Presenting Suspicious Details in User-Facing E-mail Headers Does Not Improve Phishing Detection

Sarah Y. Zheng
UCL

Ingolf Becker
UCL

Abstract

Phishing requires humans to fall for impersonated sources. Sender authenticity can often be inferred from e-mail header information commonly displayed by e-mail clients, such as sender and recipient details. People may be biased by convincing e-mail content and overlook these details, and subsequently fall for phishing. This study tests whether people are better at detecting phishing e-mails when they are only presented with user-facing e-mail headers, instead of full e-mails. Results from a representative sample show that most phishing e-mails were detected by less than 30% of the participants, regardless of which e-mail part was displayed. In fact, phishing detection was worst when only e-mail headers were provided. Thus, people still fall for phishing, because they do not recognize online impersonation tactics. No personal traits, e-mail characteristics, nor URL interactions reliably predicted phishing detection abilities. These findings highlight the need for novel approaches to help users with evaluating e-mail authenticity.

1 Introduction

Phishing is a form of deceiving humans to obtain sensitive information in cyberspace. For example, people may receive e-mails that ostensibly come from genuine sources. Nearly half of all security breaches in 2021 involved some form of phishing [53]. Public campaigns and organizational policies have been warning people about it for years. Yet, individuals still receive and fall for them [27, 18, 3, 38]. With the steady growth of global digitization efforts, phishing appears to remain a powerful threat that is unlikely to decline [23, 19, 53]. It is therefore essential to understand why people fall for it.

Previous works suggest that people disproportionately infer e-mail legitimacy from e-mail message content and less so from details in typical *user-facing e-mail header information* (e.g., subject, sender e-mail address, sender display name, timestamp) [24, 57, 40, 60]. For instance, one study found that only the presence or absence of e-mail message features and none of the e-mail header-based features predicted whether participants processed e-mails as genuine or phishing [40]. Moreover, less self-reported attention to sender details was found to predict higher phishing susceptibility [57]. Qualitative studies with general users and IT experts also found that they primarily process e-mails and websites based on content relevance, rather than header details [24, 60]. Since e-mail messages can easily be manipulated and sender details in e-mail headers less so, e-mail headers often contain more reliable indicators of e-mail authenticity. It is thus conceivable that general phishing susceptibility could be driven by users' inattention to e-mail header information typically displayed in e-mail user interfaces. Remarkably, this hypothesis has not been tested empirically before.

This study aims to see if people are better at detecting phishing e-mails when they can only see e-mail header details. If so, simple changes in inbox user interfaces (UIs) that shift people's attention to e-mail headers could help to reduce phishing susceptibility, e.g., by highlighting an external sender's e-mail address. Participants are expected to be better at detecting phishing e-mails with suspicious source details when they can only see the e-mail headers, compared to when the full e-mails are displayed. Participants who are presented with full e-mails are expected to be at least as accurate as those who only see the e-mail message contents and subject lines, if people indeed are generally ignorant toward e-mail headers. This approach gives users the benefit of the doubt on their ability to recognize e-mail impersonation tactics.

{sarah.zheng.16, i.becker}@ucl.ac.uk

USENIX Symposium on Usable Privacy and Security (SOUPS) 2022.
August 7–9, 2022, Boston, MA, United States.

1.1 Contributions

- This is the first study to test whether people fall for phishing because they have overlooked suspicious details in common user-facing e-mail headers. Results show that most participants were not able to detect most of the phishing e-mails in the face of suspicious signals, regardless of whether they saw full e-mails or just e-mail header details. This strongly suggests that most people do not recognize deception tactics that are often used in phishing, even when they cannot be distracted by persuasive e-mail messages. This finding has important implications for tool developments to support users in phishing detection.
- This study used a realistic e-mail filtering task with rendered e-mails instead of screenshots. This allowed for more reliable measurements of phishing detection ability and tracking natural user interactions with e-mails, e.g., hovering over links.
- Participants performed a task where phishing detection was a secondary task.
- The sample was representative for age and gender of the UK population (N=252).
- It provides additional evidence that demographics, personality traits and privacy concerns do not reliably predict individual phishing susceptibility.

The next section discusses prior research on phishing and misinformation susceptibility, as well as anti-phishing interventions that motivated the design of the present study. Section 3 details the study's setup and analysis approach. The results in Section 4 are structured around three key findings. Their implications are described in Section 5, before concluding the paper in Section 6.

2 Related literature

Prior studies explored online deceptions such as fake news and phishing through information processing theories, and factors that may explain individual differences in phishing susceptibility.

2.1 Inattention and online deception susceptibility

People use e-mail to communicate about relevant issues and not to detect phishing, so they may primarily read the e-mail message and use cues such as linguistic errors to infer an e-mail's authenticity from [24, 15, 8, 40, 60] and overlook suspicious indicators in source details.

The recent surge of human fake news detection research provides an interesting parallel for understanding how people process digital information and detect suspicious online content. Studies in lab settings as well as with real life Twitter data have provided evidence that user inattentiveness may drive belief in fake news [44, 43]. These findings suggest that susceptibility to fake news is mainly driven by "peripheral" cognitive processing.

Some theoretical models of phishing susceptibility align on the same notion of two distinct human information processing routes: 1. a systematic or "central" route based on careful assessment of phishing features that makes people less likely to fall for phishing, and 2. a less careful, "peripheral" route that increases people's susceptibility to phishing [33, 56, 37, 17]. As people's capacity for central information processing may be bound by their cognitive functioning, phishing susceptibility may be particularly related to markers of cognitive functioning (e.g., attention, memory) and not necessarily someone's age [16].

It may neither be realistic to expect people to use central processing for all e-mails they receive, i.e., carefully checking all details of every incoming e-mail, when most of their e-mails are genuine. Indeed, lower phishing e-mail prevalence has been associated with worse phishing detection [49, 50]. Participants in Singh et al. performed a phishing training task in which either 25, 50 or 75 percent of the e-mails were phishing. Those exposed to higher phishing proportions detected more of the phishing e-mails, but were less precise in doing so. That is, they also marked more legitimate e-mails as phishing. Sawyer and Hancock found a similar effect with more realistic proportions of phishing attacks and termed it the "prevalence paradox": participants who responded to 300 e-mails of which 1% were phishing, performed worse than participants who were presented with 5% or 20% phishing [49].

2.2 Measuring individual phishing susceptibility

There is arguably no single quintessential phishing e-mail with which people's general phishing susceptibility can be measured [28]. Consequently, various phishing e-mail tactics have been described and used in phishing e-mail studies. For example, tactics based on Cialdini and Goldstein's six principles of persuasion [62, 32, 39, 12] are: authoritativeness and urgency of e-mail messages [7], e-mails adapted to recipients' contexts [22], and positive (e.g., monetary gain) versus negative (e.g., losing something valuable if not complying with the sender) e-mail content [14, 61].

The only common tactic used in online deceptions such as phishing seems to be impersonation, where adversaries manipulate information to create the impression that the digital content came from the claimed source. Still, many studies aimed at finding personal traits associated with higher phishing susceptibility relied on one type of phishing e-mail sent out to non-representative participant samples [63, 21, 4, 55, 36, 9, 1]. While these works all found associations between various personal factors and engagement with the simulated phishing e-mails (e.g., clicking on the phishing link or entering personal details), these results need to be interpreted in light of their specific phishing types and sample contexts.

A general theory for online deception susceptibility, whether in the context of phishing or other scams, would predict higher susceptibility through *any* factor known to encourage inattentive information processing. In this view, personal traits such as age and gender are not the most reliable determinants of phishing susceptibility, since they do not necessarily indicate overall inattentiveness. Situational factors such as stress, distractions and user interfaces more likely affect people's information processing capacity [11], and thence, phishing susceptibility.

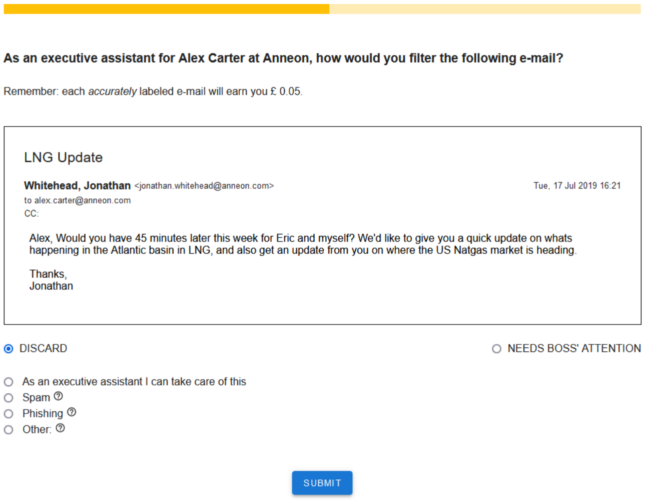
The present study tests if presenting users only with commonly displayed e-mail header information increases their ability to recognize phishing e-mails. It uses a task design that addresses the aforementioned methodological limitations in five ways: 1. participants were presented with a diverse set of phishing e-mails, 2. it used a more naturalistic phishing proportion of approximately 17% compared to previous survey-based studies with 50% phishing [52, 13], 3. it rendered all e-mails from HTML instead of screenshots, which allowed tracking user interactions with the e-mails, 4. the participants sample was representative of the UK population, 5. it used a task context in which phishing detection was not the primary task, to avoid measuring biased phishing detection abilities [41].

3 Methods

This study examines if people's phishing susceptibility is driven by inattention to suspicious e-mail source details. If so, merely presenting them with common e-mail header displays should improve their phishing detection ability. It also tests what factors (personal traits, e-mail characteristics, user interactions with URLs) could reliably predict phishing detection abilities. This section describes the experimental conditions, participants recruitment, task flow, e-mails selection, ethics approval and analysis approach. The task application, e-mail stimuli, scripts, processed data sets and supplementary materials can all be found on the OSF project page: https://osf.io/j9dm8/?view_only=212393f11473447dbea74be547afbd17.

3.1 Study design

This study followed a between-subjects design with three e-mail display conditions: "Control", "Headerless" and "Bodyless". In the Control condition, full e-mails were displayed with typical *user-facing* e-mail header information and body content. That is, headers consisted of the subject line, sender name and e-mail address, recipient(s) e-mail address(es), time it was sent and carbon copy (CC) e-mail address(es) where applicable. In Headerless, only the subject line and body content were displayed. In Bodyless, only said header information was displayed. See Figure 1 for an example full genuine e-mail display and labeling options in Control. Figure 2 shows the same e-mail as displayed in Headerless and Bodyless. E-mail headers of five of the curated phishing variants included sufficient information indicative of malice, see Table 1 and Section 3.4. Participants were randomly allocated to one of the three conditions.



As an executive assistant for Alex Carter at Anneon, how would you filter the following e-mail?

Remember: each *accurately* labeled e-mail will earn you £ 0.05.

LNG Update

Whitehead, Jonathan <jonathan.whitehead@anneon.com> Tue, 17 Jul 2019 16:21
to alex.carter@anneon.com
CC:

Alex, Would you have 45 minutes later this week for Eric and myself? We'd like to give you a quick update on whats happening in the Atlantic basin in LNG, and also get an update from you on where the US Natgas market is heading.

Thanks,
Jonathan

● DISCARD ○ NEEDS BOSS' ATTENTION

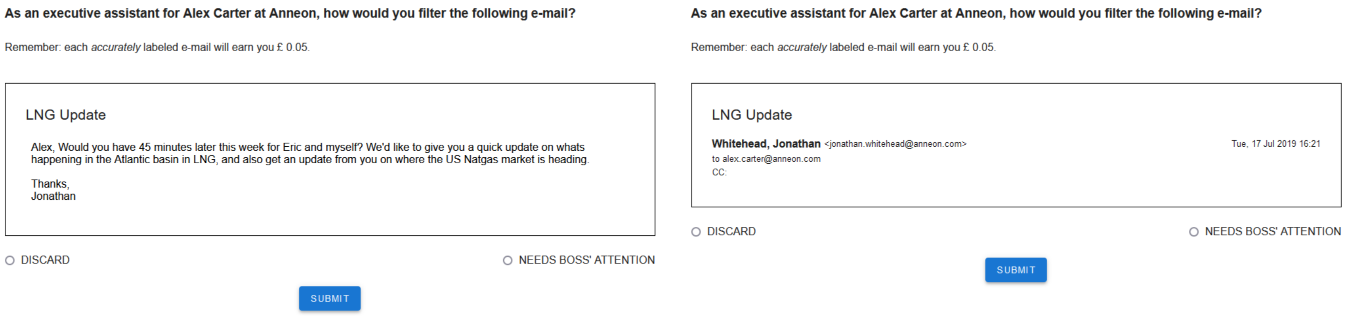
○ As an executive assistant I can take care of this
○ Spam ⓘ
○ Phishing ⓘ
○ Other: ⓘ

SUBMIT

Figure 1: Example display of a genuine e-mail in the main task in the Control condition. The formatting of e-mail header information mirrored that of Gmail. E-mails were displayed one at a time in a solid black frame below the main task question ("As an executive assistant to Alex Carter at Anneon, how would you filter the following e-mail?"). To encourage participants to perform well in the task, a reminder of bonus payment for each correctly labeled e-mail was displayed as well. Initially, participants only see the two primary labeling options ("DISCARD" and "NEEDS BOSS' ATTENTION"). When they select "DISCARD", they need to select a specific reason for why they would discard the given e-mail. This stepwise approach is meant to avoid potentially priming users with the "spam" and "phishing" labels, and mimics people's tendency to filter e-mails based on relevance first.

3.2 Participants

Eighty-four participants were recruited through Prolific for each display condition (total N=252). The sample was representative of the British population in terms of age (μ=46.46, SD=16.99, range=18–68) and gender (50% male). The required sample size was informed by a desired statistical power of 0.80 with a 95% confidence interval for correctly interpreting a medium-sized difference in phishing detection accuracy between three groups. See supplementary materials on OSF



As an executive assistant for Alex Carter at Anneon, how would you filter the following e-mail?

Remember: each *accurately* labeled e-mail will earn you £ 0.05.

LNG Update

Alex, Would you have 45 minutes later this week for Eric and myself? We'd like to give you a quick update on whats happening in the Atlantic basin in LNG, and also get an update from you on where the US Natgas market is heading.

Thanks,
Jonathan

As an executive assistant for Alex Carter at Anneon, how would you filter the following e-mail?

Remember: each *accurately* labeled e-mail will earn you £ 0.05.

LNG Update

Whitehead, Jonathan <jonathan.whitehead@anneon.com> Tue, 17 Jul 2019 16:21
to alex.carter@anneon.com
CC:

Figure 2: Example genuine e-mail display in Headerless (left) and Bodyless (right) conditions. No primary label selection was made yet, hence no further "DISCARD" reasons are displayed here.

for a complete overview of sample statistics. Two participants failed at least two out of three attention check questions in the questionnaires after the labeling task and were excluded from hierarchical regression analyses. This left 83 participants in Control and Headerless and 84 in Bodyless. All participants indicated that they responded honestly and to the best of their ability.

3.3 Task

After giving consent and solving a reCAPTCHA challenge to prevent bot responses, participants saw the task instructions and answered experiential questions on how much professional experience they have with executive assistance tasks and how many e-mails they receive on a daily basis. To cater to the secondary nature of security behaviors [41], the study was disguised as research for a new job application assessment.

Cover story Participants were told to imagine they are working as an executive assistant (EA) for their boss Alex Carter at a fictive petrochemical company called Anneon. Their main task was to filter e-mails for their boss by labeling each e-mail as "Needs boss' attention" or to "Discard" it. This distinction was made to further avoid giving participants the impression that the task was about phishing detection. Only when participants chose "Discard", a specific discard reason had to be selected: "As an executive assistant, I can take care of this", "Spam", "Phishing" or "Other". If they selected "Other", they had to provide a brief free-text format explanation. The task instructions explained the scope of work of the executive assistant and the fictitious boss and showed example e-mails for what should be labeled "Needs boss' attention" or "As an executive assistant, I can take care of this". Tooltips were added to the "Spam", "Phishing" and "Other" options with definitions of the respective labels, identical to the ones participants saw in the instructions. See Figure 1 and Appendix A for the full task instructions.

Task display E-mails were displayed one at a time, center-aligned, with a maximum width of 960 pixels and a 3 pixels thick solid black border with 10 pixels padding all around. The e-mails display order was randomized to avoid successively presenting phishing e-mails. All participants saw this pseudo-randomized order. The e-mail header format mimicked that of Gmail, since Gmail is the most used private e-mail provider. Labeling options were presented with radio buttons beneath each e-mail and a "submit" button to go to the next e-mail. Revisiting previously labeled e-mails was disabled by blocking backward navigation. A progress bar and the lines "As an executive assistant for Alex Carter at Anneon, how would you filter the following e-mail? Remember: each accurately labeled e-mail will earn you £0.05." were continuously displayed at the top of the screen during the labeling task to encourage participants to perform well in the task. The task was developed as a single page application using Vue.js to minimize potential connectivity-related latency issues, and hosted on Google Firebase.

Post-task surveys After labeling all e-mails, participants filled in the short Big Five Inventory (BFI-S [31]) and Internet Users' Information Privacy Concerns (IUIPC [34]) questionnaires. Three attention check questions of the form "Please select (option)" with 7-point Likert scale answer categories were added pseudo-randomly between the main questionnaire items. Finally, participants answered questions on their age, education level, gender, occupational status, income, estimated knowledge about cybersecurity (7-point scale, 1="No knowledge at all", 7="Very knowledgeable"), likelihood to fall for phishing ("Very unlikely", "Unlikely", "Likely" or "Very likely"), frequency of receiving phishing e-mails ("Multiple times per day", "Daily", "Weekly", "Monthly" or "Rarely"), whether they responded honestly and to the best of their ability in the study ("yes" or "no") and whether the study could be improved in any way. Responses to the latter showed that participants found the task "straightforward" and "easy to navigate". See Supplementary Materials for details.

Table 1: Displayed header details of the phishing e-mails set. For the full e-mail bodies, please refer to the Supplementary Materials.

e-mail	**subject**	**sender name**	**sender e-mail**	**recipient e-mail**
p1	URGENT	Sam Jones	sam.jones@annéon.com	alex.carter@anneon.com
p2	are you available?	Jeffrey Skilling	j.skilling@gmail.com	alex.carter@anneon.com
p3	From Mrs.Ameena Essa.	hillb439@gmail.com	hillb439@gmail.com	alex.carter@anneon.com
p4	Alex Carter	Barrister Paul Heywood	office.heywood@gmail.com	alex.carter@anneon.com
p5	Re: [Daily News Update Report] [Account Service] Microsoft account unusual sign-in activity: An order was issued grazie ordine on 06/11/2020.DLIBVCZA	Apple	ponco-gaming2443724@fajardoyandustone.com	mailtdsecure@m-lidscured.com
p6	Anneon File Cash Position Report - Oct19 (1).xlsx has been shared with you	SharePoint Online <no-reply@sharepointonline.com>	esrtn365@microsofia.com	alex.carter@anneon.com
p7	Action Required: Update your payment information now	Microsoft Online Services	no-reply@email.microsoftonline.com	alex.carter@anneon.com
p8	ZOOM Conference Call - April 06, 2020 @ 8:30 - 9:15am	anneon.com ZoomCall	zoom@anneon.formidable.it	alex.carter@anneon.com

3.4 E-mail stimuli

To present participants a realistic proportion of genuine versus phishing e-mails, 47 e-mails were selected of which eight (ca. 17%) were phishing and two were spam. Table 1 shows the header information of the phishing e-mails. See Supplementary Materials for the full e-mails set, including their bodies.

Legitimate e-mails The 37 legitimate e-mails were adapted from the Enron e-mails data set (as retrieved from `http://www.cs.cmu.edu/~enron/`), which contains actual business e-mails from former Enron employees. These were sanitized by substituting all original mentions of "Enron" with "Anneon" to prevent any potential response bias in the case of knowing the Enron scandal and bankruptcy from 2001. Personally addressed e-mails were made to target a gender-neutral executive named Alex Carter. Hyperlinks were replaced by links that opened a blank page in a new window. Despite the age of these business e-mails, they still resemble a natural source of electronic communication representative of corporate e-mails today, as they were mostly sent in plain text and concern realistic ongoing business topics.

Phishing e-mails Four of the phishing e-mails and the spam e-mails were adapted from actual e-mails received by the researchers. Four additional phishing e-mails were selected from various online sources with actual phishing examples. Participants in Bodyless were specifically expected to recognize the suspicious header details in phishing e-mails 1, 5, 6, 7 and 8. In Headerless, phishing e-mails 3–8 were expected to be detected by most participants. Phishing e-mail 2 was only reasonably detectable in Control.

Phishing e-mails 1 and 2 exemplified spear phishing e-mails, which could be detected through careful appraisal of the domain of the sender's e-mail address. If these two phishing e-mails came from an anneon.com e-mail domain, they would have been virtually impossible for participants to detect. However, the domains were annéon.com, representing a homograph attack that should be detectable in Bodyless, and gmail.com, respectively. Phishing e-mails 3 and 4 were "Nigerian prince"-style phishing e-mails in which the sender tells a story about a diseased or deceased relative, after which they seek some form of financial help. The latter two e-mails could not reasonably be detected in Bodyless.

Phishing e-mails 5, 6 and 7 impersonated Apple or Microsoft and asked recipients to log in to secure their account, update payment details or to view a file that was shared with them through SharePoint. Phishing e-mail 5 contained mismatching sender details, a phishing sign that should be detectable in Bodyless. Phishing e-mails 6 and 7 came from suspicious sender e-mail domains that resembled Microsoft's. Phishing e-mail 8 exemplified a Zoom phishing e-mail that surfaced during the COVID-19 pandemic in 2020 and also used a non-sensible e-mail domain. Phishing e-mails 5–8 all contained malicious links, of which the original URLs were made visible on hover through the HTML link "title" attribute. Clicking on them would open a blank page in a new window. Together, this selection represented a diverse set of phishing sub-types.

3.5 Ethics

This study was reviewed and received approval prior to any data collection by the authors' institution's Research Ethics Committee. Participants were compensated at a recommended rate of £7.50 per hour and typically completed the study in 20–35 minutes. Each correctly labeled e-mail yielded a bonus payment of £0.05. Participants in the Headerless condition always received an additional bonus of £0.10 and participants in Bodyless always received an additional £0.15, given that phishing e-mails 1–2 and 2–4 were not reasonably detectable in the respective conditions. The average bonus payment across conditions was £1.25. All responses were collected anonymously.

3.6 E-mail characteristics

Twenty-five e-mail characteristics were computed for all 47 e-mails. This allowed for testing if participants consistently use common e-mail characteristics to infer e-mail authenticity from. For example, whether the sender e-mail address domain was the fictive company name (Anneon), whether the e-mail contained a personal greeting and how urgent the message sounds. E-mail body and subject valence, arousal and dominance (VAD) scores were based on the NRC dictionary [35] to give an indication of authoritativeness and emotional weight of each e-mail's content. For each e-mail, the VAD-scores of all words in the subject line or body that were found in the dictionary were summed and divided over the total number of words in the subject or e-mail body, respectively. Language quality of e-mail body content was based on the number of linguistic errors as found with Beautiful Soup [48], divided over the total number of words in the e-mail body. An overview of all computed e-mail characteristics is included in the Supplementary Materials.

3.7 Analyses

Phishing detection by display condition. To test for the effect of e-mail display on each phishing e-mail's detection proportion, χ^2-tests were run with equal expected detection proportions for all display conditions under the null hypothesis. Detection proportion is the number of participants that labeled the given phishing e-mail as "phishing", divided by the total number of participants in the respective condition. Phishing detection ability is computed as participants' phishing detection precision (i.e., the proportion of e-mails the

Table 2: **Detection proportions for each phishing e-mail per display condition.** Boldfaced expectations are supported by the respective χ^2-test for equal proportions.

e-mail	expectation	Control	Headerless	Bodyless	χ^2	p
p1	worst detection in Headerless	0.16	0.06	0.14	3.80	0.15
p2	highest detection in Control	0.11	0.04	0.13	4.52	0.10
p3	**worst detection in Bodyless**	0.87	0.79	0.16	42.5	<.001
p4	**worst detection in Bodyless**	0.80	0.75	0.12	43.4	<.001
p5	**equal detection for all conditions**	0.81	0.70	0.70	0.87	0.65
p6	equal detection for all conditions	0.29	0.06	0.19	12.1	<0.01
p7	**equal detection for all conditions**	0.46	0.55	0.51	0.58	0.75
p8	**equal detection for all conditions**	0.06	0.04	0.02	1.40	0.50

participant labeled as phishing, that indeed were phishing) and phishing detection recall (i.e., the proportion of all phishing e-mails that the participant detected). Using both metrics allows for more thorough estimates of phishing detection ability, given there were less phishing than legitimate e-mails. Participants with low phishing detection precision and/or low phishing detection recall are regarded as particularly susceptible to phishing. Additional analyses were run with both “Phishing” and “Spam” as true positive labels in the detection ability metrics.

E-mail characteristics regressions. To see if participants used “rule of thumb” tactics in deciding which e-mails had to be discarded as phishing, multiple regressions were computed in R to predict phishing detection proportions for all 47 e-mails, i.e., including false positives, per condition. E-mail characteristics based on sender information (e.g., if the company name was present in the sender e-mail domain) were not included in the Headerless model and body content features (e.g., body content dominance score) were not included in the Bodyless model. The Control model incorporated all e-mail characteristics.

Hierarchical regressions. Hierarchical linear regressions were run to predict phishing detection precision and recall in each condition. This step-wise approach allows for examining the added predictive value of every set of personal traits.

Step 1 included all demographic traits as predictors of phishing detection ability (age, gender, education level, occupational status, income). Step 2 added experiential question responses (prior professional experience with executive assistance, self-reported knowledge of cybersecurity, expectation to fall for phishing, self-reported phishing reception frequency and estimated daily amount of e-mails received). Step 3 added participants’ mean scores on the three IUIPC dimensions (awareness, collection, control). Step 4 added mean scores on the Big Five personality traits (openness, conscientiousness, extraversion, agreeableness, neuroticism). Gender and occupational status were treated as categorical variables, all others as continuous.

Normality of residuals checks were done visually with QQ plots. When additional steps did not significantly reduce the residual sum of squares (RSS), only the effects in the more parsimonious step were interpreted. All hierarchical regressions were analyzed in R [51]. χ^2-, t- and correlation tests were performed using *scipy* version 1.5.4 [54] in Python 3. All results were interpreted against a two-tailed significance level of 0.05, unless noted otherwise.

4 Results

The results are divided into three key analyses. Section 4.1 shows the overall detection rates for each phishing e-mail per display condition. Section 4.2 describes the fitted e-mail feature regressions. Section 4.3 describes the hierarchical regressions on personal traits to predict phishing susceptibility.

4.1 Phishing detection varies widely by e-mail type and is the worst in Bodyless

According to the overall hypothesis, higher phishing detection proportions would be expected in the Bodyless condition for phishing e-mails 1, 5, 6, 7 and 8, which contained clearly suspicious header details. Table 2 shows the phishing detection proportions per condition per e-mail, and corresponding χ^2-tests for proportional equality. Phishing e-mail 5 was detected by the majority of participants in all conditions, as well as the “Nigerian prince”-style phishing e-mails (3 and 4) in Control and Headerless ($\chi^2(2,N=252)=42{,}49,\ p<.001$). Phishing e-mail 7 about updating Microsoft payment details was detected at around chance level in all conditions. The remaining four phishing e-mails were at most detected by 29% of participants across all three conditions.

Most relative detection proportions were as expected, except for phishing e-mails 1, 2 and 6, although the lower detection proportions in Headerless for phishing e-mails 1 and 2 could also be considered as expected at trend level ($\chi^2(2,N=252)=3.80,\ p=.150$; $\chi^2(2,N=252)=4.52,$

Table 3: **Number of participants who hovered over phishing URLs.**

e-mail	**condition**	**N hovered**	**N labeled as phishing or spam**	**URL in body**
p5	Headerless	1	1	https://apple.ngrok.io/3p8sf9JeGzr60+haC9F9mxANtLM
p6	Headerless	2	0	http://25.245.256.02/excel/3p8sf9JeGzr60+haC9F9mxANtLM
p7	Control	3	1	http://office365.microsoft.netgriokgth.com
p8	Control	1	0	https://ngrok.io/b31d032cfdcf47a399990a71e43c5d2a

$p = .100$). Detection of phishing e-mail 6 was worst in Headerless ($\chi^2(2, N = 252) = 12.13$, $p = .010$), while a detection proportion comparable to Control would have been expected in the best case scenario. That is, if participants in Headerless hovered over the hyperlink and recognized the suspicious URL.

Of note is that in Bodyless, one participant labeled phishing e-mail 1 as "Other", despite correctly identifying the homograph attack. They commented "No informative subject, accent on the e in sam.jones anneon email address". This suggests that people may perfectly spot the discrepancy in sender details, but lack the knowledge that these are intentional deception tactics.

4.2 Phishing detection is not predicted by e-mail characteristics

To see if people use consistent rules to infer suspiciousness from e-mail characteristics, linear regressions were run with e-mail characteristics to predict the phishing detection proportions in each display condition. In Control, the full model yielded a significant regression ($F(24,22) = 2.187$, $p = .035$, $R^2 = 0.7047$, $R^2_{adj} = 0.3825$) where more linguistic errors ($\beta = -3.334$, $p = .023$) and presence of the Anneon company name in the sender e-mail address ($\beta = -0.200$, $p = .038$) predicted lower phishing proportions. The former can be explained by the absence of linguistic errors in the phishing e-mails and presence of some grammatical errors and typos in some legitimate e-mails. In Headerless, a multiple regression predicting phishing detection proportions with only e-mail header-based features was not significant ($F(19,27) = 1.425$, $p = .195$, $R^2 = 0.500$, $R^2_{adj} = 0.149$).

In Bodyless, a significant regression ($F(9,37) = 3.186$, $p = .006$, $R^2 = 0.450$, $R^2_{adj} = 0.298$) was fit with all e-mail header-based features. Longer subject lines were found to predict higher phishing detection proportions ($\beta = 0.002$, $p < 0.001$). Phishing e-mail 5 had the longest subject line of all e-mails and had the highest detection rate in Bodyless, which explains this small, but highly significant effect. Since none of the other e-mail characteristics significantly predicted phishing detection across conditions, people do not seem to use consistent strategies in differentiating phishing from genuine e-mails.

Even when participants hovered over phishing URLs, most did not raise suspicion. One common piece of security advice is to check the true URLs of links in e-mails, by hovering over them [46]. To see if people do so, this study tracked and analyzed user interactions with e-mail links. Phishing e-mails 5, 6, 7 and 8 contained malicious URLs. Seven participants hovered over at least one of them. In two cases, the e-mail was labeled as "Phishing". One of the three participants who hovered over the URL in phishing e-mail 7, labeled the e-mail as "Other". For phishing e-mails 5 and 6, no URL hovers were observed in Control, nor for phishing e-mails 7 and 8 in Headerless—see Table 3. This suggests that most participants who labeled phishing e-mails as phishing did not base their judgments on the true URL of linked e-mail contents or did not know what to do with this information.

4.3 Phishing detection is not reliably predicted by personal traits

Overall, phishing detection accuracies varied greatly between participants. The mean phishing detection recall score was 0.46 (SD=0.2) in Control, 0.25 (SD=0.18) in Bodyless and 0.37 (SD=0.16) in Headerless, showing that most people detected less than half of all phishing e-mails. The mean phishing detection precision score was 0.93 (SD=0.18) in Control, 0.73 (SD=0.37) in Bodyless and 0.81 (SD=0.27) in Headerless, suggesting that when people think an e-mail is phishing, they are mostly correct.

To investigate individual differences in phishing susceptibility, hierarchical linear regressions were run to see if personal traits can reliably predict participants' phishing detection recall and precision scores. Table 4 shows ANOVA results that compare the added value of each hierarchical regression step in predicting phishing detection recall from personal traits. None of the steps significantly reduced model RSS compared to step 1 in Control and Headerless, and step 1 regressions did not significantly predict phishing detection recall in Control ($F(7,75) = 1.292$, $p = .266$, $R^2 = 0.108$, $R^2_{adj} = 0.024$), nor in Headerless ($F(7,75) = 1.590$, $p = .152$, $R^2 = 0.130$, $R^2_{adj} = 0.048$). Only adding experiential question responses in step 2 in Bodyless significantly reduced RSS compared to step 1 and showed a significant regression ($F(12,71) = 3.014$, $p = .002$, $R^2 = 0.338$, $R^2_{adj} = 0.226$).

Table 4: **Hierarchical regressions predicting phishing detection recall.** Only step 2 in Bodyless significantly improved phishing detection recall predictions compared to step 1. None of the hierarchical regression steps significantly improved model fits in Control and Headerless compared to step 1. Colors indicate whether the regression fit at the respective step was significant (green) or not (yellow), e.g., step 1 regressions in Control and Headerless were non-significant. RSS = Residual Sum of Squares

		Control			**Headerless**			**Bodyless**		
step	**predictors**	RSS	F (df)	p	RSS	F (df)	p	RSS	F (df)	p
1	demographics	3.13			1.90			2.31		
2	experiential questions	2.80	1.61 (5, 70)	.171	1.82	0.679 (5, 70)	.642	1.80	3.94 (5, 71)	.004
3	privacy concerns	2.70	0.826 (3, 67)	.484	1.65	2.22 (3, 67)	.095	1.78	0.251 (3, 68)	.860
4	Big Five	2.55	0.696 (5, 62)	.629	1.55	0.875 (5, 62)	.503	1.64	1.08 (5, 63)	.379

Table 5: **Hierarchical regressions predicting phishing detection precision.** None of the hierarchical regression steps in any condition significantly improved fit results compared to step 1. Colors indicate whether the regression fit at the respective step was significant (green) or not (yellow), e.g., only the step 1 regression in Control was significant. RSS = Residual Sum of Squares

		Control			**Headerless**			**Bodyless**		
step	**predictors**	RSS	F (df)	p	RSS	F (df)	p	RSS	F (df)	p
1	demographics	2.94			5.29			9.66		
2	experiential questions	2.82	0.659 (5, 70)	.659	5.18	0.332 (5, 70)	.892	8.65	1.71 (5, 71)	.144
3	privacy concerns	2.53	2.63 (3, 67)	.058	4.89	1.36 (3, 67)	.263	8.20	1.27 (3, 68)	.294
4	Big Five	2.30	1.24 (5, 62)	.304	4.40	1.36 (5, 62)	.251	7.42	1.31 (5, 63)	.271

Lower phishing detection recall was predicted by higher age ($\beta = -0.003$, $p = .040$), less frequent self-reported phishing reception ($\beta = -0.057$, $p = .003$) and more professional experience with executive assistance work ($\beta = -0.054$, $p < 0.001$).

None of the steps in the hierarchical regressions showed significant reductions in model residuals when predicting phishing detection precision from personal traits (see Table 5). Therefore, only step 1 regressions are reported further. In Control, the step 1 multiple regression with only demographic traits was significant ($F(7,75) = 2.436$, $p = .026$, $R^2 = 0.185$, $R^2_{adj} = 0.109$). Higher education level predicted higher phishing detection precision ($\beta = 0.037$, $p = .013$) and higher income predicted lower phishing detection precision ($\beta = -0.001$, $p = .018$). Step 1 regressions did not significantly predict phishing detection precision in Bodyless ($F(7,76) = 1.620$, $p = .143$, $R^2 = 0.130$, $R^2_{adj} = 0.050$), nor in Headerless ($F(7,75) = 1.309$, $p = .258$, $R^2 = 0.109$, $R^2_{adj} = 0.026$) conditions. Effect sizes of all significant predictors were small and arguably of limited meaningful value. Note that using a different order of regression steps did not change the results.

Higher age was associated with slower labeling responses in Control ($r = 0.355$, $p = .002$) and Bodyless ($r = 0.341$, $p = .002$). That is, older participants were slower at the task overall. However, no significant associations were found between mean labeling RTs and phishing detection recall or precision in any of the display conditions. This further implies that demographics do not reliably predict phishing detection ability.

Adding “spam” as an accurate phishing detection label does not lead to more consistent results. Some participants may have confused the meaning of “spam” and “phishing”. Hence, additional regressions with personal traits were run where both “spam” and “phishing” were regarded as accurate (true positive) labels for phishing e-mails and false positive labels for legitimate e-mails. This approach yielded prediction improvements for a step 2 regression in Control and step 4 regression in Headerless. Both regressions were significant. In the step 2 regression in Control (demographics and experiential questions; $F(12,70) = 1.973$, $p = .040$, $R^2 = 0.253$, $R^2_{adj} = 0.125$), older participants had a somewhat higher phishing detection recall score ($\beta = -0.004$, $p = .014$). The step 4 regression in Headerless (including all personal traits) predicted phishing detection recall at trend level ($F(20,62) = 1.637$, $p = .072$, $R^2 = 0.346$, $R^2_{adj} = 0.135$), where higher phishing detection recall was predicted by higher mean extraversion ($\beta = 0.040$, $p = .003$) and neuroticism ($\beta = 0.033$, $p = .048$). Higher mean agreeableness predicted lower phishing detection recall ($\beta = -0.040$,

$p = .012$). None of the steps in the hierarchical regressions for the Bodyless condition were significant, meaning none of the personal traits predicted phishing detection recall in Bodyless, even when "spam" was considered as an accurate phishing detection label.

In Control, phishing detection precision was significantly predicted in a step 3 regression with demographics, experiential questions and privacy concerns ($F(7,76) = 1.995$, $p = .029$, $R^2 = 0.309$, $R^2_{adj} = 0.154$). Less frequent self-reported phishing reception ($\beta = -0.041$, $p = .017$) and higher IUIPC "control" dimension scores ($\beta = -0.043$, $p = .014$) predicted lower phishing detection precision. None of the hierarchical regressions predicting phishing detection precision in Bodyless and Headerless were significant.

Altogether, whereas more personal traits were found to predict phishing detection recall by including "spam" as an accurate phishing detection label, the effects remained inconsistent over conditions and effect sizes were of limited meaning. These analyses strongly suggest that personal traits (e.g., demographics, personality traits, privacy concerns) do not consistently relate to how susceptible people are to phishing e-mails.

5 Discussion

Given the rising and increasingly sophisticated threat of phishing e-mails, it is essential to understand why people fall for them and to develop new solutions that reduce phishing victimization. This study highlights the possibility that phishing susceptibility is caused by inattention to suspicious source details found in e-mail headers. It tested the phishing detection ability of a representative sample in an e-mail processing task with different display conditions. Contrary to expectations, participants were not better at detecting phishing when only e-mail header details were displayed. Since the vast majority did detect phishing e-mail 5 in all conditions and the "Nigerian prince" scams in Control and Headerless, low participant motivation to perform well at the task is an unlikely explanation for the low overall detection rates. These findings show that people do not necessarily have a blind spot for e-mail source details, but instead do not recognize deception tactics commonly used in phishing.

The lack of e-mail characteristics predictive of phishing detection proportions confirm the idea that users do not rely on consistent tactics to gauge e-mail authenticity. One heuristic to do so, for instance, is checking if the sender e-mail address domain corresponds with the organization the sender claims to be from. If people used this rule, most participants should have detected phishing e-mails 1, 2 and 6 in Control. Another often given advice to avoid getting phished is to always inspect the actual URL of links in e-mail content, by hovering over them. If people adopted this advice, phishing e-mails 5, 6, 7 and 8 should have been detected by the majority as well. The lack of participants who did so suggests that common anti-phishing advice is not used or that they do not know what to do with the gathered information [46], and may reflect people's general misreading of URL domains when they hover over links [42, 2].

Many existing efforts to reduce phishing victimization rely on some form of training and are widely implemented in organizations and public campaigns already [5, 25, 46, 47, 6, 45, 52, 20, 58, 27, 30, 30, 29]. If the general public followed common cybersecurity advice, this study should have found higher average phishing detection proportions. The low detection rates imply the need for alternative solutions that help people recognize deception tactics used in phishing e-mails.

A strategy would be to target at-risk individuals with personalized anti-phishing interventions. If traits such as demographics were reliable predictors of phishing susceptibility, cybersecurity training could be targeted more specifically at certain demographic groups. However, the present study used an improved sample and task design, and still found no consistent relations between phishing detection and demographics, personality traits, privacy concerns, self-reported cybersecurity knowledge, nor self-reported phishing susceptibility. This accord with results from studies that also used role-playing tasks with a larger variety of phishing e-mails [13, 28, 26]. Whereas more research is needed to see if people with certain traits may be more susceptible to specific types of phishing (e.g., see [32]), interventions solely based on personal traits are not well-justified by the current body of research. Another under-researched direction is to profile within-individual differences in attention and situational changes to predict phishing susceptibility.

Taken together, this study emphasizes the need for research on user-centered techniques to reduce phishing susceptibility. In this approach, knowledge about online deception tactics needs to be accessible and usable for users in real-time. This moves away from conventional time-limited training programs and calls for more interdisciplinary collaboration between software developers and social scientists. An encouraging example from work on fake news detection showed that simply asking Twitter users to think about the veracity of social media articles reduced content sharing from untrustworthy sources [43]. More studies are needed to test similar tactics in e-mail inbox interfaces. Examples include explaining URLs to users when they hover over them [45, 58, 5] and Outlook's external sender warnings. New experiments are being conducted by the authors on new e-mail functionalities in this realm, e.g., showing explainable suspicion scores and changing text colors for suspicious e-mails. Such interventions could provide cost-effective alternatives to anti-phishing training programs that suffer from questionable long-term effectiveness [27, 47] or phishing simulations that bear the risk of damaging employee relationships [59].

5.1 Limitations

The business context of this task may have been difficult to empathize with for participants without corporate experience, although business experience was not needed to recognize the suspicious details in the provided phishing e-mails. Moreover, knowing the business context may not necessarily lead to better phishing detection. Current phishing attacks are often deliberately adapted to organizational contexts, as in phishing e-mails 1, 2, 6 and 8, and real employees have fallen for them. Unfamiliarity with the business context may in fact have prompted people to read the e-mail contents more carefully before deciding what to do with them.

Next, the task interface did not fully mimic an actual inbox. Only the single e-mail display mimicked e-mail displays in Gmail. The task may have been more convincing if situated in an actual inbox, although similar setups were used in previous works [52, 41, 13]. It is also possible that people in Bodyless still ignored sender e-mail addresses and merely based their judgments on the subject line and sender name. Various online and offline eye tracking methods were considered to measure participants' visual attention, but none were able to differentiate users' gaze at such granular levels.

Lastly, this study only asked for participants' self-reported cybersecurity knowledge and not their actual amount of prior cybersecurity education or anti-phishing training. However, equal distributions of variance in cybersecurity training can be expected in each experimental condition, since participants were randomly allocated to either of them.

6 Conclusion

Phishing e-mails are a growing and increasingly sophisticated threat in our daily lives. Whereas e-mail messages can easily be manipulated, altering actual source details is more difficult to achieve. Consequently, phishing e-mails will often show suspicious signs in e-mail header details. When people fail to pay attention to them, they may especially be prone to falling for phishing e-mails. The present study compared people's phishing susceptibility when only e-mail headers were displayed, to when they saw full e-mail messages in a realistic task context. Presenting people merely with e-mail header details was expected to improve phishing detection. Surprisingly, this was not the case. Phishing susceptibility did not seem to be caused by blindness to source details. Instead, the results imply that people do not recognize deception tactics that are often used in phishing. The findings also affirmed that personal traits do not reliably predict phishing susceptibility. Altogether, this study encourages more interdisciplinary development of alternative user-centered tools that help us in the challenge against phishing.

7 Acknowledgments

We thank the anonymous reviewers for their valuable feedback. This research was funded by the Dawes Centre for Future Crime at UCL.

References

[1] Hossein Abroshan, Jan Devos, Geert Poels, and Eric Laermans. Phishing Happens beyond Technology: The Effects of Human Behaviors and Demographics on Each Step of a Phishing Process. *IEEE Access*, 9:44928–44949, 2021. ISSN: 21693536. DOI: 10.1109/ACCESS.2021.3066383.

[2] Sara Albakry, Kami Vaniea, and Maria K. Wolters. What is this URL's Destination? Empirical Evaluation of Users' URL Reading. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, 2020. ISBN: 9781450367080. DOI: 10.1145/3313831.3376168.

[3] Hussain Aldawood and Geoffrey Skinner. Reviewing cyber security social engineering training and awareness programs-pitfalls and ongoing issues. *Future Internet*, 11(3), 2019. ISSN: 19995903. DOI: 10.3390/fi11030073.

[4] Ibrahim Alseadoon, M. F.I. Othman, and Taizan Chan. What is the influence of users' characteristics on their ability to detect phishing emails? *Lecture Notes in Electrical Engineering*, 315:949–962, 2015. ISSN: 18761119. DOI: 10.1007/978-3-319-07674-4_89.

[5] Kholoud Althobaiti, Nicole Meng, and Kami Vaniea. I don't need an expert! making url phishing features human comprehensible. *Conference on Human Factors in Computing Systems - Proceedings*, 2021. DOI: 10.1145/3411764.3445574.

[6] Aurélien Baillon, Jeroen De Bruin, Aysil Emirmahmutoglu, Evelien Van De Veer, and Bram Van Dijk. Informing, simulating experience, or both: A field experiment on phishing risks. *PLoS ONE*, 14(12), 2019. ISSN: 19326203. DOI: 10.1371/journal.pone.0224216.

[7] Maxim Baryshevtsev and Joseph McGlynn. Persuasive Appeals Predict Credibility Judgments of Phishing Messages. *Cyberpsychology, Behavior, and Social Networking*, 23(5):297–302, 2020. ISSN: 21522723. DOI: 10.1089/cyber.2019.0592.

[8] Mark Blythe, Helen Petrie, and John Clark. F for fake: four studies on how we fall for phish. In pages 3469–3478, 2011. DOI: 10.1145/1978942.1979459.

[9] Frank Kun Yueh Chou, Abbott Po Shun Chen, and Vincent Cheng Lung Lo. Mindless response or mindful interpretation: Examining the effect of message influence on phishing susceptibility. *Sustainability (Switzerland)*, 13(4):1–25, 2021. ISSN: 20711050. DOI: 10.3390/su13041651.

[10] Robert B. Cialdini and Noah J. Goldstein. Social influence: Compliance and conformity. *Annual Review of Psychology*, 55:591–621, 2004. ISSN: 00664308. DOI: 10.1146/annurev.psych.55.090902.142015.

[11] Ronald V. Clarke. Situational crime prevention. In redacted by Richard Wortley and Michael Townsley, *Environmental Criminology and Crime Analysis*, Crime Science Series. Routledge, New York, 2nd edition edition, 2008.

[12] Marco De Bona and Federica Paci. A real world study on employees' susceptibility to phishing attacks. In *ACM International Conference Proceeding Series*. Association for Computing Machinery, 2020. ISBN: 9781450388337. DOI: 10.1145/3407023.3409179.

[13] Rachna Dhamija, J. Doug Tygar, and Marti Hearst. Why phishing works. In *SIGCHI Conference on Human Factors in Computing Systems*, pages 581–590, 2006. DOI: 10.1145/1124772.1124861.

[14] Alejandra Diaz, Alan T. Sherman, and Anupam Joshi. Phishing in an academic community: A study of user susceptibility and behavior. *Cryptologia*, 44(1):53–67, 2020. ISSN: 0161-1194. DOI: 10.1080/01611194.2019.1623343.

[15] Julie S. Downs, Mandy Holbrook, and Lorrie Faith Cranor. Behavioral response to phishing risk. Technical report, 2007, pages 37–44. DOI: 10.1145/1299015.1299019.

[16] Natalie C. Ebner et al. Uncovering Susceptibility Risk to Online Deception in Aging. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, 75(3):522–533, 2020. ISSN: 10795014. DOI: 10.1093/geronb/gby036.

[17] Edwin Donald Frauenstein and Stephen Flowerday. Susceptibility to phishing on social network sites: A personality information processing model. *Computers and Security*, 94, 2020. ISSN: 01674048. DOI: 10.1016/j.cose.2020.101862.

[18] Steven Furnell, Kieran Millet, and M. Papadaki. Fifteen years of phishing: can technology save us? *Computer Fraud and Security*, 2019(7):11–16, 2019. ISSN: 13613723. DOI: 10.1016/S1361-3723(19)30074-0.

[19] Adam Kavon Ghazi-Tehrani and Henry N. Pontell. Phishing Evolves: Analyzing the Enduring Cybercrime. *Victims and Offenders*, 16(3):316–342, 2021. ISSN: 15564991. DOI: 10.1080/15564886.2020.1829224.

[20] C. J. Gokul, Sankalp Pandit, Sukanya Vaddepalli, Harshal Tupsamudre, Vijayanand Banahatti, and Sachin Lodha. Phishy - A serious game to train enterprise users on phishing awareness. In *CHI PLAY 2018 - Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, pages 169–181, 2018. ISBN: 9781450359689. DOI: 10.1145/3270316.3273042.

[21] Tzipora Halevi, James Lewis, and Nasir Memon. A pilot study of cyber security and privacy related behavior and personality traits. *WWW 2013 Companion - Proceedings of the 22nd International Conference on World Wide Web*:737–744, 2013. DOI: 10.1145/2487788.2488034.

[22] Farkhondeh Hassandoust, Harminder Singh, and Jocelyn Williams. The role of contextualization in users' vulnerability to phishing attempts. *Australasian Journal of Information Systems*, 24, 2020. ISSN: 14498618. DOI: 10.3127/AJIS.V24I0.2693.

[23] Joseph M. Hatfield. Social engineering in cybersecurity: The evolution of a concept. *Computers and Security*, 73:102–113, 2018. ISSN: 01674048. DOI: 10.1016/j.cose.2017.10.008.

[24] Markus Jakobsson. The Human Factor in Phishing. *Privacy Security of Consumer Information*, 7:1–19, 2007. URL: http://markus-jakobsson.com/papers/jakobsson-psci07.pdf.

[25] Daniel Jampen, Gürkan Gür, Thomas Sutter, and Bernhard Tellenbach. Don't click: towards an effective anti-phishing training. A comparative literature review. *Human-centric Computing and Information Sciences*, 10(1):1–41, 2020. ISSN: 21921962. DOI: 10.1186/s13673-020-00237-7.

[26] Helen S. Jones, John N. Towse, Nicholas Race, and Timothy Harrison. Email fraud: The search for psychological predictors of susceptibility. *PLoS ONE*, 14(1):1–15, 2019. ISSN: 19326203. DOI: 10.1371/journal.pone.0209684.

[27] Iacovos Kirlappos and M. Angela Sasse. Security education against Phishing: A modest proposal for a Major Rethink. *IEEE Security and Privacy*, 10(2):24–32, 2012. ISSN: 15407993. DOI: 10.1109/MSP.2011.179.

[28] Sabina Kleitman, Marvin K.H. Law, and Judy Kay. It's the deceiver and the receiver: Individual differences in phishing susceptibility and false positives with item profiling. *PLoS ONE*, 13(10), 2018. ISSN: 19326203. DOI: 10.1371/journal.pone.0205089.

[29] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong. Lessons from a real world evaluation of anti-phishing training. *eCrime Researchers Summit*, 2008. DOI: 10.1109/ECRIME.2008.4696970.

[30] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong. Teaching johnny not to fall for phish. *ACM Transactions on Internet Technology*, 10(2), 2010. ISSN: 15335399. DOI: 10.1145/1754393.1754396.

[31] Frieder R. Lang, Dennis John, Oliver Lüdtke, Jürgen Schupp, and Gert G. Wagner. Short assessment of the Big Five: Robust across survey methods except telephone interviewing. *Behavior Research Methods*, 43(2):548–567, 2011. ISSN: 1554351X. DOI: 10.3758/s13428-011-0066-z.

[32] Tian Lin, Daniel E. Capecci, Donovan M. Ellis, Harold A. Rocha, Sandeep Dommaraju, Daniela S. Oliveira, and Natalie C. Ebner. Susceptibility to spear-phishing emails: Effects of internet user demographics and email content. *ACM Transactions on Computer-Human Interaction*, 26(5):32, 2019. ISSN: 15577325. DOI: 10.1145/3336141.

[33] Xin (Robert) Luo, Wei Zhang, Stephen Burd, and Alessandro Seazzu. Investigating phishing victimization with the heuristic–systematic model: a theoretical framework and an exploration. *Computers & Security*, 38:28–38, 2013. ISSN: 0167-4048. DOI: 10.1016/j.cose.2012.12.003.

[34] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*, 15(4):336–355, 2004. ISSN: 10477047. DOI: 10.1287/isre.1040.0032.

[35] Saif M. Mohammad. Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text, 2016. DOI: 10.1016/B978-0-08-100508-8.00009-6.

[36] Gregory D. Moody, Dennis F. Galletta, and Brian Kimball Dunn. Which phish get caught An exploratory study of individuals susceptibility to phishing. *European Journal of Information Systems*, 26(6):564–584, 2017. ISSN: 14769344. DOI: 10.1057/s41303-017-0058-x.

[37] Paula M.W. Musuva, Katherine W. Getao, and Christopher K. Chepken. A new approach to modelling the effects of cognitive processing and threat detection on phishing susceptibility. *Computers in Human Behavior*, 94:154–175, 2019. ISSN: 07475632. DOI: 10.1016/j.chb.2018.12.036.

[38] Adam Oest, Penghui Zhang, Brad Wardman, Eric Nunes, Jakub Burgis, Ali Zand, Kurt Thomas, Adam Doupé, and Gail Joon Ahn. Sunrise to sunset: Analyzing the end-to-end life cycle and effectiveness of phishing attacks at scale. *Proceedings of the 29th USENIX Security Symposium*:361–377, 2020. DOI: 10.5555/3489212.3489233.

[39] Kathryn Parsons, Marcus Butavicius, Paul Delfabbro, and Meredith Lillie. Predicting susceptibility to social influence in phishing emails. *International Journal of Human Computer Studies*, 128(February):17–26, 2019. ISSN: 10959300. DOI: 10.1016/j.ijhcs.2019.02.007.

[40] Kathryn Parsons, Marcus Butavicius, Malcolm Pattinson, Agata McCormac, Dragana Calic, and Cate Jerram. Do users focus on the correct cues to differentiate between phishing and genuine emails? In *ACIS 2015 Proceedings - 26th Australasian Conference on Information Systems*, 2015. ISBN: 9780646953373.

[41] Kathryn Parsons, Agata McCormac, Malcolm Pattinson, Marcus Butavicius, and Cate Jerram. Phishing for the truth: A scenario-based experiment of users' behavioural response to emails. *IFIP Advances in Information and Communication Technology*, 405:366–378, 2013. ISSN: 1868422X. DOI: 10.1007/978-3-642-39218-4_27.

[42] Ed Pearson, Cindy L. Bethel, Andrew F. Jarosz, and Mitchell E. Berman. "To click or not to click is the question": Fraudulent URL identification accuracy in a community sample. In *2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017*, pages 659–664, 2017. ISBN: 9781538616451. DOI: 10.1109/SMC.2017.8122682.

[43] Gordon Pennycook, Ziv Epstein, Mohsen Mosleh, Antonio A. Arechar, Dean Eckles, and David G. Rand. Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855):590–595, 2021. ISSN: 14764687. DOI: 10.1038/s41586-021-03344-2.

[44] Gordon Pennycook, Jonathon McPhetres, Yunhao Zhang, Jackson G. Lu, and David G. Rand. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science*, 31(7):770–780, 2020. ISSN: 14679280. DOI: 10.1177/0956797620939054.

[45] Justin Petelka, Yixin Zou, and Florian Schaub. Put your warning where your link is: Improving and evaluating email phishing warnings. *Conference on Human Factors in Computing Systems*, 2019. DOI: 10.1145/3290605.3300748.

[46] Elissa M. Redmiles, Noel Warford, Amritha Jayanti, Aravind Koneru, Sean Kross, Miraida Morales, Rock Stevens, and Michelle L. Mazurek. A comprehensive quality evaluation of security and privacy advice on the web. In *Proceedings of the 29th USENIX Security Symposium*, pages 89–108, 2020. ISBN: 9781939133175. URL: https://www.usenix.org/conference/usenixsecurity20/presentation/redmiles.

[47] Benjamin Reinheimer, Lukas Aldag, Peter Mayer, Mattia Mossano, Reyhan Duezguen, Bettina Lofthouse, Tatiana von Landesberger, and Melanie Volkamer. An investigation of phishing awareness and education over time: When and how to best remind users. *Proceedings of the 16th Symposium on Usable Privacy and Security, SOUPS 2020*:259–284, 2020. URL: https://www.usenix.org/conference/soups2020/presentation/reinheimer.

[48] Leonard Richardson. Beautiful Soup Documentation. *Media.Readthedocs.Org*:1–72, 2016. URL: https://media.readthedocs.org/pdf/beautiful-soup-4/latest/beautiful-soup-4.pdf.

[49] Ben D. Sawyer and Peter A. Hancock. Hacking the Human: The Prevalence Paradox in Cybersecurity. *Human Factors*, 60(5):597–609, 2018. ISSN: 15478181. DOI: 10.1177/0018720818780472.

[50] Kuldeep Singh, Palvi Aggarwal, Prashanth Rajivan, and Cleotilde Gonzalez. Training to Detect Phishing Emails: Effects of the Frequency of Experienced Phishing Emails. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 63(1):453–457, 2019. ISSN: 2169-5067. DOI: 10.1177/1071181319631355.

[51] Team R Development Core. A Language and Environment for Statistical Computing, Vienna, Austria, 2018. URL: http://www.r-project.org.

[52] René van Bavel, Nuria Rodríguez-Priego, José Vila, and Pam Briggs. Using protection motivation theory in the design of nudges to improve online security behavior. *International Journal of Human Computer Studies*, 123:29–39, 2019. ISSN: 10959300. DOI: 10.1016/j.ijhcs.2018.11.003.

[53] Verizon. 2021 data breach investigations report, 2021.

[54] Pauli Virtanen et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, 2020. ISSN: 15487105. DOI: 10.1038/s41592-019-0686-2.

[55] Arun Vishwanath. Examining the Distinct Antecedents of E-Mail Habits and its Influence on the Outcomes of a Phishing Attack. *Journal of Computer-Mediated Communication*, 20(5):570–584, 2015. ISSN: 10836101. DOI: 10.1111/jcc4.12126.

[56] Arun Vishwanath, Brynne Harrison, and Yu Jie Ng. Suspicion, Cognition, and Automaticity Model of Phishing Susceptibility. *Communication Research*, 45(8):1146–1166, 2018. ISSN: 15523810. DOI: 10.1177/0093650215627483.

[57] Arun Vishwanath, Tejaswini Herath, Rui Chen, Jingguo Wang, and H. Raghav Rao. Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems*, 51(3):576–586, 2011. ISSN: 01679236. DOI: 10.1016/j.dss.2011.03.002.

[58] Melanie Volkamer, Karen Renaud, Benjamin Reinheimer, and Alexandra Kunz. User experiences of TORPEDO: TOoltip-poweRed Phishing Email DetectiOn. *Computers and Security*, 71:100–113, 2017. ISSN: 01674048. DOI: 10.1016/j.cose.2017.02.004.

[59] Melanie Volkamer, Martina Angela Sasse, and Franziska Boehm. Analysing Simulated Phishing Campaigns for Staff:312–328, 2020. DOI: 10.1007/978-3-030-66504-3_19.

[60] Rick Wash. How Experts Detect Phishing Scam Emails. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2), 2020. ISSN: 25730142. DOI: 10.1145/3415231.

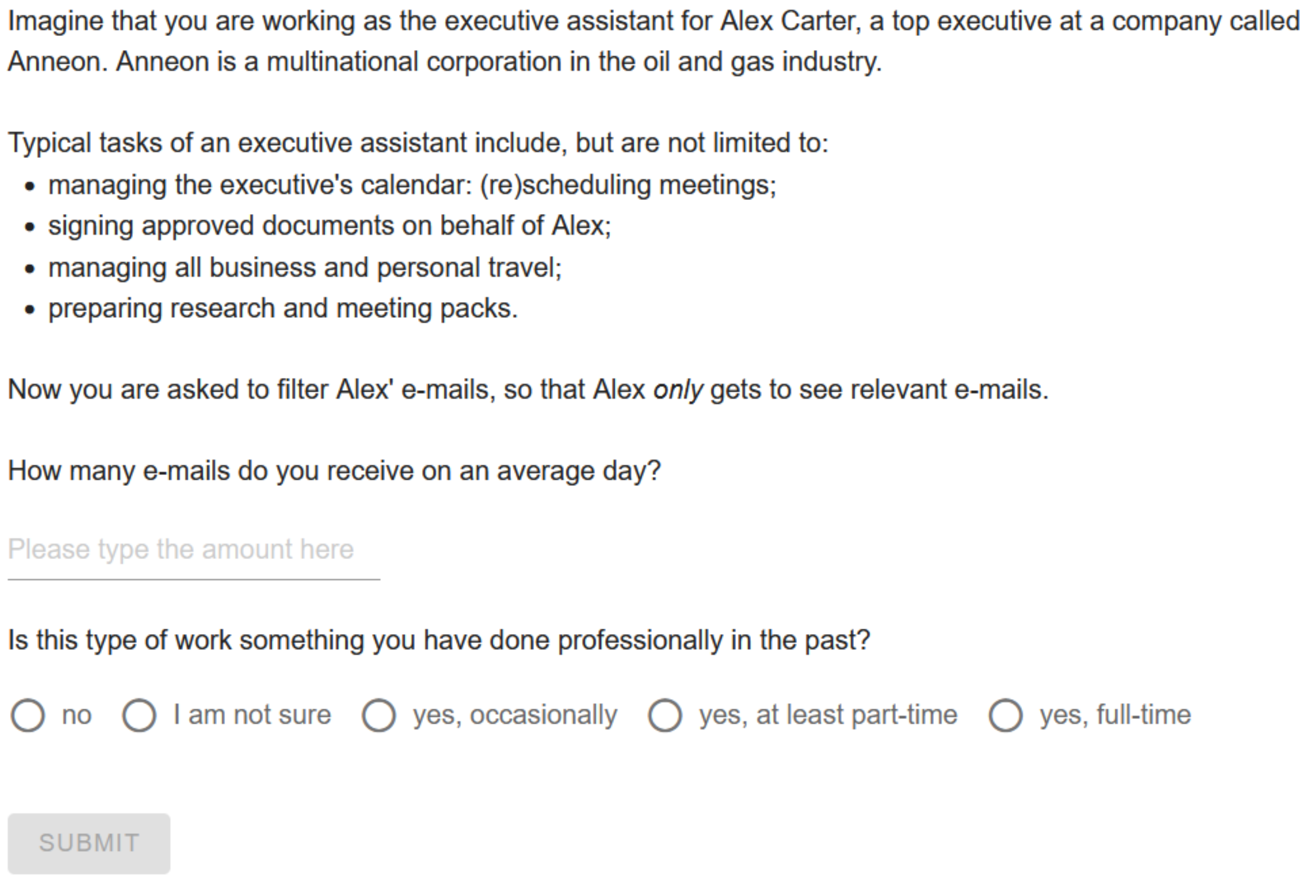
[61] Emma J. Williams and Danielle Polage. How persuasive is phishing email? The role of authentic design, influence and current events in email judgements. *Behaviour and Information Technology*, 38(2):184–197, 2019. ISSN: 13623001. DOI: 10.1080/0144929X.2018.1519599.

[62] Ryan T. Wright, Matthew L. Jensen, Jason Bennett Thatcher, Michael Dinger, and Kent Marett. Influence techniques in phishing attacks: An examination of vulnerability and resistance. *Information Systems Research*, 25(2):385–400, 2014. ISSN: 15265536. DOI: 10.1287/isre.2014.0522.

[63] Ryan T. Wright and Kent Marett. The influence of experiential and dispositional factors in phishing: An empirical investigation of the deceived. *Journal of Management Information Systems*, 27(1):273–303, 2010. ISSN: 07421222. DOI: 10.2753/MIS0742-1222270111.

A Task instructions

Display of e-mail examples in the instructions was adapted according to the participant's randomly assigned display condition. The screenshots below are taken from the Bodyless condition.



Imagine that you are working as the executive assistant for Alex Carter, a top executive at a company called Anneon. Anneon is a multinational corporation in the oil and gas industry.

Typical tasks of an executive assistant include, but are not limited to:

- managing the executive's calendar: (re)scheduling meetings;
- signing approved documents on behalf of Alex;
- managing all business and personal travel;
- preparing research and meeting packs.

Now you are asked to filter Alex' e-mails, so that Alex *only* gets to see relevant e-mails.

How many e-mails do you receive on an average day?

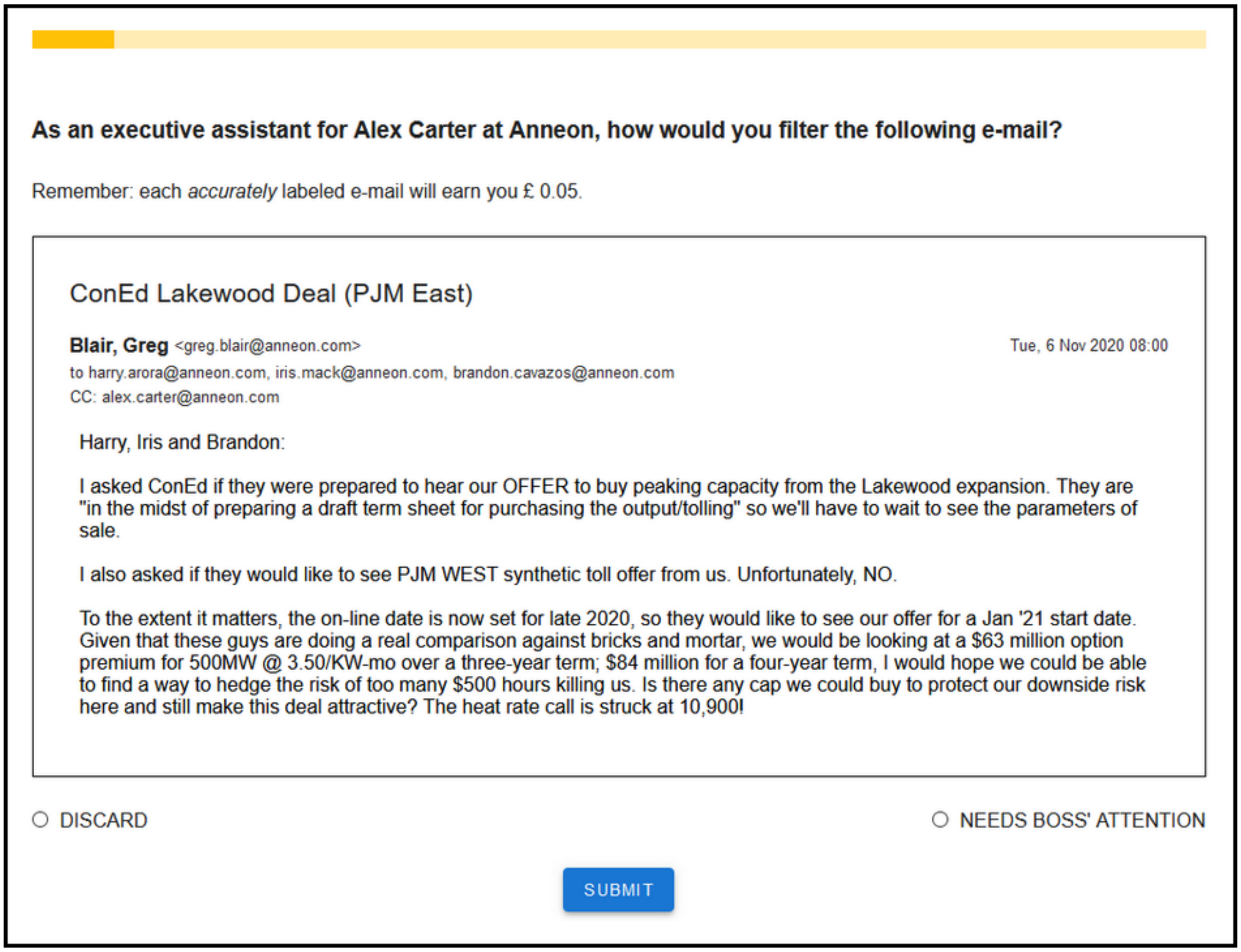
Please type the amount here

Is this type of work something you have done professionally in the past?

○ no ○ I am not sure ○ yes, occasionally ○ yes, at least part-time ○ yes, full-time

You will see one e-mail at a time.
For each e-mail, you need to indicate whether you think it needs your boss' attention or whether to "discard" it.
To do so, you need to **select the applicable label beneath the e-mail** and click "SUBMIT".

Your screen will look like this:



As an executive assistant for Alex Carter at Anneon, how would you filter the following e-mail?

Remember: each *accurately* labeled e-mail will earn you £ 0.05.

ConEd Lakewood Deal (PJM East)

Blair, Greg <greg.blair@anneon.com> Tue, 6 Nov 2020 08:00
to harry.arora@anneon.com, iris.mack@anneon.com, brandon.cavazos@anneon.com
CC: alex.carter@anneon.com

Harry, Iris and Brandon:

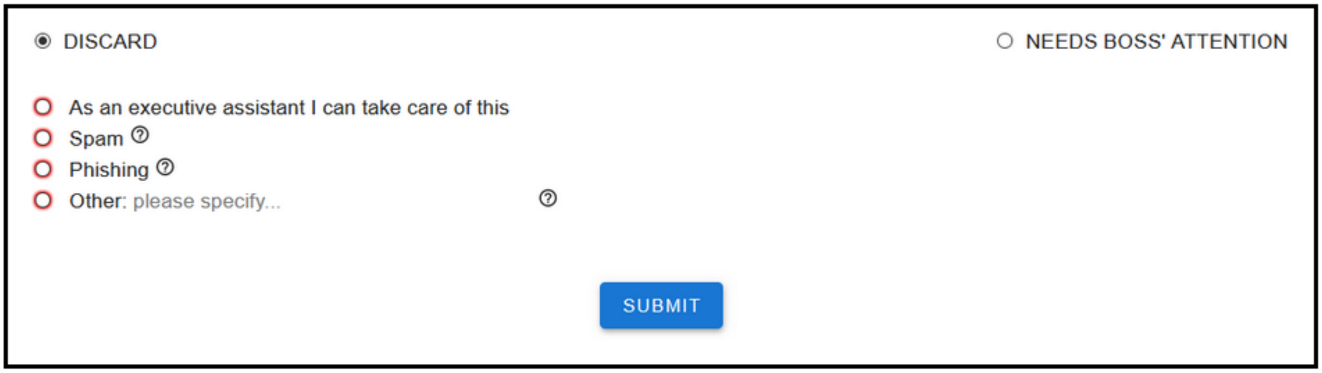
I asked ConEd if they were prepared to hear our OFFER to buy peaking capacity from the Lakewood expansion. They are "in the midst of preparing a draft term sheet for purchasing the output/tolling" so we'll have to wait to see the parameters of sale.

I also asked if they would like to see PJM WEST synthetic toll offer from us. Unfortunately, NO.

To the extent it matters, the on-line date is now set for late 2020, so they would like to see our offer for a Jan '21 start date. Given that these guys are doing a real comparison against bricks and mortar, we would be looking at a $63 million option premium for 500MW @ 3.50/KW-mo over a three-year term; $84 million for a four-year term, I would hope we could be able to find a way to hedge the risk of too many $500 hours killing us. Is there any cap we could buy to protect our downside risk here and still make this deal attractive? The heat rate call is struck at 10,900!

The yellow bar at the top indicates your progress.

If you label an e-mail as “NEEDS BOSS’ ATTENTION”, it will go to your boss Alex Carter.
If you label an e-mail as “DISCARD”, you will be asked to specify the reason:



● DISCARD ○ NEEDS BOSS' ATTENTION

○ As an executive assistant I can take care of this
○ Spam ⓘ
○ Phishing ⓘ
○ Other: please specify... ⓘ

As a top executive, Alex is responsible for making strategic and financial decisions to keep the business running as profitably as possible. These decisions can range from foreign investment deals, finding new business partners, to changing local plant operations.
Next to this, Alex acts as a representative of Anneon who ocassionally needs to present about the company at internal or external events.
Any e-mail that fits the scope of these responsibilities needs to be labeled as "NEEDS BOSS' ATTENTION".

"As an executive assistant, I can take care of this" indicates that the e-mail falls within the scope of the typical tasks of an executive assistant as outlined before.

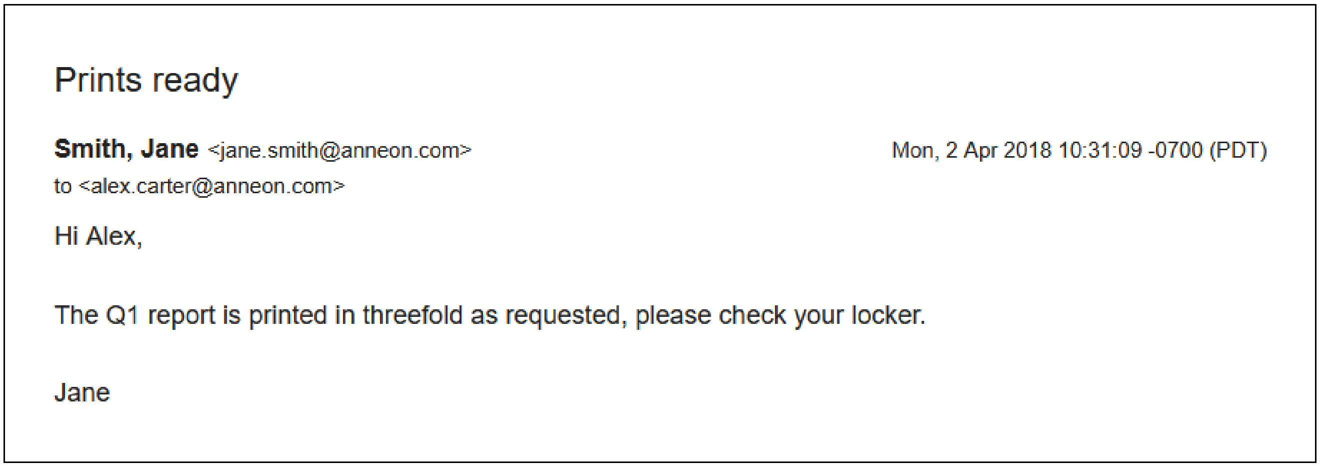
Spam e-mails are messages that the recipient did not ask for and are sent to many people at once, often with commercial aims.

Phishing e-mails are fraudulent communications that appear to come from a reputable source. Their objective is to obtain sensitive data from their victim.

Select "Other" if you believe none of the other categories apply, with a brief comment with your reasoning.

After each submit, you will be presented with the next e-mail.
You cannot revise e-mails that you have already labeled.
The yellow bar at the top indicates your progress.

This example e-mail would be labeled as "DISCARD" with the reason “As an executive assistant, I can take care of this”:



Prints ready

Smith, Jane <jane.smith@anneon.com> Mon, 2 Apr 2018 10:31:09 -0700 (PDT)
to <alex.carter@anneon.com>

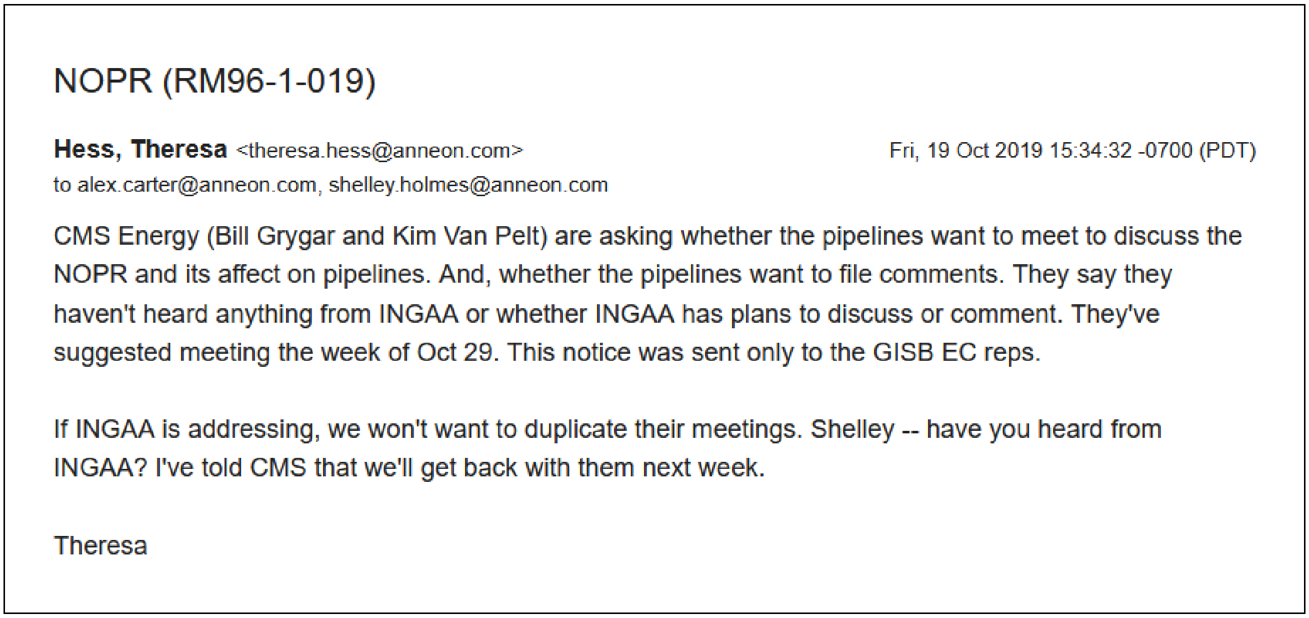
Hi Alex,

The Q1 report is printed in threefold as requested, please check your locker.

Jane

Printed reports for Alex can be managed by the executive assistant.

This example e-mail would be labeled as “NEEDS BOSS' ATTENTION”:



NOPR (RM96-1-019)

Hess, Theresa <theresa.hess@anneon.com> Fri, 19 Oct 2019 15:34:32 -0700 (PDT)
to alex.carter@anneon.com, shelley.holmes@anneon.com

CMS Energy (Bill Grygar and Kim Van Pelt) are asking whether the pipelines want to meet to discuss the NOPR and its affect on pipelines. And, whether the pipelines want to file comments. They say they haven't heard anything from INGAA or whether INGAA has plans to discuss or comment. They've suggested meeting the week of Oct 29. This notice was sent only to the GISB EC reps.

If INGAA is addressing, we won't want to duplicate their meetings. Shelley -- have you heard from INGAA? I've told CMS that we'll get back with them next week.

Theresa

NOPR stands for Notice of Proposed Rulemaking. Alex needs to be informed about NOPRs, as they legally affect Anneon's business operations.