Trust me and Click! A Pilot Study of Cognitive Walkthrough for Phishing Emails

Pierre-Emmanuel Arduin, Kathia Marcai de Oliveira, Christophe Kolski

Presented at: HSI 2021, 14th International Conference on Human System Interaction (July 08-10), IEEE, Gdansk, Poland, pp. 1-6, 2021

Trust me and Click! A Pilot Study of Cognitive Walkthrough for Phishing Emails

Pierre-Emmanuel Arduin
Université Paris-Dauphine – PSL
DRM – UMR CNRS 7088
Paris, France
pierre-emmanuel.arduin@dauphine.psl.eu

Káthia Marçal de Oliveira
Univ. Polytechnique Hauts-de-France
LAMIH – UMR CNRS 8201
Valenciennes, France
kathia.oliveira@uphf.fr

Christophe Kolski
Univ. Polytechnique Hauts-de-France
LAMIH – UMR CNRS 8201
Valenciennes, France
christophe.kolski@uphf.fr

***Abstract*—The growing importance of new markets on the Internet in the 1990s led to the development of new guidelines and interface design principles ensuring online consumer trust. These guidelines have been rapidly spoofed by attackers who integrated interface design principles maliciously to deceive users with phishing emails. In this paper we propose to adapt an inspection method, Cognitive Walkthrough, to understand how users walk through the processing of a phishing email. 23 experienced evaluators used the proposed method, Cognitive Walkthrough for phishing emails, in a pilot study and gave a feedback on it. The method allows to analyze the areas of a phishing email eliciting users to trust, to act, and to distrust. These areas are then commented allowing to understand the interface design principles exploited maliciously by attackers when they are saying to users: Trust me and click!**

***Index Terms*—Cognitive Walkthrough, Phishing, Online Deception, Usability Inspection.**

I. Introduction

The "businessization" of the Internet [18] in the early 1990s led to new markets and implied new challenges for Human System Interactions. Mainframe computers interacting with experts in the 1970s became personal computers interacting with employees in the 1980s. These became connected to the Internet in the 1990s and interacting with citizens of the world through to the World Wide Web. Researchers developed then models and guidelines ensuring online consumer trust [4], [13]. Originally developed for legitimate companies purposes, these guidelines have been rapidly spoofed by phishers to deceive individuals [12].

Past studies highlighted that users from different socioeconomic groups or different countries have different privacy-related experiences and perceptions [31]. An experiment made by Butavicius et al. [8] concluded that firewalls and anti-viruses are more trusted than spam filters and social media privacy settings. In 2021, more than 300 billion emails are sent everyday in the world [29] and when facing an email, decision-making relies on simple cues making most phishing emails being peripherally processed [34]. The pilot study presented in this paper focuses on the first step of attacks beginning with phishing emails as an entry point. We argue that emails may be regarded as Human System Interactions phenomena, where user actions may be requested by the system.

Considering an increasing malicious adoption of interface design principles by attackers, we propose to adapt Cognitive Walkthrough [20], an inspection method historically developed for usability inspection, to the case of phishing emails. An originality of this work is to focus on the evaluation of a kind of maliciousness quality, *i.e.* the extend to which the attacker has integrated interface design principles in a phishing email to deceive the user.

In the second section of this paper, we present the background theory and assumptions: usability inspection and Cognitive Walkthrough, phishing and online deception. In the third section, we present how Cognitive Walkthrough can be adapted to phishing emails. In the fourth section, we present a pilot study that involved 23 participants: the experiment procedure and the participants, the preliminary results, and a discussion on the limits and ethical implications of this research. The overall purpose of this paper is to raise awareness on how usability principles may be misused by attackers targeting legitimate users, *i.e.* maliciousness quality. Our proposition leads to understand the interface design principles exploited by attackers when they are saying to users: Trust me and click!

II. Background theory and assumptions

In this section, we first present usability inspection and Cognitive Walkthrough. Second, we draw from the literature to present phishing and how users may be deceived through online deception.

A. Usability inspection and Cognitive Walkthrough

Past research highlighted early the idea that software evaluation was not well accepted within organizations [27], often considered as too expensive, time-consuming, and difficult to implement. In the mid-1990s, inspection methods appeared as a potential way to persuade professionals to adopt evaluation: they are convenient, affordable, and manageable. Traditionally, these methods intervene early in a user interface design to identify, qualify and quantify usability problems [33]. In this paper, we propose a paradigm switch by adapting and studying how a method such as Cognitive Walkthrough (CW) could be applied *a posteriori* to inspect maliciousness quality, *i.e.* how usability principles have been maliciously used by attackers to deceive users.

CW is an inspection method inspired by Norman's Action Theory [28] with the idea of learning through exploration, allowing the detection of usability defects and the estimation of their degree of seriousness. It relies on: (i) the ease of performing a task with a minimum of knowledge of the system, and (ii) the ease of learning through exploration of the user interface. During a preparation phase, the evaluator specifies series of tasks to be assessed and each task is broken down into action sequences. Then, during an evaluation phase, the evaluator inspects each action by walking through the interface; he/she answers to questions, identifies problems, and reports them in problem description sheets.

Actually, even if when receiving an email a user do not need to walk through an interface, we argue that emails in general and phishing emails in particular may be regarded as Human System Interactions phenomena: user actions are requested such as clicking on a link or downloading an attachment for example. This constitutes an originality of the study presented in this paper. Indeed, the scientific community proposed several variants and evolutions of CW [21] and, when looking at a variation such as CW for the Web, we discovered a method for detecting and fixing errors occuring when browsing a website [6]. First, the user parses the page into a range of subregions and generates a brief description of each subregion (attention phase). Second, he/she generates descriptions of all graphic widgets in the target subregion and acts on the closest to his or her goals [19].

Dhamija et al. [12] performed a CW on 200 sample attacks. The 22 participants browsed legitimate and phishing websites with no information on their trustworthiness, had to say if the site was legitimate or not, state their confidence in their evaluation, and explain their reasoning. The authors highlighted three main dimensions in attackers' strategies: (i) lack of knowledge, (ii) visual deception, and (iii) lack of attention. This approach is rather different from the one of the work presented in this paper: not only we focus on emails as user interfaces requiring user actions, but also the participants are not acting as common users receiving emails but they have the role of evaluators, as if they were inspecting usability. Collected data will lead to understand usability problems and we argue that such data could also lead to a better understanding of the interface design principles used maliciously by attackers to deceive users.

B. Phishing and online deception

Matheson and Zanna [22] already proposed in the late 1980s a computer-mediated persuasion study. Participants had to complete decision-making problems in two situations: (i) face-to-face, reading persuasive communication printed on paper, (ii) computer-mediated, reading persuasive communication on a screen. The results showed then an increased feeling of privacy in the computer-mediated situation compared to the face-to-face situation.

Phishing, *i.e.* the practice of directing users to fraudulent websites [12], is well discussed in the literature through psychometric studies [8], context studies [3], or training studies [32]. Yu et al. [36] presented recently that slow mouse movements are good predictors of phishing threat detection. Nevertheless, the literature seems to lack of an adapted usability inspection method, such as Cognitive Walkthrough (CW, see Section II-A), to evaluate a kind of maliciousness quality, *i.e.* the extend to which the attacker has integrated interface design principles in a phishing email to deceive the user.

Indeed, online interactions as well as a constant multitasking led individuals to rely more on mental shortcuts such as cognitive biases [23]. Such biases may be used as online weapons of influence [25] when a user interface is designed to integrate persuasion principles [9] and leads to effective phishing [14]. Users often suffer from engaging in an effortful and careful way of thinking [15], particularly when interactions are online [16]. In the early 2000s authors as Cialdini et al. [10] warned on the potentiality of persuasion principles used maliciously on the Internet, asking for what they called "ethics of influence". Authors such as Xiao et al. [35] go deeper recently with a call to understand comprehensively online persuasion mechanisms and processes. As the reader may guess, this is not what we are going to propose in this paper, focusing on the evaluation of interface design principles integrated maliciously in a phishing email to deceive users.

If Zuckerman et al. [37] proposed in the early 1980s studies to detect deception through videotape experiments, recent research considers phishing susceptibility as a likelihood of being phished [17] and opposes it to phishing resistance [8]. Some authors also consider trust and distrust [2], even if others such as Moody et al. [24] argue that neither the disposition to trust nor to distrust influence phishing susceptibility prediction. Authors as Nicho et al. [26] proposed to summarize the literature on variables affecting phishing susceptibility and we note that very few concern is given to usability inspection.

We propose in this paper to go further in studying how attackers adopted interface design principles to deceive users. By adapting an inspection method, CW, to phishing emails, we are studying how users walk through the analysis of the content, how they decide to trust, to follow a call-to-action, and to distrust a phishing email.

III. Cognitive Walkthrough for phishing emails

We argue that phishing emails are entry points of phishing attacks and that the presence of user-action calls within these emails led us to consider them as Human System Interactions phenomena.

The considered task of the walkthrough is the processing of an email by a user when he/she receives it. Such a task is broken down into actions such as the identification of trust-eliciting cues and reasoning explanations, the identification of calls to potentially risky actions and consequent damages, and the identification of distrust-eliciting design cues and reasoning explanations.

First, the evaluator identifies in the email a set of **trust-eliciting areas** for the user: T. The factors eliciting trust for each area $t \in T$ are then noted by the evaluator in open text comments fields. Second, the evaluator identifies in the email a set of **areas calling the user to act** and potentially posing a

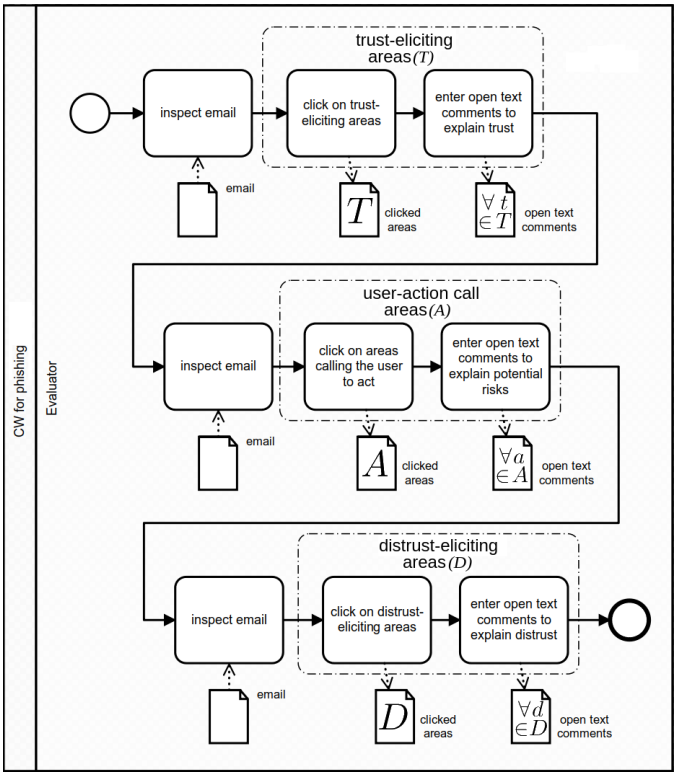

Fig. 1. Actions completed by the evaluator during the inspection of a phishing email.

risk (click on a link, open an attachment, etc.): A. The potential risks for each area $a \in A$ are then noted by the evaluator in open text comments fields. Third, the evaluator identifies in the email a set of **distrust-eliciting areas** for the user: D. The factors eliciting distrust for each area $d \in D$ are then noted by the evaluator in open text comments fields.

The areas T, A, and D selected by several evaluators will allow to produce heatmaps [7] showing the most trust-eliciting, calling a user action, and distrust-eliciting areas in the email, whereas an analysis of the open text comments will allow to highlight the most occurring explanations on reasons eliciting trust, on possible risks following a user action, and on reasons eliciting distrust. Some of the areas calling the user to act (A) may be malicious and lead him/her to constitute an insider threat [1]

Indeed, we consider that a problem occurs with phishing emails when an area calling the user to act is also eliciting his/her trust. This is the case when the intersection between the sets A (the areas calling the user to act) and T (the trust-eliciting areas) is not empty: $A \cap T \neq \emptyset$. The areas eliciting user distrust provide also useful comparative data. Fig. 1 presents the actions completed by the evaluator during an inspection with CW for phishing emails.

IV. Pilot Study

In this section, we first present the experiment procedure and the 23 evaluators. Second, we discuss the preliminary results and third, the limits and ethical implications of this research.

A. Experiment procedure and participants

This pilot study has been conducted with 23 evaluators (11 male, 12 female, 15 PhD, 6 MSc, 2 BSc). In the following we refer to these evaluators involved in the study as *the participants*. The average age of the participants is 42.1 years. 30.4% of the participants reported having inspected ten or more user interfaces (52.2% between three and ten, 47.8% none). Thus participants show different levels of expertise in terms of inspection. In terms of phishing, they were 52.2% to recognize having already been a victim of phishing (39.1% never). 78.3% declared themselves as Computer Scientists and 21.7% as Managers. Data collection was managed with the online reaction time experiment software Qualtrics [5].

The experiment procedure was as follows: a brief text informed the participants about the purpose and the context of the study: they were not users receiving phishing emails but evaluators acting as if they were inspecting usability. Participants were informed that a variation of CW for phishing emails will be used and they had then to give their informed consent to take part.

In this pilot study, two emails were sequentially shown to the participants with no information about their trustfulness. They inspected each email one at a time and faced, for each, three pages with questions structured following the method presented in Section III and by completing the actions presented in Fig. 1. Then, the participants were asked questions about their impressions on the inspection procedure and their profile (age, gender, academic level, area of expertise, number of usability inspections already conducted).

At the end of the experiment, the research team manually determined for each email the set $A \cap T$ showing the existence of areas calling the user to act also trusted by him/her.

B. Preliminary results of the pilot study

Through this experiment, we collected two categories of data: (i) clicked areas in the emails and (ii) open text comments. Participants had no time constraint and the average response time was 26.06 minutes (min. = 00:04:35, max. = 01:23:44). The participant with the higher response time was a cybersecurity expert who admitted having deeply analyzed the emails. The second higher response time was 01:06:20 (no explanation given).

1) Heatmaps of the clicked areas: For the 23 participants and the 2 emails, 3 images with clicked areas were collected (trust-eliciting, user-action-call, distrust-eliciting), namely: $23 \times 2 \times 3 = 138$ images of emails with clicked areas. These 138 images were aggregated by the Qualtrics software to $2 \times 3 = 6$ heatmaps: 2 trust-eliciting, 2 user-action-call, and 2 distrust-eliciting areas. Fig. 2 presents the heatmaps showing the trust-eliciting areas (T), the areas calling a user action (A), and the distrust-eliciting areas (D) for the two considered emails.

The participants identified an average of 3.6 areas as trust-eliciting (min. = 0, max. = 10), 2.3 areas as calling a user action (min. = 1, max. = 6) and 3.3 areas as distrust-eliciting (min. = 1, max. = 10), by taking the two emails together. 166

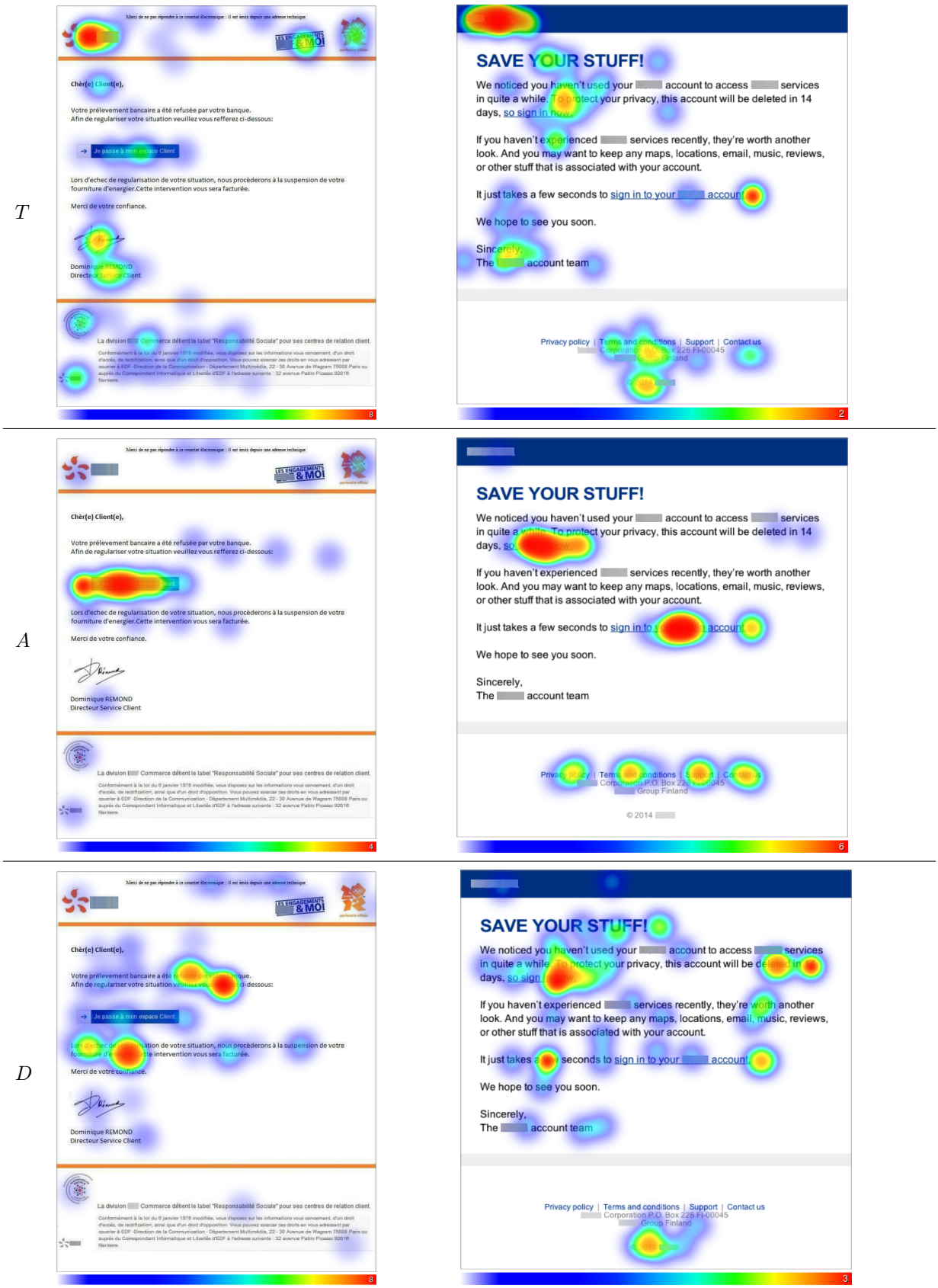

Fig. 2. Trust-eliciting (T), calling a user action (A), and distrust-eliciting (D) areas for the two phishing emails considered in this pilot study ($N = 23$). The trademarks were hidden after the experiment.

areas were identified as trust-eliciting, 105 as calling a user action, and 154 as distrust-eliciting by the 23 participants for the two emails.

In this pilot study, the average number as well as the total number of trust-eliciting areas are higher than for distrust-eliciting areas, even if the participants were aware of the nature of the inspection: CW for phishing emails. Indeed, distrust-eliciting areas, as shown in Fig. 2.D, highlight a focus of participants on the body of the text when seeking for distrust-eliciting cues, but are less numerous than trust-eliciting areas. These results also show that the intersections between the sets A and T for the two emails are not empty (see Figs. 2.A and 2.T), meaning that the attackers have successfully integrated interface design principles to call a user action while eliciting him/her to trust such a call.

2) Analysis of the open text comments: As presented in Sections III and IV-A, the participants should explain why they selected these areas. In this pilot study, up to 10 areas may be clicked by the participants and thus up to 10 open text fields (one explanation per click) for each question may be completed or left empty. On a potential total of $23 \times 2 \times 3 \times 10 = 1,380$ responses, we collected actually 351 responses corresponding to the areas clicked by the participants. These 351 responses were manually tagged by the research team and 303 tags were affected, some responses having no tag: *e.g.* "*it leaves me in doubt...*" or "*see previous answer*". On these 303 tags, 131 were assigned for trust-eliciting explanations, 58 for potential risks of user actions, and 114 for distrust-eliciting explanations.

Trust-eliciting areas are explained by the presence of a usual and good-looking logo (31), the presence of a signature or human contact (16), a suitable formatting (14) and phrasing (12), as well as a verifiable address (12). The possible risks of a user action identified by the participants are a hidden redirection (23) and personal data theft (20). Finally, distrust-eliciting areas are explained by spelling mistakes (24), an irrelevant content (17), a threatening message (15), the presence of a suspicious link (11), as well as a poor phrasing (10), a generic message (9) or a poor formatting (8).

C. Limits and ethical implications of this research

After the inspection, 78.3% of the participants described the procedure as "simple to realize". We noted that 43.5% would not have wanted to be more guided, whereas 39.1% would have preferred. Nevertheless 73.9% agreed that being more guided would have influenced their responses. We noted a 65.2% agreement that the proposed inspection method allowed them to transcribe how a user processes a phishing email and a 39.1% indifference with the feeling of missing something (34.8% disagree, 26.1% agree).

The general comments of the 23 evaluators involved in this pilot study showed an interest for the proposed adaptation of CW for phishing emails. Some of them pointed out limits such as the context of the experiment: "*I would have liked to receive both emails on my address to experience it just as in real life*". Others pointed out the technical constraints of the tool: "*I would have preferred to select areas and not just click on them*", "*not seeing the URLs associated with the links does not really allow to analyze the emails*" . Indicating how the user contextualizes the email to his/her topicality is also presented as a lack for some participants: "*More than specific elements, the links between these elements usually help me to know if the email is fraudulent*", "*I would check if the signatory actually works for the company sending me the email*" . As the reader may guess, the research team is now working on how to improve the method from these comments on the pilot study presented in this paper.

The ethical implications of this research have to be considered. Indeed, the risk of a malicious use of our results or the proposed inspection method should not be neglected. Attackers may use this research as a way to develop a "how to" guide and the risk of dual use [30] is important. Phishing emails are particularly interesting because their efficiency increases not only with the quality of their maliciouness, but also when organizational countermeasures show their limits in mitigating sloppiness and ignorance [11]. Consequently, identifying the interface design principles used maliciously by attackers would be a way to understand how users walk through these cues in particular and the processing of an email in general.

The importance of awareness and formation remains then crucial as a way to improve the processing of emails. We argue that education, higher education, and management could be effective vehicles to raise awareness among users on how they process emails. Such awareness will lead users to better understand how they may be deceived by an attacker through the system. A simple email may appear so informative, so passive, and consequently so harmless. Nevertheless, we have to be aware and make users aware that HTML code and hypertext browsing allow attackers not only to deceive us with a simple email, but also to call and to lead us to act.

V. Conclusion

In this paper we propose a new procedure relying on CW to inspect the maliciousness quality of phishing emails, *i.e.* the extend to which the attacker has integrated interface design principles to deceive the user.

With the presented pilot study involving 23 evaluators, we highlighted trust-eliciting areas, areas calling for a potentially risky user action, and distrust-eliciting areas. We observed that some of the trust-eliciting areas were also calling for a potentially risky user action, showing maliciousness quality. We raised a call for further investigation notably on links between the identified areas, the profile of participants using CW for phishing emails, and the final trust-and-click user action. We then analyzed the open-text explanations filled by the participants to highlight key trends on: why users trust? what do they risk? and why do they distrust? We finally proposed to open and keep open controversy as the considered research material and results may be maliciously used as a "how to" guide by attackers.

The idea of a "design for successful guessing" proposed by Lewis et al. [20] was to facilitate the problem solving

mechanisms with the design of an interface easy to learn by exploration. With this pilot study, we focused on the first step of attacks beginning with a phishing email. This work is currently continuing, as the research team thoroughly analyzes the data per participant to highlight trends on explanations and reasons eliciting trust, user actions and distrust. Moreover, analyzing overlapped areas and quantitative measures as well as crossing data from the open text comments and the heatmaps are other research directions. Indeed, like any other cognitive process, guessing may be deceived by attackers who integrated maliciously interface design principles, as if they were saying to users: Trust me and click!

ACKNOWLEDGMENT

The authors would like to thank the 23 evaluators who performed this pilot study and provided valuable insight into this work.

REFERENCES

[1] P.-E. Arduin, *Insider Threats*. John Wiley & Sons, 2018.

[2] P.-E. Arduin, B. Rajaonah, and K. M. de Oliveira, "Trusting security when sharing knowledge?" in *Knowledge, People, and Digital Transformation*, F. Matos, V. Vairinhos, I. Salavisa, L. Edvinsson, and M. Massaro, Eds. Springer, 2020, pp. 163–181.

[3] E. Ayaburi and F. K. Andoh-Baidoo, "Understanding phishing susceptibility: An integrated model of cue-utilization and habits," in *Proceedings of the 2019 International Conference on Information Systems*, 2019.

[4] S. Ba and P. A. Pavlou, "Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior," *MIS quarterly*, vol. 26, no. 3, pp. 243–268, 2002.

[5] J. S. Barnhoorn, E. Haasnoot, B. R. Bocanegra, and H. van Steenbergen, "Qrtengine: An easy solution for running online reaction time experiments using qualtrics," *Behavior Research Methods*, vol. 47, no. 4, pp. 918–929, 2015.

[6] M. H. Blackmon, P. G. Polson, M. Kitajima, and C. Lewis, "Cognitive walkthrough for the web," in *Proceedings of the SIGCHI conference on human factors in computing systems, ACM*, 2002, pp. 463–470.

[7] A. Bojko, "Informative or misleading? heatmaps deconstructed," in *Human-Computer Interaction. New Trends*, J. A. Jacko, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 30–39.

[8] M. Butavicius, K. Parsons, M. Lillie, A. McCormac, M. Pattinson, and D. Calic, "When believing in technology leads to poor cyber security: Development of a trust in technical controls scale," *Computers & Security*, vol. 98, p. 102020, 2020.

[9] R. B. Cialdini, *Influence: The psychology of persuasion*. New York: Collins, 2007.

[10] R. B. Cialdini, B. J. Sagarin, and W. E. Rice, "Training in ethical influence," in *LEA's organization and management series. Social influences on ethical behavior in organizations*, J. Darley, D. M. Messick, and T. R. Tyler, Eds. Lawrence Erlbaum Associates Publishers, 2001, pp. 137–153.

[11] R. E. Crossler, A. C. Johnston, P. B. Lowry, Q. Hu, M. Warkentin, and R. Baskerville, "Future directions for behavioral information security research," *Computers & Security*, vol. 32, pp. 90–101, 2013.

[12] R. Dhamija, J. D. Tygar, and M. Hearst, "Why phishing works," in *Proceedings of the SIGCHI conference on Human Factors in computing systems, ACM*, 2006, pp. 581–590.

[13] F. N. Egger *et al.*, "Affective design of e-commerce user interfaces: How to maximise perceived trustworthiness," in *Proceedings of the international conference on affective human factors design*, 2001, pp. 317–324.

[14] A. Ferreira and G. Lenzini, "An analysis of social engineering principles in effective phishing," in *2015 Workshop on Socio-Technical Aspects in Security and Trust*. IEEE, 2015, pp. 9–16.

[15] S. T. Fiske and S. E. Taylor, *Social cognition: From brains to culture*. Sage, 2013.

[16] R. E. Guadagno, B. M. Okdie, and N. L. Muscanell, "Have we all just become "robo-sapiens"? reflections on social influence processes in the internet age," *Psychological inquiry*, vol. 24, no. 4, pp. 301–309, 2013.

[17] K. W. Hong, C. M. Kelley, R. Tembe, E. Murphy-Hill, and C. B. Mayhorn, "Keeping up with the joneses: Assessing phishing susceptibility in an email task," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 57, no. 1, 2013, pp. 1012–1016.

[18] M. Imase, "As an element of the regional innovation cluster, the citizen/non-profit sector fulfills the "seedbed function" of the new industry," in *Diversity, Innovation and Clusters*, I. Bernhard, U. Gråsjö, and C. Karlsson, Eds. Edward Elgar Publishing, 2020, pp. 186–206.

[19] M. Kitajima, M. H. Blackmon, and P. G. Polson, "A comprehension-based model of web navigation and its application to web usability analysis," in *People and computers XIV – Usability or else!* Springer, 2000, pp. 357–373.

[20] C. Lewis, P. G. Polson, C. Wharton, and J. Rieman, "Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces," in *Proceedings of the SIGCHI conference on Human factors in computing systems, ACM*, 1990, pp. 235–242.

[21] T. Mahatody, M. Sagar, and C. Kolski, "State of the art on the cognitive walkthrough method, its variants and evolutions," *International Journal of Human–Computer Interaction*, vol. 26, no. 8, pp. 741–785, 2010.

[22] K. Matheson and M. P. Zanna, "Persuasion as a function of self-awareness in computer-mediated communication." *Social Behaviour*, vol. 4, pp. 99–111, 1989.

[23] J. McAlaney and V. Benson, "Cybersecurity as a social phenomenon," in *Cyber Influence and Cognitive Threats*, V. Benson and J. Mcalaney, Eds. Elsevier, 2020, pp. 1–8.

[24] G. D. Moody, D. F. Galletta, and B. K. Dunn, "Which phish get caught? an exploratory study of individuals' susceptibility to phishing," *European Journal of Information Systems*, vol. 26, no. 6, pp. 564–584, 2017.

[25] N. L. Muscanell, R. E. Guadagno, and S. Murphy, "Weapons of influence misused: A social influence analysis of why people fall prey to internet scams," *Social and Personality Psychology Compass*, vol. 8, no. 7, pp. 388–396, 2014.

[26] M. Nicho, H. Fakhry, and U. Egbue, "Evaluating user vulnerabilities vs phisher skills in spear phishing." *IADIS International Journal on Computer Science & Information Systems*, vol. 13, no. 2, pp. 93–108, 2018.

[27] J. Nielsen, "Guerrilla HCI: Using discount usability engineering to penetrate the intimidation barrier," Tech. Rep., 1994.

[28] D. Norman, "Cognitive engineering," in *User Centred System Design: New Perspectives on Human Computer Interaction*. Erlbaum, Hillsdale, NJ, 1986, pp. 31–61.

[29] Radicati-Group, "Email statistics report, 2020-2024," Radicati Group Inc. February, Tech. Rep., 2020.

[30] J. Rath, M. Ischi, and D. Perkins, "Evolution of different dual-use concepts in international and national law and its implications on research ethics and governance," *Science and engineering ethics*, vol. 20, no. 3, pp. 769–790, 2014.

[31] M. Saleh, M. Khamis, and C. Sturm, "What about my privacy, habibi?" in *IFIP Conference on Human-Computer Interaction*. Springer, 2019, pp. 67–87.

[32] K. Singh, P. Aggarwal, P. Rajivan, and C. Gonzalez, "Training to detect phishing emails: Effects of the frequency of experienced phishing emails," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 63, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2019, pp. 453–457.

[33] R. Virzi, "Usability inspection methods," in *Handbook of Human-Computer Interaction*, M. Helander, T. Landauer, and P. Prabhu, Eds. Elsevier, 1997, pp. 705–715.

[34] A. Vishwanath, T. Herath, R. Chen, J. Wang, and H. R. Rao, "Why do people get phished? testing individual differences in phishing vulnerability within an integrated, information processing model," *Decision Support Systems*, vol. 51, no. 3, pp. 576–586, 2011.

[35] L. Xiao, R. Guadagno, J. Hemsley, and A. Quan-Haase, "Online persuasion mechanisms and processes–a research agenda," in *The 9th Annual International Conference on Social Media and Society*, 2018.

[36] K. Yu, R. Taib, M. A. Butavicius, K. Parsons, and F. Chen, "Mouse behavior as an index of phishing awareness," in *IFIP Conference on Human-Computer Interaction*. Springer, 2019, pp. 539–548.

[37] M. Zuckerman, R. Koestner, and A. O. Alton, "Learning to detect deception," *Journal of Personality and Social Psychology*, vol. 46, no. 3, pp. 519–528, 1984.