Checking, nudging or scoring?
Evaluating e-mail user security tools

7 August 2023 – SOUPS

Sarah Zheng, Ingolf Becker
sarah.zheng.16@ucl.ac.uk
Scope for usable e-mail security tools

➢ Even automated detection tools with high accuracies will still let some phishing e-mails through to users, e.g. Oest et al. (2020)

➢ Most methods to help users detect phishing e-mails rely on training (Franz et al., 2021)

➢ Phishing awareness training may not be effective enough (Hillman et al., 2023; Lain et al., 2022; Zheng & Becker, 2022; Reinheimer et al., 2020), because they do not provide guidance during critical decision-making moments

➢ Underexplored use of nudges (Franz et al., 2022); URL checking tools embedded in inboxes showed promising results (Petelka et al., 2019; Volkamer et al., 2017)
How usable are nudging and checking tools for e-mail security?

1. “Check” button
   - **Objective**: assist user when in doubt
   - Shows parsed sender details, links and past correspondence – applies to legitimate e-mails, too
   - Advice in case of mismatching details

2. Nudge I: Collegiate phishing report
   - **Objective**: raise phishing awareness
   - “This e-mail was reported as suspicious today by one of our colleagues”
   - Shows phishing e-mail example with suspicious signals annotated

3. Nudge 2: Suspicion score
   - **Objective**: raise phishing awareness
   - “Are you sure you can trust this e-mail?”
   - Shows e-mail suspicion score + recommends user actions
Study design

Qualitative think-aloud task
- Reflective thematic analysis on users’ reasoning about e-mails in simulated Outlook web inbox without and then with each tool
- Questions before & after main task; rated which design they found most useful

Implementation-focused formative evaluation
- How do the tools affect users’ e-mail processing behaviour?
- Iterative design: tools were updated after 5 users gave same feedback

Professional e-mail users (N=27)
- UCL staff; mean age: 33.3, 48% male, 18 recalled cybersecurity training; 19 studied technical subject
- Sessions ran consecutively
- Ethics approval from UCL department
Simulated Outlook web inbox

Try it yourself: https://mock-inbox.web.app/ - 33 legitimate & 6 phishing e-mails, academic context
## Evaluation of check button

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Check button</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ((N = 8))</td>
<td>The majority of users were unaware of the button until nudged towards it (after 2–3 minutes); users did not explore all sub menu items</td>
</tr>
<tr>
<td>2 ((N = 7))</td>
<td>Users remained unaware of the button until the researcher pointed it out, but also often did not see the benefit of the provided information.</td>
</tr>
<tr>
<td>3 ((N = 5))</td>
<td>Users remained unaware of the button until the researcher pointed it out; the ‘past correspondence’ element was deemed useful</td>
</tr>
<tr>
<td>4 ((N = 7))</td>
<td>Most users who noticed and started using the button found it very useful</td>
</tr>
</tbody>
</table>
Evaluation of collegiate phishing report nudge

**Iteration 1: Collegiate phishing report**

1. \(N = 8\) (Users tended not to click on the warning banner or got confused about which e-mail the warning is referring to)

2. \(N = 7\) (Users did not like pop-up windows and often felt urged to close it right away)

3. \(N = 5\) (Design did not change)

4. \(N = 7\) (More users skimmed over the warning content, some users found this and the suspicion score generally useful as they alerted them of suspicious e-mails)
Evaluation of suspicion score nudge

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Nudge 2: Suspicion score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (N = 8)</td>
<td>Users did not read all provided information, but found the orange colour positively alerting and useful</td>
</tr>
<tr>
<td>2 (N = 7)</td>
<td>(design did not change)</td>
</tr>
<tr>
<td>3 (N = 5)</td>
<td>(design did not change)</td>
</tr>
<tr>
<td>4 (N = 7)</td>
<td>Users did not read all provided information, but found the orange colour positively alerting and useful; subtle text formatting edits did not lead to significantly more users applying the recommended actions</td>
</tr>
</tbody>
</table>
Overall usability drivers

“Most useful”: suspicion score nudge (N=9), past correspondence check (N=7), both (N=4)

1. **Usability of security information**: technical security-related information was perceived as too much, difficult to understand and/or often ignored; users did adopt intuitive cues of legitimacy, e.g. past correspondence check

2. **Productivity vs. security**: users did not engage with tools that seem irrelevant to get the primary task done

3. **Concerns on false positives**: suspicion score nudge let users actively think about e-mail legitimacy; may not fully prevent wrong conclusions

4. **Ignorance toward security features**: when users find the tool’s functionality unclear or unnecessary, they did not explore it
Conclusion

Guidelines for future usable e-mail security tool development

Embedded e-mail user security tools can be effective if they:

1. **Highlight cues of desired (i.e. legitimate) communication** instead of what is undesired (e.g. phishing)
2. **Enhance users' existing behaviour** instead of technical knowledge
   - To avoid warning fatigue, provide contextually relevant information only when helpful
3. **Do not** interfere with users' productivity (i.e., primary task)
Checking, nudging or scoring? Evaluating e-mail user security tools

7 August 2023 – SOUPS

Sarah Zheng, Ingolf Becker
sarah.zheng.16@ucl.ac.uk