and the type of phishing emails encountered, and as such, it can be considered an emergent behavior arising from the interactions of these factors.

2.2 Phishing believability

We draw from previous research on reporting and phishing sophistication the concept of 'phishing believability', i.e. the extent the receiver considers a (phishing) email as a credible message from the source [28]. The idea behind phishing believability is that the higher the believability, the lower the chances that a phishing email will be detected as phishing and, eventually, reported to an IT department. Several factors have been shown to influence the believability of a phishing email, such as technical specifications of the email (e.g. sender, payload URL [15, 25, 36, 40]), contextual alignment between the email and target (e.g., impersonation and

Table 1: Mail believability features.

	Feature	Description
Technical believability	*Payload URL*	The link requiring target interaction should be plausible or difficult to distinguish from a genuine one [36, 50].
	Sender	Non-trivial, coherent spoofing or masquerading of Sender name and address (username and/or domain) [39, 50].
	Attachment	The type of document attached to the email does not appear to be malicious or obviously harmful [50]. The name of the attachment indicates a harmless file [14].
	Other headers	Other email headers, such as To: and Reply-To:, can be spoofed as part of a phishing tactic [45].
Contextual believability	*Impersonation*	The manipulation of email features in such a way the email seems to originate from a legitimate source. For example, mentioning or purporting to be the target organization in the sender, subject, or body of the email [15, 25, 40].
	Premise alignment	The email content aligns with target experiences and expectations, such as references to events or activities happening in the target's environment, or aligning the pretext with typical internal procedures or jargon at an organization [41, 50].
	Timing	The delivery of the attack at a useful time, such as festivities, specific events or a busy time of the day [4].

pretext alignment [10, 50]), appropriate language and tone (e.g., communication style at organizations [10, 55]), the visual aspects resembling a legitimate email (e.g., graphics, layout [58, 62]), and the usage of persuasion techniques (e.g., authority or scarcity persuasion principles [53, 54]).

Although each of these factors contributes to the sophistication and believability of phishing emails, we focus on those factors that, at the same time, can increase the believability of a phishing email and are commonly found in phishing emails. Specifically, contextualized emails are often used in spear phishing campaigns, which are generally considered more believable [6, 16]. On the other hand, many phishing attacks attempt to forge technical specifications, such as URLs or attachments, of an email to make it difficult to distinguish it from a legitimate one [13, 22]. We, therefore, focus on the factors affecting 'technical believability' and 'contextual believability'. Technical believability refers to how convincing or realistic the email appears from a technical standpoint [28, 50]. This includes the spoofing of the sender address or other header details, usage of homograph techniques to mask the sender domain or payload URL, hiding the URL payload behind legitimate services, such as well-known URL shorteners or file hosting services, or disguising email attachments. Contextual believability concerns the alignment with the context or expectations of the recipient [20, 28]. This includes tailoring the contents of the email to the targets, such as impersonating the target's organization or providing details that seem relevant to the recipient's recent activities and interactions, such as with internal departments or relevant external entities. Table 1 describes the email features often used to improve technical and contextual believability.

2.3 Research gap and contribution

It has been shown that phishing reporting at organizations depends on a variety of factors, such as attitude, job role, or the type of phishing encountered [29, 32, 49], as well as different employee motivations [7, 8, 16]. However, only a few works investigate employees' reasons for reporting, often with a handful of participants and following simulated phishing campaigns, suggesting that a full assessment and understanding of how employees rationalize reporting is still lacking.

To overcome this gap in the understanding of user reporting motivations, research is needed to identify emergent behaviors in reporting, and identify the factors that condition such behavior (e.g., what do they report). For instance, individuals' decisions to report can be influenced by the perceived sophistication and consequent danger of the encountered phishing email [7, 8]. Previous research suggests that observing reports of more sophisticated emails is less likely than reports of 'generic' phishing [28]. To gain such an understanding, in this work, we explore reasons and motivations to report phishing emails by interviewing 49 employees of a mid-sized European university who reported phishing between 2019 and 2021.

3 METHODOLOGY

Since our research question is exploratory in nature, suitable research methods include those that offer rich, qualitative data about a phenomenon, allowing building hypotheses and tentative theories [17]. To this end, we adopted semi-structured interviews [5] to collect data. The advantage of interviews over, e.g., surveys, is the possibility to ask follow-up questions and clarify respondents' answers.

To sample participants, we analyzed a dataset comprising emails reported to the IT department of a mid-sized European university (from now on, *UNI*) by its employees (Section 3.1). The reporting mechanism at UNI at the time of the study is to forward suspicious emails to a dedicated shared inbox, called '*abuse inbox*', that IT security employees monitor to detect and take action (e.g., domain/IP blocking) on reported threats. Reporters receive a feedback note from the IT security team once the investigation is completed reporting on the outcome of the investigation and any actions taken, where relevant. For our study, we gained authorized access to the *abuse inbox* of UNI.

The analysis of the *abuse inbox* allows us to identify employees with similar emergent behavior in terms of reporting emails with common characteristics, i.e., similar contextual and technical believability. From these clusters, we performed stratified sampling [3] to interview employees exhibiting the identified different emerging behaviors (Section 3.2). We asked questions about their motivations for reporting suspicious emails (Section 3.3) and applied the initial stage of the Socio-Technical Grounded Theory (STGT) method [24] to synthesize large amounts of rich, qualitative data (Section 3.4). In STGT, data collection and analysis occur by means of theoretical sampling and thematic analysis. An overview of the overall approach is presented in Fig. 1.

3.1 Data characterization

To understand the distribution of reports and reporters, we first characterized the dataset of the emails reported to the IT department of UNI (hereafter, *abuse inbox*). This step aims to determine employees with similar emergent behaviors (cf. Section 3.2). The

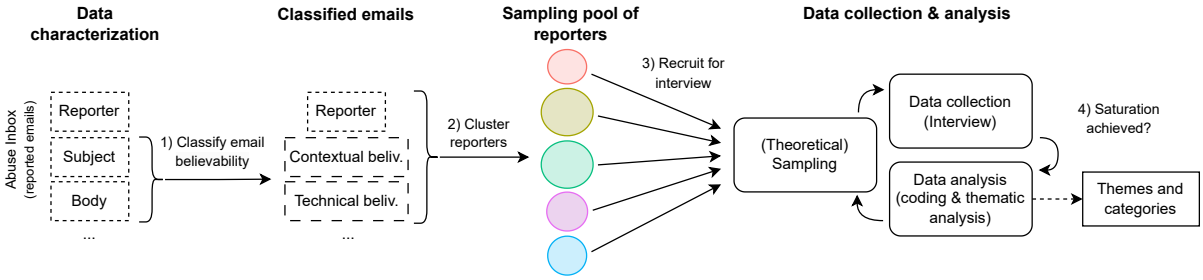


Figure 1: Methodology overview.

Table 2: Abuse inbox variable descriptions, unique counts, and examples.

Variable	Description	Count	Example
id	Unique id for each reporting instance	8369	012c8a61
reporterAddress	Aliased email address of a reporter	1460	UZYRP65M
toAddress	Recipient address from the reported email	1921	NAME@fewell.cf
fromAddress	Sender address from the reported email	3178	NAME@fewell.cf
subject	Subject of the reported email (if available)	3119	Re: Request for Quotation
body	Body of the reported email	6503	Good day, Please find attached ...
attachmentName	Attachments' names (if any)	687	Quotation.iso
attachment	SHA256 hash of attachments (if any)	1118	dbd06719dcea540153d76d7ac9 ...
receivedTime	Timestamp of received time of the reported email (if available)	5102	Tuesday, 19 November 2019 18:50
reportedTime	Timestamp of reported time of the reported email	8356	20-Nov-19 8:57:52 AM

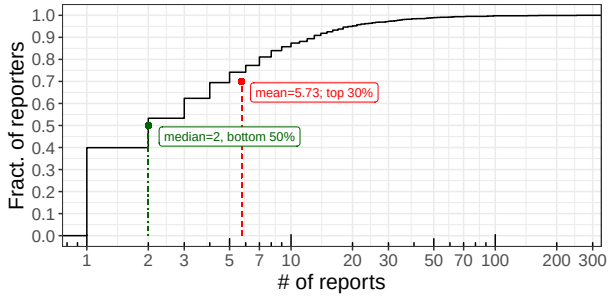


Figure 2: Fraction of reporters and # of reports (log10 scaled).

dataset extracted from the abuse inbox at UNI contains 8369 individual emails spanning from 2019-02-08 to 2021-03-15 for a total of 766 days. Due to the location of UNI, the dataset features emails in both Dutch and English.[1]

A summary of the dataset is given in Table 2. All variables but `id` contain duplicate values, e.g., the same emails from a phishing campaign can be reported by multiple employees. The variable `reporterAddress` represents the 1460 unique reporters. The distribution of the number of reports per reporter is skewed towards a small number of reports per reporter (Mdn = 2, mean = 5.73), with the top 25% of employees reporting at least 6 emails (Q3 = 6). Fig. 2 presents the fraction of reporters per number of reports.

3.2 Sampling and recruitment

An adequate sample in theoretical sampling is achieved when data analysis does not generate new insights after several iterations of sampling and analysis (thematic saturation) [24]. Following established guidelines [19, 21], we iteratively extended a starting sample of participants based on the outcomes of the thematic analysis until thematic saturation was achieved, i.e., no new insights were obtained. However, by sampling participants directly from the reporters' population, e.g., with random sampling, we could achieve thematic saturation before interviewing individuals whose emergent behavior is less prevalent [24]. Specifically, it could exclude from the interviews employees who reported rare, sophisticated phishing emails. Thus, we opted for stratified sampling: we first classified emails over features related to contextual and technical believability (Section 3.2.1); then clustered the employees that reported a similar fraction of email classes, including potentially highly believable emails (Section 3.2.2); and finally recruited the participants to interview from each cluster (Section 3.2.3).

3.2.1 Email classification. We first randomly sampled 20 reporters and manually reviewed 131 emails reported by them to determine which believability features (cf. Section 2.2) can be used to automatically classify emails. To ensure the quality of the review [33], four investigators applied the criteria defined in Table 1, discussed each sampled email, identified and resolved points of disagreements, and iteratively updated the review of emails. Given the limited amount of emails to be coded, two co-coding sessions were enough to resolve disagreements and update the criteria definitions.

After excluding reports of non-phishing emails[2], the review suggests that the sampled emails vary largely in content and mostly present a low believability on either the contextual dimension, technical dimension, or both. Emails deemed highly believable on both dimensions are relatively rare (≈14%, see Appendix A). Importantly, we observe a high variability of features affecting contextual and

[1]One of the authors is a native Dutch speaker. The automated processing described in Section 3.2.1 accounts for the language of the reported email.

[2]Not phishing: 42 out of the 131 considered reported emails (32%), of which 28 are spam emails and 14 are legitimate emails.

Table 3: Email classification criteria. The full list of domains and file types is reported in Table 8 (Appendix A).

Criterion	Type	Class	Example
Mention of UNI and its variants in body or subject	contextual	high	Dear UNI employee
Otherwise	contextual	low	
If any, the payload URL contains:			
a domain of popular URL shorteners	technical	high	bit.ly, t.co
a domain of popular file hosting services	technical	high	dropbox.com
a homograph attack on UNI	technical	high	Vniversity
a domain of popular free web hosting services*	technical	low	vnivesity.weebly.com
If any, the attachment file type is:			
uncommon file type in UNI office setting	technical	high	iso, html, msi, etc.
Otherwise	technical	low	

* Overrides previous URL criteria.

technical believability (i.e., `fromAddress`, `subject`, and `body`). This makes the implementation of automated solutions for accurately labeling reported emails hard or impossible, as it would require manually labeling thousands of email features by trained agents with contextual knowledge of UNI's environment. However, to build our employee sampling pool, it suffices to have an *approximate* representation of email believability. We assigned each email a binary value, *high* or *low*, indicating the presence or absence of contextual believability and technical believability features. Table 3 shows the feature matching rules used for the classification stemming from Table 1. The classifier achieves an accuracy of 88.6% and a precision of 76.1% for contextual believability, and an accuracy of 72.5% and a precision of 78.1% for technical believability (see Appendix A for further details).

We find that emails with low technical and contextual believability are the most common in our dataset (4805, 57%). Emails with high technical believability are the least common, equal split between high (745, 9%) and low contextual believability (750, 9%). High contextual believability and low technical believability are relatively common (2069, 25%).

3.2.2 Clustering of reporters. The sheer majority of reporters in our dataset reported suspicious emails only once or twice (Mdn = 2). As our goal is to understand users' motivation to decide to report or not report a suspicious email, we are interested in identifying subjects that do show repeated reporting behavior. The average number of reports per person is 5.7, corresponding to the top 30% of the reporters. Hence, we chose five reports as a threshold for prospective interviewees. This results in a pool of 445 subjects. As we are interested in individuals' behavior and decision-making, we filtered out reporters whose email address corresponds to a shared functional account, such as 'library@UNI.edu', because it is unfeasible to trace back individuals who reported a specific message in the past from a given shared address. Among 445 subjects, 336 used a personal, as opposed to functional, email address.

Each reporter can be characterized in terms of the classes of emails they have reported, i.e., each reporter can be described using four variables: `Ch_Th` (fraction of high contextual and high technical believability emails among all emails they have reported), `Cl_Th` (contextual low and technical high), `Ch_Tl` (contextual high and technical low) and `Cl_Tl` (contextual low and technical low). Using these four variables as a representation of reports, we perform clustering. We do not expect clearly separated groups as it is likely that almost all reporters have reported the most prevalent type of phishing at some point (i.e., contextual low and technical low). To determine a suitable number of clusters, we applied the elbow method and chose five clusters (Fig. 3). Since the features (i.e., `Ch_Th`, `Cl_Th`, `Ch_Tl` and `Cl_Tl`) are numeric and at the same scale (i.e., bounded between 0 and 1), to perform the clustering we applied the Hartigan and Wong k-means algorithm [37] with 25 random sets for the initialization and the Euclidean distance.[3]

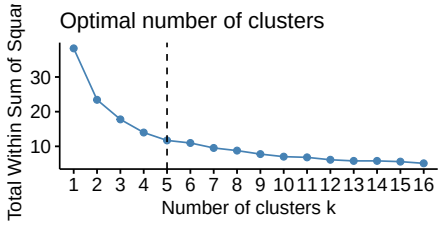


Figure 3: Optimal number of clusters following the elbow method. The plot shows the within the sum of square cost function [51] at the varying of the number of clusters k, for an arbitrary max of 16 clusters. The reduction of the within sum of squares was negligible for $k > 5$. Hence, we choose five clusters.

Fig. 4 shows a visualization (reduced to two dimensions) of the five clusters, capturing approximately 70% of the overall variance. Cluster 1 overlaps with Clusters 2, 3, and 5. Other clusters appear disjoint over these two dimensions; given the high fraction of captured variance, this suggests the clustering succeeded in meaningfully separating users based on emergent reporting behavior. Clusters 1, 3, and 4 have approximately 45 subjects each, whereas clusters 2 and 5 have 99 subjects (cf. Table 4). Clusters 2 and 5 include users reporting mostly (i.e., approx 80% of the time) 'technical low' emails (`Ch_Tl` or `Cl_Tl`). The other three clusters show a prevalence of 'high believability' emails over either technical, contextual, or both dimensions.

3.2.3 Recruitment of participants. The recruitment was performed in several iterations following the theoretical sampling approach. We send invitations to the institutional email address the prospective interviewees used for reporting to the abuse inbox. From the 336 reporters in the five clusters, we randomly sample subjects from each cluster. To keep the scheduling of interviews manageable, for each iteration, we sample no more than 10% of the cluster size.

[3]The algorithm is implemented in the R package `stats`.

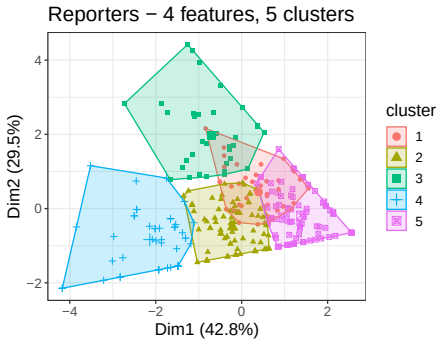


Figure 4: Reporter clusters projected on the first two dimensions of principal component analysis of the email features explaining 69.4% of variance. See Table 4 for the total number of reporters in each cluster and the mean values of `Ch_Th`, `Cl_Th`, `Ch_Tl` and `Cl_Tl` per cluster.

Table 4: Clusters of reporters. C=contextual and T=technical, l=low and h=high.

Cluster	Ch_Th	Cl_Th	Ch_Tl	Cl_Tl	Total	Invited	Interviewed
1	0.054	0.270	0.161	0.515	46	20	10
2	0.075	0.042	0.336	0.547	99	16	12
3	0.385	0.074	0.162	0.380	44	23	8
4	0.068	0.075	0.613	0.244	48	27	10
5	0.059	0.061	0.103	0.777	99	27	9
Total					336	90	49

Following the theoretical sampling procedure [24], we invited participants until thematic saturation was reached *for each cluster separately*. Based on Guest et al. [21], we deem saturation to be achieved when no new themes emerge from the analysis of the last two interviews. We reached thematic saturation after four rounds, and 49 interviews (cf. Table 4).

3.3 Interviews

During the interviews, we asked high-level questions on the rationale and motivations for reporting suspicious emails to the abuse inbox. Following best interview practices [5], a one-page interview guide was tested and revised after a pilot interview with a PhD student in our research group. A detailed interview guide is available in Appendix C. As recommended by Bird [5], the semi-structured interviews were carried out by one of the investigators in a confidential setting, either in-person (individual offices, or reserved conference rooms) or online via the institutional video conferencing service. Before the interview, participants were reminded that they could withdraw from the study without explanation and were handed an information sheet together with the consent form. English was the primary language, except for one case where the national language was used (a native speaker assured the validity of interview questions and answers).

3.4 Data analysis

The interviews were recorded, transcribed with the help of specialized software, and paraphrased following the approach taken in a similar study [16]. The subsequent analysis of the transcripts is largely based on the initial stage of the STGT method interleaving data collection and data analysis steps: 1) we identified emerging codes, 2) formed and assessed themes, 3) checked the saturation of themes, and 4) eventually sampled additional participants from our sampling pool (see Section 3.2). The first iteration of coding was carried out in person with hard-copy printouts by four investigators (one of them with extensive experience with qualitative methods). The following iterations were carried out by the first author in a virtual environment. To ensure the quality of the analysis, all four investigators regularly met and discussed the coded transcripts, identified ambiguities, and resolved disagreements. Furthermore, to ensure validity of the codes, the remaining authors have independently applied the codes created by the first author to three randomly selected paraphrased transcripts each. The inter-rater agreement [34] of the first author with the second one was 80%, with the third one was 68%, and with the fourth one was 77%, for an average agreement of 75% (a 'moderate' agreement level [34]). More details on the codes are available in Appendix D.

3.5 Ethical considerations

Data collection and analysis were carried out under ethical approval `ERB2020MCS13` by our institution's ethical review board. Potentially sensitive information willingly or unwillingly contained in the reported messages was removed from the analysis whenever possible. Participants were thoroughly informed about the research aims and methods, both orally and in writing, on a consent form; they were offered enough time to familiarize themselves with the form and ask further questions (including after the interview). The interview was carried out adhering to ethical guidelines [1]. Data collection and analysis were executed on the university's premises, through encrypted communication and storage, minimizing potential harm to the participants. Based on the consent agreement signed by the participants, we cannot publish the raw data.

4 RESULTS

4.1 Thematic analysis

We identified 13 themes characterizing the rationale for reporting phishing emails to the IT department by the employees. Table 5 summarizes the identified themes and participant contributions per cluster. We stress that the counts of participants merely reflect the opinions of the interviewees participating in our study and are not intended to reflect the prevalence of the themes among all reporters. As such, these numbers are intended to provide qualitative rather than quantitative insights.

The table shows that the intention to protect their colleagues, the organization, and themselves was mentioned most commonly by the participants. The second most common reason is helping their colleagues or the organization as a whole. Following, various, less homogeneous motivations drive employees' reporting behavior. For example, some employees report in case of doubt (`ask for confirmation`), others are aware of phishing risks due to their previous experience (`awareness and experience`). Less common reasons include `annoyance` and the will to '`fight hackers`'. Interestingly, participants mention `detection`, a distinct but related activity to reporting, as a motivation. In the following, we elaborate

in detail on the identified themes. Appendix D provides further examples in the codebook.

Protect and help. Theme `protect others` refers to the intention or desire to prevent others from receiving phishing emails, falling for phishing, or becoming subject to the negative consequences of phishing, such as causing *"harm to our computers"* (P2). This is the most prevalent reason to report, mentioned by almost 70% of the interviewees. The majority of participants are concerned that others might not recognize a phishing email and can be deceived or "*be hacked*" (P7). These considerations are often drawn from participants' previous experience with phishing (see the `awareness and experience` theme). Only a few participants expect their colleagues to be less careful or not have enough detection skills. For example, "*[I report] so that nobody gets trapped in those phishing emails. I don't, but others can.*" (P1) or "*There are colleagues that would not look carefully enough at the emails (sender addresses), so might fall for it.*" (P6). In contrast, one participant observed that anyone, irrespective of vigilance, can be deceived under certain circumstances. Interestingly, some participants were concerned that other colleagues might be annoyed or bothered by phishing emails: "*Prevent other people from being annoyed by yet another phishing email.*" (P44).

The second most prevalent reason to report is to `protect UNI`. This theme aligns with the previous one in that it captures the desire to protect the university environment, however, with an emphasis on the organization as an institution or as a set of systems, networks, and data rather than as colleagues; for example, to minimize institutional risk (*"to lower risk for the organization"* (P24), *"to protect the confidentiality at UNI"* (P13)), or reputation damage (*"to keep the reputation of UNI high"* (P41)).

The `protect me` theme represents motivations that are connected to safeguarding participants' own security (often mentioned as 'safety') in terms of data, work, and ability to work. Few participants reported their understanding that the protection of others is closely related to their own because "*[...] maybe one day I'm a bit sleepy and I press on the wrong link, so I'd like to prevent others as well because it can happen to anyone*" (P44), and vice-versa.

Mentions of theme `help UNI`, including the IT department, refer to the intention to assist the organization in its mission to keep a functioning environment and avoid negative consequences of an attack, similar to `protect UNI`, but with an explicit reference to helping someone to protect themselves. Such mentions concern the intention to help the IT department achieve wider visibility of an ongoing attack and assist them in thwarting it: "*I report to make the IT department aware, so they can act on it, and they can't monitor everything.*" (P9) or "*My motivation is to let [the IT department] investigate and have more clues about the phishing emails and receive fewer emails.*" (P45). Some participants linked the desire to assist the organization to protect others or themselves, e.g., *"I think they [the IT department] may be able to block links etc., so sending it might help others and avoid them [phishing emails]"* (P34).

Awareness, experience, and doubt. We observe a connection of the themes related to protecting and helping with the theme `experience and awareness`. For instance, participants mentioned news reports of data leaks and ransomware attacks at hospitals and other universities (with the University of Maastricht[4] as a recurrent case). In particular, participant reasoning reveals a degree of awareness in direct connection with how phishing may occur and with the possible consequences of a successful attack. For instance, participants mentioned: "*Relative to previous years' attacks, current attacks are way more realistic, like no spelling mistakes*" (P45) and "*I know the danger phishing can cause, e.g., in Maastricht*" (P9).

At least three interviewees' answers referred to past events, suggesting that awareness (of risks) might derive from experience as well: *"Some colleagues were affected and their computers were locked and held hostage (crypto ransomed), with their PCs unavailable for two weeks. Therefore, the impact can be big"* (P39).

In case of doubt over the legitimacy of an email, participants said they report to `ask for confirmation` as a 'default strategy'. 'Report in case of doubt' is often a common recommendation within organizations and in awareness material. However, this may be at odds with efficient security processes at the organization: too many reports can increase the workload of IT teams and risk delaying actions on high priority reports [8].

In one instance, a participant said they would ask their colleagues and, eventually, send it to the IT department: *"I also ask for confirmation from my colleagues if they received it too. If so, they often tell they 'Throw it away' and don't report"* (P29); another participant mentions reporting as a 'default strategy' to deal with suspicious messages: *"Always, if I don't trust, I send it to 'abuse' [the reporting inbox]"* (P47). This 'default strategy' appears closely related to the efficacy to report, which relates to knowledge of what to do with suspicious messages (and how to do it) from the `efficacy to report` theme discussed next.

Efficacy and responsibility. The `efficacy to report` theme gathers mentions related to the knowledge and efficacy in reporting suspicious emails, i.e., being confident in the need to report rather than knowing what exactly constitutes phishing. The majority of codes in this theme indicate that participants report because they are told to do so (by the IT department). This connects to the 'default strategy' from the `ask for confirmation` theme (e.g., *"I just send them to [the IT department] and I get a response later that this is indeed phishing and that I don't have to do anything"*, P34), which is enabled by knowing what to do or how to report, as it emerges from the `efficacy to report` theme: *"I was told if I get a suspicious email, I should report, so I did."* (P34) or *"Because I know that there is the abuse [inbox]"* (P19). One participant mentioned that reporting takes *"very little effort"* (P23), suggesting that another enabling factor to report is the easiness of reporting.

Many participants reported suspicious emails driven by a `sense of responsibility`: *"if someone threatens via my email, I think it's my duty to inform [the IT department]"* (P11); comparing reporting to a civic habit (*"like, throwing garbage in a civil way"*, P48) and in one case acknowledging that protection of others is related to their own (*"Security is our all responsibility"*, P32). These answers highlight a sense of conscientiousness towards the organization as a community and appear to be an external motivation driver to report for some employees (as also expressed by one participant in the `loyalty` theme). In one instance, a participant felt responsible

[4]The University of Maastricht was hit by ransomware through a phishing campaign; the case was extensively covered on the national news.

Table 5: Overview of the identified themes per cluster of reporters. The numbers of contributing participants are not intended to reflect the prevalence of the themes among reporters from UNI.

Theme	Contributing participants					
	Tot	clst1	clst2	clst3	clst4	clst5
`Protect others`: colleague or community, their systems, confidentiality and data	34	6	9	6	7	6
`Protect UNI`: organization, employer, their systems network and data	24	4	6	5	3	6
`Protect me: myself and my system, confidentiality and data`	18	5	4	3	2	4
`Help UNI/IT`: intention to assist the university, the IT department in handling the issue of phishing	29	5	8	5	5	6
`Help others`: intention to assist colleagues in handling the issue of phishing	3	1	1	0	1	0
`Loyalty`: to the organization and community as a reason to protect	1	1	0	0	0	0
`Awareness & experience`: awareness of phishing risks and consequences stemming from personal and third party experience	8	2	1	2	2	1
`Ask for confirmation`: reporting in case of doubt over the legitimacy of an email to IT or colleagues	6	2	0	1	1	2
`Sense of responsibility`: the feeling of responsibility or duty to report as a norm, civic duty, or conscientious behavior.	10	3	3	0	2	2
`Efficacy to report`: knowing and being confident about how or why to report	7	2	1	1	3	0
`Annoyance`: the feelings of being annoyed or angry by the unwanted emails or the sender	8	0	4	1	1	2
`Fight hackers`: the desire to fight or punish the attackers, or a feeling of disdain towards the perpetrator as a reason to report	5	2	1	1	1	0
`Detection`: reporting because an email was detected as (suspicious) phishing	5	2	0	0	0	3

for "*the data in the emails*" (P4) as a direct function of their role as secretary.

Annoyance and fight hackers. A considerable number of participants mentioned that they report phishing emails because of the `annoyance` caused by receiving unwanted emails. Some participants express anger at the sender of a phishing email: *"Do they think I'm stupid, or something?"* (P15) and *"get a life, a job, do something better"* (P44). The rest of the mentions are milder and involve junk email in their reasoning (*"If this is the spam emails, they annoy me; therefore, I reported some"*, P31), however, with a distinction between phishing and spam: *"It's pretty rare to receive [phishing] emails, so it's a very low level of being annoyed"*, (P20).

At least in two cases `annoyance` is linked with the desire to `fight hackers` where participants feel disdain or disapproval towards 'hackers' (*"I hate them"*, P9 or *"they should stay out of it"*, P38); or because of the consequences of tolerating such attempts (*"To minimize the chance to enrich themselves by our means"*, P26). This signals that the act of reporting potential phishing emails is not only driven by specific, objective rationales or expectations but also by feelings [7].

Detection. The `detection` theme is a special case of reasons to report, as answers belonging to this theme somewhat elude the scope of reporting, which is an action that typically follows detection. For example, participants stated that they report emails *"When I get an email with a strange request or the email address is not correct"*, (P5) or *"When I don't believe it, it's too good to be true."* (P9), indicating that "they report because they detect". It is worth noting that all respondents in this theme provided additional reasons, such as to `protect` or `help UNI`. Yet, the immediate answer to why they do report suspicious emails was related to detection. This can be a sign of unbalanced prowess between detection and reporting activities [8].

4.2 Differences across clusters

Cluster-wise, we observe limited differences in the themes mentioned by clusters 2, 3, and 4, on the one hand, and by clusters 1 and 5, on the other hand. The latter two are the only clusters containing mentions of `detection`-related reasons to report. Interestingly, participants in these clusters reported significantly more contextual low and technical low (`Cl_Tl`) emails than the other types (see Table 4). This might suggest that the participants' reasoning 'report everything detected' is reflected in their emergent behavior of reporting 'any' suspicious email.

Another observable, qualitative difference in themes concerns cluster 5 where the `ask for confirmation` theme is prevalent as a 'default strategy' when encountering suspicious emails. This may indicate that cluster 5, which reports mostly `Cl_Tl` emails, might show a somewhat lower prowess in terms of handling suspicious phishing emails, or at least lower confidence in evaluating whether reporting for that email is necessary. A possible explanation could be that employees in cluster 5 may lack confidence regarding phishing detection or prefer to delegate every decision-making, act conscientiously by adhering to the rules at the organization, or a combination of both.

5 DISCUSSION

Summary of results. All participants mentioned protecting others/UNI or assisting the organization as a reason to report, making these themes the main factors that drive the reporting of suspicious emails. These factors can be directly linked to the altruistic tendency of individuals in organizations (an antecedent of OCBs), which is a known predictor of intention to report phishing [32]. The third most common theme concerns employees' own protection and, in some cases, it is mentioned as an indirect effect of the protection of others. The remaining themes vary considerably in presence across participants, underlining the high dimensionality of factors that constitute the rationale for reporting. The main themes (protecting and helping) are often, but sparsely, related to the rest of the themes, such as awareness and experience (e.g., in terms of understanding the severity of consequences [56]), to know how and be able to report (e.g., in terms of self-efficacy [29, 32]), or feeling responsible (as in commitment trait [32]) or even angry (as an emotional driver for reporting [7]).

A complex mix of motivations to report. Our results suggest that whereas some themes, such as the ones related to helping and protecting, are common among most reporters, different subjects report a wide and disparate range of additional motivations. These

motivations often appear to interrelate, such as in the themes related to helping and `sense of responsibility`, depicting a complex scenario whereby some subjects feel both an obligation from a sense of duty (e.g., from an internal policy) and the desire to be proactive and helpful within their organization and/or towards their colleagues. Interestingly, we find that *uncertainty* plays a role in actively reporting (as opposed to not taking action), with some subjects reporting 'just in case' a threat is present (`ask for confirmation`). Further, our results uncover previously unseen motivations to report phishing emails: `annoyance` and `fight hackers`. The former falls in the category of intrinsic rewards whereby the person wants to get rid of unwanted emails (while still distinguishing phishing from spam). The latter reveals a more visceral motivation to act, signaling a certain level of personal relevance and, thus, a higher likelihood to act on it [7]. Since these motivations have not been investigated previously and are not directly linked with existing models, researchers may test such hypotheses in future experiments to understand whether these factors can contribute to explain reporting behaviors. Finally, we did not observe significant differences in themes across clusters (possible explanations are discussed in Section 5.1). Qualitative observations concern clusters 1 and 5 (reporting mostly low contextual and technical believability emails) vs. other clusters: motivations to report mentioned in clusters 1 and 5 are mainly linked to themes `ask for confirmation` and `detection`. As seen in Section 4.2, this relation might suggest a lower prowess in dealing with suspicious phishing emails for clusters 1 and 5, as opposed to cluster 3 (mostly high believability emails), assuming similar base rates of (type of) received phishing emails. If this relation is confirmed, a phishing reporting mechanism able to gauge the reasons for reporting a given email can provide a series of 'reasons to report' options in the user interface of an email client upon the reporting action (e.g., similar to [49]), that include an 'ask for confirmation' option. Such a mechanism might be valuable for users who keep selecting 'ask for confirmation' as a reason to report whereby they might be presented (targeted) training opportunities to uplift their phishing detection skills [12]. Further, whereas this study focuses on reasons to report phishing emails, different motivations may emerge for reporting other types of suspicious, yet not phishing, emails such as spam. Whereas these reports may not be about an actual security threat, follow-up studies on this may reveal additional nuances on the relation between threat perception and (preventative) user actions.

The importance of keeping the reporter 'in the loop'. Another promising research venue on reporting concerns how feedback and (public) acknowledgment of employees successfully reporting attacks might benefit organizational security. Previous work showed that acknowledging reported incidents and validating reported emails can facilitate reporting as a crowd-sourced defense [26, 30, 57]. In this sense, our findings lead to ask the question: given employees' motivations and reasons, what feedback could be provided to encourage reporting again in the future? For example, one interviewee stated: *"Sometimes, when reporting, there is a lack of feedback [...] I might lose motivation because nobody reads this [the report]."* (P13). Indeed, personal feedback on true positive reports encourages employees to report more [30]. Moreover, P13 added *"It would be interesting to have internal statistics once a year or so, maybe this will inspire people to report more"*. This suggests that, by reporting public statistics on recent phishing campaigns and report efforts, employees whose main concern is protecting colleagues might consolidate their self-efficacy as well as motivate others [26, 27, 32]. However, apart unintended effects from incentivized reporting [8, 27], not everyone cares to receive feedback and might instead perceive it as a nuisance: *"When I report an email to [the IT department], they open a ticket and inform me what they've done with this, but I don't care, I'm not curious what they do with it. I just want to give them information."* (P7). Therefore, the question of the *type* of feedback to give reporters assumes a new relevance on its own.

Links to theoretical underpinnings. Our investigation reveals a wide range of reasons and motivations for reporting phishing emails. A first highlight is that several themes concern protecting and helping. These themes closely relate to the protection motivation construct utilized in the Protection Motivation Theory (PMT), which explains a fair share of intentions related to cyber security behaviors [48]. The PMT posits that individuals form their behavior from a cost-benefit analysis where risks associated with the behavior (threat appraisal) are compared to the costs of trying to reduce the risks (coping appraisal) [44, 48]. These factors can lead to a high protection motivation, resulting in protective behaviors, such as reporting phishing emails. Additional themes linked to protecting and helping include `sense of responsibility` and `efficacy to report`, which are clearly related to coping appraisals mechanisms of subjective norms (e.g., civic duty to report) [52] and self-efficacy (e.g., easy to report) [56]; also, the `awareness and experience` theme can be related to threat appraisals antecedents, such as threat severity (e.g., aware of the severe consequences of a successful phishing attack) [48]. Our findings endorse the applicability of PMT to explain the protective security behavior of reporting. However, whereas a large body of research on PMT and information security behaviors focuses on the perceptions and avoidance of phishing (and other threats), no previous work employed PMT to model and predict employees' (intention of) reporting phishing emails, which our findings suggest to be a suitable protective behavior. The only PMT-related research that deals with reporting, does so within the scope of generic security behaviors, such as following information security policies (ISP) that include phishing reporting [32]. However, evidence suggests that models mapping antecedent factors to generic behaviors (ISP that include reporting) may not represent a specific behavior (of intention to report) well [32, 47]. Therefore, understanding what drives reporting behaviors from the well-established PMT point of view can be crucial to identify potential interventions aimed at individuals that may, but *do not* report phishing; for instance, individuals who are less aware of the threat severity in their organization or of the efficacy of their response to the threat (reporting to the IT department) might benefit from training interventions, such as tailored awareness programs [32] and role-playing training [12].

Similarly to PMT, many themes appear to reflect the relationship between individuals' traits that affect Organization Citizenship Behaviors (OCBs) and phishing reporting. For example, the distinction between protecting colleagues and the organization as a whole can be framed in terms of the individual- vs. organization-directed Organizational Citizenship Behaviors (OCB) [38, 59], which include

reporting phishing to the IT department. Participants who direct their reasoning toward helping and protecting the organization as a whole might share traits and attitudes related to organizational-directed OCBs, whereas participants mentioning protecting others might score higher on traits related to individual-level OCB. For example, among the individual OCB predictors that were shown to influence intention to report [32], altruism aligns well with the (relatively prevalent in our data) themes related to *protecting* and *helping*.

Other organization-level OCB traits, such as commitment and civic virtue, appear connected to reporting behaviors motivated by the desire to `protect UNI` and by the `sense of responsibility`, although previous work did not confirm such relations yet [32]. Moreover, our findings on the `ask for confirmation` theme (as a 'default strategy' when encountering suspicious emails) opens the question of whether more reports equals more security, as the efficiency of IT teams might suffer due to too many reports [8]. Indeed, there may be individuals who choose to not report to avoid creating an additional burden for their colleagues, as seen with the sportsmanship OCB trait [32]. Therefore, further research on drivers of phishing reporting is necessary to foster a sustainable security culture [32] and develop efficient collective defenses against phishing [8, 12].

5.1 Limitations/Threats to validity

As is the case with any empirical study, the validity of our conclusions might be threatened by various reasons. Our research combines a quantitative (identification of the clusters) and a qualitative component (analysis of interviews). For the quantitative component we adhere to the well-established threats to validity framework of Shadish, Cook, and Campbell [46] and guidelines based on it [60].[5] This framework is inappropriate for the qualitative component, and hence in our reflection, we adhere to the guidelines of Lincoln and Guba [31]. **Quantitative.** One of the main *constructs* of our study is believability; the validity of our conclusions can, hence, be threatened by the operationalization of this construct. These threats are partly inherited from the previous work on this topic [20, 28] and partly stem from the joint identification of the believability features suited for automation (Section 3.2.1). To address the latter threats, we have to ensure that agreement has been reached between all four investigators. This threat is closely related to the instrumentation, i.e., errors introduced by the believability classifier (Section 3.2.1), that threaten the *internal validity* of the study. To mitigate this, the development of the classifier followed best practices involving iteratively analyzing false positive and false negative outcomes over independent training sets until classification performance was deemed sufficient. Further, UNI conducts internal phishing awareness campaigns that may affect the reporting rates considered by our sampling strategy. As the degree to which awareness campaigns may affect behaviour is unknown, we rely on our internal knowledge of UNI and employ a best-effort approach and remove sampled reported emails that are likely to belong to an internal awareness campaign. We expect any residual effect to be minimal and not impacting study results. *External validity* related to the criteria used to classify emails that may depend on the type or frequency of emails normally received at UNI (e.g. whose employees may be used to emails from, for example, predatory publishers). **Qualitative.** To ensure *credibility* of the study, we focus on UNI, a university we have been familiar with for an extended period. To ensure *transferability* of our findings, in this report, we provide a detailed description of the interview and analysis process such that the person who might be interested in transferring the study insights can decide whether they might apply to their situation. Finally, to ensure *dependability*, i.e., the ability to audit the process, we provide the audit trail from the interview data to process notes (themes and examples of quotes corresponding to each theme).

Finally, we discuss the **limitations** of our work. The exploratory nature of the study required us to focus on the qualitative analysis and no quantitative insights should be derived from it. The overview of the themes per cluster in Table 5 does not necessarily reflect the prevalence of different reasons for reporting. Follow-up studies should investigate several complementary explanations for the lack of differences across clusters: a strong variability of motivations (other than protecting) between subjects, fundamentally different base rates of the type of phishing emails received by participants, and the classification of phishing believability is not able to reflect the real believability of emails to derive the clusters.

[5]We are mindful of the ongoing debate on trade-offs in research study design and threats to validity induced by these trade-offs, taking place in many scientific disciplines [11, 35, 61]. However, in the absence of a commonly accepted alternative to the threats to validity framework of Shadish, Cook, and Campbell, we adhere to it, acknowledging its limitations [42, 43].

6 CONCLUSION

In this work, we investigated what motivations drive users to report suspicious emails and which types of emails they report. To this end, we sampled and interviewed $n = 49$ employees from the pool of phishing reporters at a medium-sized European technical university. The results show that protecting and helping the organization and others are the main factors that drive the reporting of suspicious emails. Other factors for reporting phishing emails are a sense of responsibility, awareness of the negative consequences they can lead to, or insecurity. Interestingly, our results show phishing reporting is also driven by feelings such as annoyance and anger towards the attackers. By relating our findings with PMT, we identified relevant insights and promising directions for future work.

Acknowledgments. This work is supported by the INTERSCT project, Grant No. NWA.1162.18.301, funded by the Netherlands Organisation for Scientific Research (NWO).

REFERENCES

[1] University of Chicago 2020. *Ethical Guideline for Online Interviews - Virtual Ethnographic Methods | Class Research Portfolio.* University of Chicago. https://voices.uchicago.edu/202003sosc20224/2020/06/25/ethical-guidelines-for-online-interviews/

[2] Luca Allodi, Tzouliano Chotza, Ekaterina Panina, and Nicola Zannone. 2020. The Need for New Antiphishing Measures Against Spear-Phishing Attacks. *IEEE Security & Privacy* 18, 2 (2020), 23–34.

[3] Sebastian Baltes and Paul Ralph. 2022. Sampling in software engineering research: a critical review and guidelines. *Empirical Softw. Engg.* 27, 4 (2022), 31 pages. https://doi.org/10.1007/s10664-021-10072-8

[4] Adam Binks. 2019. The art of phishing: past, present and future. *Computer Fraud & Security* 2019, 4 (2019), 9–11. https://doi.org/10.1016/S1361-3723(19)30040-5

[5] C. Bird. 2016. Interviews. In *Perspectives on Data Science for Software Engineering.* Morgan Kaufmann, 125–131. https://doi.org/10.1016/B978-0-12-804206-9.09992-X

[6] Jan-Willem Bullee, Lorena Montoya, Marianne Junger, and Pieter Hartel. 2017. Spear phishing in organisations explained. *Information & Computer Security* 25, 5 (2017), 593–613. https://doi.org/10.1108/ICS-03-2017-0009

[7] Pavlo Burda, Luca Allodi, Abdul Malek Altawekji, and Nicola Zannone. 2023. The Peculiar Case of Tailored Phishing against SMEs: Detection and Collective Defense Mechanisms at a Small IT Company. In *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 232–243.

[8] Pavlo Burda, Luca Allodi, and Nicola Zannone. 2020. Don't Forget the Human: a Crowdsourced Approach to Automate Response and Containment Against Spear Phishing Attacks. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 471–476. https://doi.org/10.1109/EuroSPW51379.2020.00069

[9] Pavlo Burda, Luca Allodi, and Nicola Zannone. 2024. Cognition in Social Engineering Empirical Research: A Systematic Literature Review. *ACM Trans. Comput. Hum. Interact.* 31, 2 (2024), 19:1–19:55.

[10] Pavlo Burda, Tzouliano Chotza, Luca Allodi, and Nicola Zannone. 2020. Testing the Effectiveness of Tailored Phishing Techniques in Industry and Academia: A Field Experiment. In *Proceedings of the 15th International Conference on Availability, Reliability and Security (ARES '20)*. ACM, New York, NY, USA, 1–10. https://doi.org/10.1145/3407023.3409178

[11] Giliberto Capano and Isabelle Engeli. 2022. Using Instrument Typologies in Comparative Research: Conceptual and Methodological Trade-Offs. *Journal of Comparative Policy Analysis: Research and Practice* 24, 2 (2022), 99–116. https://doi.org/10.1080/13876988.2020.1871297

[12] Xiaowei Chen, Margault Sacré, Gabriele Lenzini, Samuel Greiff, Verena Distler, and Anastasia Sergeeva. 2024. The Effects of Group Discussion and Role-playing Training on Self-efficacy, Support-seeking, and Reporting Phishing Emails: Evidence from a Mixed-design Experiment. https://doi.org/10.1145/3613904.3641943 arXiv:2402.11862 [cs].

[13] Kang Leng Chiew, Kelvin Sheng Chek Yong, and Choon Lin Tan. 2018. A survey of phishing attacks: Their types, vectors and technical approaches. *Expert Systems with Applications* 106 (2018), 1–20. https://doi.org/10.1016/j.eswa.2018.03.050

[14] Prateek Dewan, Anand Kashyap, and Ponnurangam Kumaraguru. 2014. Analyzing social and stylometric features to identify spear phishing emails. In *Symposium on Electronic Crime Research*. IEEE, 4–8. https://doi.org/10.1109/ECRIME.2014.6963160

[15] R. Dhamija, J. Tygar, and M. Hearst. 2006. Why Phishing Works. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, 581–590.

[16] Verena Distler. 2023. The Influence of Context on Response to Spear-Phishing Attacks: an In-Situ Deception Study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM, 1–18. https://doi.org/10.1145/3544548.3581170

[17] Steve Easterbrook, Janice Singer, Margaret-Anne Storey, and Daniela Damian. 2008. Selecting Empirical Methods for Software Engineering Research. In *Guide to Advanced Empirical Software Engineering*. Springer London, London, 285–311. http://link.springer.com/10.1007/978-1-84800-044-5_11

[18] A. Ferreira, L. Coventry, and G. Lenzini. 2015. Principles of Persuasion in Social Engineering and Their Use in Phishing. In *Human Aspects of Information Security, Privacy, and Trust (LNCS)*. Springer, 36–47.

[19] Jill J. Francis, Marie Johnston, Clare Robertson, Liz Glidewell, Vikki Entwistle, Martin P. Eccles, and Jeremy M. Grimshaw. 2010. What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology & Health* 25, 10 (2010), 1229–1245. https://doi.org/10.1080/08870440903194015

[20] Kristen Greene, Michelle P Steves, Mary F Theofanos, and Jennifer A Kostick. 2018. User Context: An Explanatory Variable in Phishing Susceptibility. In *Network and Distributed Systems Security (NDSS) Symposium*. Internet Society, 14 pages.

[21] Greg Guest, Emily Namey, and Mario Chen. 2020. A simple method to assess and report thematic saturation in qualitative research. *PLOS ONE* 15, 5 (2020), e0232076. https://doi.org/10.1371/journal.pone.0232076

[22] Ryan Heartfield and George Loukas. 2016. A Taxonomy of Attacks and a Survey of Defence Mechanisms for Semantic Social Engineering Attacks. *Comput. Surveys* 48, 3 (2016), 1–39. https://doi.org/10.1145/2835375

[23] G. Ho, A. Cidon, L. Gavish, M. Schweighauser, V. Paxson, S. Savage, G. Voelker, and D. Wagner. 2019. Detecting and Characterizing Lateral Phishing at Scale. In *USENIX Security Symposium*. USENIX Association, 1273–1290.

[24] Rashina Hoda. 2022. Socio-Technical Grounded Theory for Software Engineering. *IEEE Transactions on Software Engineering* 48, 10 (2022), 3808–3832. https://doi.org/10.1109/TSE.2021.3106280

[25] Markus Jakobsson, Alex Tsow, Ankur Shah, Eli Blevis, and Youn-Kyung Lim. 2007. What Instills Trust? A Qualitative Study of Phishing. In *Financial Cryptography and Data Security (LNCS, Vol. 4886)*. Springer, Berlin, Heidelberg, 356–361.

[26] Matthew Jensen, Alexandra Durcikova, and Ryan Wright. 2017. Combating Phishing Attacks: A Knowledge Management Approach. In *Proceedings of the 50th Hawaii International Conference on System Sciences*. 10. https://doi.org/10.24251/HICSS.2017.520

[27] Matthew L. Jensen, Ryan T. Wright, Alexandra Durcikova, and Shamya Karumbaiah. 2022. Improving Phishing Reporting Using Security Gamification. *Journal of Management Information Systems* 39, 3 (2022), 793–823.

[28] Leon Kersten, Pavlo Burda, Luca Allodi, and Nicola Zannone. 2022. Investigating the Effect of Phishing Believability on Phishing Reporting. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 117–128. https://doi.org/10.1109/EuroSPW55150.2022.00018

[29] Youngsun Kwak, Seyoung Lee, Amanda Damiano, and Arun Vishwanath. 2020. Why do users not report spear phishing emails? *Telematics and Informatics* 48 (2020), 101343. https://doi.org/10.1016/j.tele.2020.101343

[30] Daniele Lain, Kari Kostiainen, and Srdjan Čapkun. 2022. Phishing in Organizations: Findings from a Large-Scale and Long-Term Study. In *Symposium on Security and Privacy (SP)*. IEEE, 842–859.

[31] Yvonna S. Lincoln and Egon G. Guba. 1985. *Naturalistic Inquiry*. Sage Publications.

[32] Ioana Andreea Marin, Pavlo Burda, Nicola Zannone, and Luca Allodi. 2023. The Influence of Human Factors on the Intention to Report Phishing Emails. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3544548.3580985

[33] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 72:1–72:23. https://doi.org/10.1145/3359174

[34] Mary L. McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia Medica* 22, 3 (2012), 276–282. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/

[35] Shawna L. Mercer, Barbara J. DeVinney, Lawrence J. Fine, Lawrence W. Green, and Denise Dougherty. 2007. Study Designs for Effectiveness and Translation Research: Identifying Trade-offs. *American Journal of Preventive Medicine* 33, 2 (2007), 139–154. https://doi.org/10.1016/j.amepre.2007.04.005

[36] K. Molinaro and M.. Bolton. 2018. Evaluating the applicability of the double system lens model to the analysis of phishing email judgments. *Computers & Security* 77 (2018), 128–137.

[37] Laurence Morissette and Sylvain Chartier. 2013. The k-means clustering technique: General considerations and implementation in Mathematica. *Tutorials in Quantitative Methods for Psychology* 9, 1 (2013), 15–24. https://doi.org/10.20982/tqmp.09.1.p015

[38] Dennis Organ. 1997. Organizational Citizenship Behavior: It's Construct Clean-Up Time. *Human Performance* 10 (1997), 85–97.

[39] K. Parsons, M. Butavicius, M. Pattinson, A. McCormac, D. Calic, and C. Jerram. 2015. Do Users Focus on the Correct Cues to Differentiate Between Phishing and Genuine Emails?. In *Australasian Conference on Information Systems*. AISEL, 10.

[40] K. Parsons, A. McCormac, M. Pattinson, M. Butavicius, and C. Jerram. 2015. The design of phishing studies: Challenges for researchers. *Computers & Security* 52 (2015), 194–206.

[41] Javier Pastor-Galindo, Pantaleone Nespoli, Félix Gómez Mármol, and Gregorio Martínez Pérez. 2020. The Not Yet Exploited Goldmine of OSINT: Opportunities, Open Challenges and Future Trends. *IEEE Access* 8 (2020), 10289–10292. https://doi.org/10.1109/ACCESS.2020.2965257

[42] Charles S. Reichardt. 2011. Criticisms of and an alternative to the Shadish, Cook, and Campbell validity typology. *New Directions for Evaluation* 2011, 130 (2011), 43–53. https://doi.org/10.1002/ev.364

[43] Martin P. Robillard, Deeksha M. Arya, Neil A. Ernst, Jin L.C. Guo, Maxime Lamothe, Mathieu Nassif, Nicole Novielli, Alexander Serebrenik, Igor Steinmacher, and Klaas-Jan Stol. 2024. Communicating Study Design Trade-offs in Software Engineering. *ACM Trans. Softw. Eng. Methodol.* 33, 5, Article 112 (2024), 10 pages. https://doi.org/10.1145/3649598

[44] Ronald W. Rogers. 1975. A Protection Motivation Theory of Fear Appeals and Attitude Change1. *The Journal of Psychology* 91, 1 (1975), 93–114.

[45] Sibi Chakkaravarthy Sethuraman, Devi Priya V s, Tarun Reddi, Mulka Sai Tharun Reddy, and Muhammad Khurram Khan. 2024. A comprehensive examination of email spoofing: Issues and prospects for email security. *Computers & Security* 137 (2024), 103600. https://doi.org/10.1016/j.cose.2023.103600

[46] William R. Shadish, Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company.

[47] Teodor Sommestad and Jonas Hallberg. 2013. A Review of the Theory of Planned Behaviour in the Context of Information Security Policy Compliance. In *Security and Privacy Protection in Information Processing Systems (IFIP Advances in Information and Communication Technology)*. Springer, 257–271.

[48] Teodor Sommestad, Henrik Karlzén, and Jonas Hallberg. 2015. A Meta-Analysis of Studies on Protection Motivation Theory and Information Security Behaviour. *International Journal of Information Security and Privacy (IJISP)* 9, 1 (2015), 26–46.

[49] Nathalie Stembert, Arne Padmos, Mortaza S. Bargh, Sunil Choenni, and Frans Jansen. 2015. A Study of Preventing Email (Spear) Phishing by Enabling Human Intelligence. In *2015 European Intelligence and Security Informatics Conference*. IEEE, 113–120. https://doi.org/10.1109/EISIC.2015.38

[50] Michelle Steves, Kristen Greene, and Mary Theofanos. 2020. Categorizing human phishing difficulty: a Phish Scale. *Journal of Cybersecurity* 6, 1 (2020), 9.

[51] Robert L. Thorndike. 1953. Who belongs in the family? *Psychometrika* 18, 4 (1953), 267–276. https://doi.org/10.1007/BF02289263
[52] Hsin-yi Sandy Tsai, Mengtian Jiang, Saleem Alhabash, Robert LaRose, Nora J. Rifon, and Shelia R. Cotten. 2016. Understanding online safety behaviors: A protection motivation theory perspective. *Computers & Security* 59 (2016), 138–150. https://doi.org/10.1016/j.cose.2016.02.009
[53] Rohit Valecha, Pranali Mandaokar, and H. Raghav Rao. 2022. Phishing Email Detection using Persuasion Cues. *IEEE Transactions on Dependable and Secure Computing* 19, 2 (2022), 747–756. https://doi.org/10.1109/TDSC.2021.3118931
[54] A. Van Der Heijden and L. Allodi. 2019. Cognitive triaging of phishing attacks. In *USENIX Security Symposium*. USENIX Association, 1309–1326.
[55] J. Wang, T. Herath, R. Chen, A. Vishwanath, and H. Rao. 2012. Research Article Phishing Susceptibility: An Investigation Into the Processing of a Targeted Spear Phishing Email. *IEEE Transactions on Professional Communication* 55, 4 (2012), 345–362. https://doi.org/10.1109/TPC.2012.2208392
[56] J. Wang, Y. Li, and H.R. Rao. 2017. Coping responses in phishing detection: An investigation of antecedents and consequences. *Information Systems Research* 28, 2 (2017), 378–396. https://doi.org/10.1287/isre.2016.0680
[57] Emma J Williams, Joanne Hinds, and Adam N Joinson. 2018. Exploring susceptibility to phishing in the workplace. *International Journal of Human-Computer Studies* 120 (2018), 1–13.
[58] Emma J. Williams and Danielle Polage. 2019. How persuasive is phishing email? The role of authentic design, influence and current events in email judgements. *Behaviour & Information Technology* 38, 2 (2019), 184–197. https://doi.org/10.1080/0144929X.2018.1519599
[59] Larry J. Williams and Stella E. Anderson. 1991. Job Satisfaction and Organizational Commitment as Predictors of Organizational Citizenship and In-Role Behaviors. *Journal of Management* 17, 3 (1991), 601–617.
[60] Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, Björn Regnell, and Anders Wesslén. 2012. *Planning*. Springer, Berlin, Heidelberg, 89–116. https://doi.org/10.1007/978-3-642-29044-2_8
[61] Manuel Wolff and Annegret Haase. 2020. Viewpoint: Dealing with trade-offs in comparative urban studies. *Cities* 96, Article 102417 (2020), 7 pages. https://doi.org/10.1016/j.cities.2019.102417
[62] Olga A. Zielinska, Allaire K. Welk, Christopher B. Mayhorn, and Emerson Murphy-Hill. 2016. A Temporal Analysis of Persuasion Principles in Phishing Emails. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 60, 1 (2016), 765–769. https://doi.org/10.1177/1541931213601

A DETAILS ON EMAIL CLASSIFICATION AND REPORTER CLUSTERING

To obtain an insight into the overall email characteristics, we sampled 20 reporters and classified the resulting 131 emails reported by them. We manually classified the emails as phishing vs. not phishing (i.e., legitimate/spam), and high contextual and/or high technical believability. Given the experience of the investigators in the context of UNI, classifying phishing and not phishing emails was a straightforward task with the dataset at hand. The classification of believability, however, was carried out deductively by four investigators by means of the a-priori defined criteria of Section 2. To ensure the quality of the analysis, the four investigators coded the emails by discussing each sampled email, identifying points of disagreements and resolving disagreements, and iteratively updating the analysis of emails [33]. Given the limited amount of emails to be coded, two co-coding sessions were enough to resolve disagreements and update the criteria definitions. Out of the 131 emails, 42 (32%) were deemed as not phishing and 89 as phishing (68%), and no ambiguous or unknown type of emails were identified. The believability classification results are reported in Table 6 for all 131 sampled emails, and in Table 7 for only phishing emails. The abuse inbox contains emails that vary largely in content as well as in contextual and technical believability. The majority of sampled phishing emails present a low believability on either one of the believability dimensions (19 and 29) or on both (25), and emails high on both dimensions are only 16 (≈18%).

Table 6: Outcome of manual classification of 131 emails.

		Technical	
		low	high
Contextual	low	56	36
	high	21	18

Table 7: Outcome of manual classification of 89 phishing emails (removing the non-phishing emails from the total sampled 131 emails).

		Technical	
		low	high
Contextual	low	25	29
	high	19	16

Importantly, we observe a high variability of features in the data that affect contextual and technical believability the most (i.e., `fromAddress`, `subject` and `body`). Previous literature on investigating characteristics of phishing attacks reports similar limitations [18, 50]. When attempting to characterize more advanced features in phishing emails, previous efforts rely on, for example, massive historical data with ground truth [23] or simplify the relevant features to satisfactory levels of approximation within the scope of the study [54].

The outcome of the review suggests that implementing a machine learning approach to identify high contextual and technical believability reports might be inappropriate for our goals. On the one hand, an unsupervised method to detect similar features might identify sufficiently large phishing or spam campaigns, but it will unlikely identify the less frequent sophisticated emails. On the other hand, *manually* labeling the features that determine a high contextual and technical believability to enable a supervised approach for the purpose of our sampling strategy and to answer the research question would be impractical. For instance, building a training set would require thousands of emails to be labeled by experienced agents with contextual knowledge. Given the relatively static structure of the emails and the specific nature of our dataset of emails that come from only one organization (e.g., identifying targetization towards UNI can be done with a regular expression matching the -short- name of UNI), we apply a rule-based classifier, with the exact rules reported in Table 8. We selected email features that showed a lower variability (e.g., mentions of UNI, as opposed to pretexts matching UNI context) and features that are more robust indicators of contextual or technical believability (e.g., free web hosting domains in the URLs signal a low technical believability).

To evaluate the classification performance of the rule-based approach, we used the manually classified emails as the ground truth and ran the naive classifier on it. Table 9 and 10 report the true positives and negatives, and false positives and negatives for the contextual and technical believability classifiers, respectively. The contextual believability classifier has an accuracy of 88.6% and a precision of 76,1%. The technical believability classifier has an accuracy of 72.5% and precision of 78,1%.

Fig. 5 shows the distribution of the fractions of reported emails across the combinations of Contextual and Technical features.

Table 8: Classification rules.

Criterion	Type	Class
A regex with UNI and variations	contextual	high
Otherwise	contextual	low
If any, the payload URL contains: a domain of popular URL shorteners: bit.ly, 1drv.ms, is.gd, tinyurl.com, bit.do, cutt.ly, s.id, rebrand.ly, ht.co, clck.ru, bit.do, rplg.co	technical	high
a domain of popular file hosting services: dropbox.com, drive.google.com, docs.google.com, box.com, mega.nz, onedrive.live.com, forms.office.com, icloud.com, nextcloud.com, spideroak.com, idrive.com, pcloud.com, mediafire.com, tresorit.com, egnyte.com, sugarsync.com, storegate.com, opendrive.com, jungledisk.com, carbonite.com, flipdrive.com, filesanywhere.com, elephantdrive.com, adrive.com, clck.ru	technical	high
a homograph attack in the link based on the strings related to UNI: [redacted for submission] with Levenstein distance is either 1 or 2 from the strings in domain and subdomain	technical	high
a domain of popular free web hosting services*: weebly.com, 000webhost.com, 000webhostapp.com, x10Hosting.com, wix.com, ucoz.com	technical	low
Not any of these: zip, rar, 7z, doc, docx, docm, xls, xlsx, ppt, pptx, pdf, jpeg, jpg, png, gif, exe	technical	high
Otherwise	technical	low

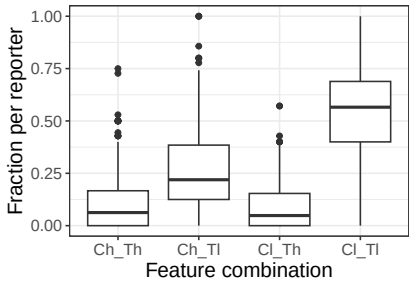
* Overrides previous URL criteria.

Table 9: Classification performance for contextual believability.

		Reference	
		low	high
Prediction	low	81	4
	high	11	35

Table 10: Classification performance for technical believability.

		Reference	
		low	high
Prediction	low	70	29
	high	7	25



	Min.	Q1	Mdn	Mean	Q3	Max.
Ch_Th	0.000	0.000	0.063	0.107	0.167	0.750
Cl_Th	0.000	0.000	0.048	0.088	0.154	0.571
Ch_Tl	0.000	0.125	0.219	0.260	0.385	1.000
Cl_Tl	0.000	0.400	0.566	0.545	0.689	1.000

Figure 5: Distribution of fractions of reports across the combinations of Contextual and Technical features (Boxplot on the left and detailed statistics on the right).

Table 11: Participants' demographics

Variable	Value	Freq.
Gender identity	Female	29
	Male	20
Role	Faculty (Assistant, Associate, Full Professor)	6
	Lecturer	1
	PhD student/PostDoc	3
	Manager	4
	Support staff	27
	Secretary	8
Reporting outside UNI	Yes	25
	No	24
Aware of UNI's ISP	Yes	11
	No	38

B DETAILS ON PARTICIPANTS

Table 11 reports the demographics of the participants.

Table 12: Codebook with examples.

Theme	Description	Example
Protect me	Protect oneself from being a victim, protect own data, computer, or system. This includes avoiding negative consequences, such as data theft or not being able to work.	I want to feel safe in my email [inbox]. (P27) I want to be in the picture. I report it because I want to know if I should worry that my laptop has a virus or what is it (P4)
Protect others	Includes the concept of protecting others, people, colleagues, or community as a reason to report a suspicious email. This extends to protecting others' systems and data.	To protect the community (P13) Because it's not only my inbox to receive these emails but also other people (P13) Warning people for certain accounts [those who received it as well] (P12)
Protect UNI	Includes the concept of protecting UNI as a reason to report a suspicious email. This extends to protecting systems, networks, and data.	Keep the [UNI] network safe (P14)
Loyalty	Explicit expression of loyalty to UNI (as a reason to report).	Loyalty to [UNI] and the community. (P29)
Help UNI/IT	Help/assist UNI and/or specifically its IT department in their effort to handle the issue of phishing emails/protecting infrastructure/employees.	Sometimes I know it's only spam, which I also send to abuse, so they block them (P27) Warn [the IT department], so they block access to a website and prevent further damage (P37)
Help others	Help/assist colleagues in their efforts to avoid falling for phishing.	Because I can help others with it. (P29)
Sense of responsibility	The person feels/knows it's their duty/responsibility to report (or to help/protect UNI/IT/colleagues/data), either because reporting is the norm, a civic duty or simply "the right thing to do".	Then it should be a normal thing to report. [...] I think this is good practice, and should be general practice. (P2)
Awareness & Experience	The person is aware of the risks and consequences of phishing/not reporting. This includes previous experience w.r.t. personal experience and/or others' or in the news.	I know the danger phishing can cause, e.g., in Maastricht. (P30) It's one of the biggest risks and nuisances nowadays (P30)
Ask for confirmation	Enquiring IT/other colleagues to confirm or refute that an email is phishing as a reason to report.	Also, to ask for confirmation that it is phishing. (P12)
Efficacy to report	Statements on (having) the efficacy to report (e.g., because it's easy). Includes the case of knowing how to report and knowing that they should do this (e.g. being told once).	Because I know that there is the abuse@UNI.edu. (P15) Once I was asked to do it, so I do it. (P27)
Annoyance	Statements about the feelings of being annoyed or angry at the attacker/attack. Sometimes refers to being annoyed by spam.	Hopefully, I will not get them back in my mailbox because it's so annoying. (P43)
Fight hackers	Statements on the desire to fight or punish the attackers, or a feeling of disdain towards the perpetrator as a reason to report.	I don't approve phishing attacks. I want to correct them; (P28)
Detection	When they answer the question "how do you detect" instead of "why do you report". For example, reporting because the pretext is strange/suspicious, they do not trust it or the sender is unknown/impersonated.	When I don't believe it, it's too good to be true. (P27) I don't know sender name (P18)

C DETAILS ON INTERVIEWS

Listing 1: Interview guide.

```
Consent being recorded
Remind they can pause or stop the interview whenever they want.

The goal of our research at the [research] group of [department] is
      to explore what and why our colleagues at [UNI] (and
      organizations in general) report suspicious messages to enable
      the development of better tools and methods against such
      threats in the future.

0.1) Have you also reported suspicious emails beyond [UNI]? - e.g.,
      bank, your email provider etc.
0.2) How do you consider the role of the employee in the protection
      of the organization?
0.3) Are you aware of the Information Security Policies at the [UNI]?
       To what extent do you believe they are relevant to your
      security, as opposed to the university's?

Main questions:

1.1) Why would you report a suspicious email?
1.2) What would you say are your main motivations that drive you to
      report an email as phishing?

Closing questions:

These were my questions. Is there any comment or feedback that you
      would like to share about this interview or phishing/reporting
      in general?
```

The interview guide is shown in Listing 1. The guide begins with a brief description of the study goals. We then asks introductory questions to familiarize the interviewee with the conversation format. The study description and introductory questions were crafted to avoid priming the participants on the follow-up questions. Further, we asked high level questions on the rationale and motivations to report suspicious emails to the IT inbox.

We wrapped up the interview by asking about the participants' role at the university from a multiple-choice list derived from the UNI organization chart (cf. Table 11), and by encouraging them to share any comments, feedback, or concerns over the interview and the study.

D CODEBOOK

The codebook contains a short description of associated codes that (will) form a theme, an inclusion/exclusion criterion, and several (counter) examples. Initial themes were formed upon new cards defining new groups, and with each newly formed theme, a re-sorting of cards was applied to meaningfully assess if and which groups need to be reformed. Codebook definitions were added and/or updated for the new themes (akin to the *constant comparison* [24]). The codebook with examples is reported in Table 12.