Behavioural Public Policy (2026), **10**, 527–539
doi:10.1017/bpp.2024.19

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Phishing feedback: just-in-time intervention improves online security

Svetlana Bender[1], Samantha Horn[2], George Loewenstein[3] and Olivia Roberts[1]

[1]GuideWell, USA, [2]University of Chicago, USA and [3]Carnegie Mellon University, USA
Corresponding author: Olivia Roberts, Email: olivia.roberts@bcbsfl.com

(Received 13 February 2023; revised 23 February 2024; accepted 27 February 2024)

Abstract
Phishing emails cost companies millions. In the absence of technology to perfectly block phishing emails, the responsibility falls on employees to identify and appropriately respond to phishing attempts and on employers to train them to do so. We report results from an experiment with around 11,000 employees of a large U.S. corporation, testing the efficacy of just-in-time feedback delivered at a teachable moment – immediately after succumbing to a phishing email – to reduce susceptibility to phishing emails. Employees in the study were sent an initial pseudo-phishing email, and those who either ignored or fell victim to the phishing email were randomized to receive or not receive feedback about their response. Just-in-time feedback for employees who fell victim to or ignored the initial pseudo-phishing email reduced susceptibility to a second pseudo-phishing email sent by the research team. Additionally, for employees who ignored the initial email, feedback also increased reporting rates.

Keywords: cybersecurity; online security; just-in-time; behavior change; phishing susceptibility; phishing identification

Introduction

As reliance on Internet communication continues to increase, both for work and leisure, computer security is of ever-increasing importance. Among the security threats facing individuals and organizations, phishing – emails designed to induce individuals to reveal personal information, such as passwords and credit card numbers – and spear phishing – emails that appear to come from a known or trusted sender – constitute one of the greatest vulnerabilities. According to IBM's 2023 Cost of Data Breach Report (IBM, 2023), phishing is not only the predominant initial attack method, accounting for 16% of all data breaches, but also the second most financially damaging, costing an average of $4.76 million per incident.[1]

[1]Costs for a phishing attack include (i) lost business due to system downtime, lost customers and reputational damage; (ii) post-breach response activities such as legal expenditures, regulatory fines and a helpdesk for affected customers; (iii) communication costs such as notifications to affected customers and

© GuideWell Mutual Holding Corporation, 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/bpp.2024.19 Published online by Cambridge University Press

Individuals are both the weakest link and the last line of defense against phishing (Sasse et al., 2001). Interventions or technologies that block phishing emails, so they never reach potential victims, are an obvious, but not currently fully effective protective strategy (Jansson and von Solms, 2013; Abbasi et al., 2015). Given the inability to fully shield users from phishing emails, interventions targeted at improving individual-level responses to phishing attacks can complement other, more upstream, interventions (Burns et al., 2019).

In this paper, we present results from a large field study with around 11,000 employees of a large U.S. health solutions organization that evaluates an intervention designed to bolster this last line of defense against phishing. Our intervention draws on two overlapping ideas in the behavioral sciences: teachable moments and just-in-time feedback.

Teachable moments and just-in-time feedback

Teachable moments

Prior research on "teachable moments" has found that the success of interventions to change behavior can depend on the timing of the intervention's delivery (see Lawson and Flocke (2009) and McBride et al. (2003) for a conceptual review). For example, hospitalization has been identified as a propitious time to implement interventions aimed at smoking cessation (Emmons and Goldstein, 1992), alcohol use disorder (Gentilello et al., 1999) and suicide (O'Connor et al., 2020). Some research suggests, moreover, that the length of time during which a teachable moment can be exploited is brief, emphasizing the importance of timing. One prior study found that the impact on the future behavior of attendance at a counselling appointment by patients who abused alcohol was twice as great when appointments occurred less than two days following an emergency department admission compared to those delayed by two days or longer (Williams et al., 2005). The authors conclude that the "half-life" of the teachable moment in this situation was two days.

Just-in-time feedback

A recent movement, mainly in employee management but also in health and education behavior change, involves giving people real-time performance feedback. The assumption underlying just-in-time feedback is that it increases the likelihood that individuals learn associations between an action and its effect. Such feedback can have a greater impact on behavior than the delayed, summarized, feedback that is typically given in, e.g., annual performance reviews. In the domain of behavior change (of which preventing succumbing to phishing would be an example), two papers, both focusing on diet, have shown that just-in-time feedback is effective in changing behavior. Researchers in one study observed a substantial reduction in food intake when dieters were given immediate feedback from a wearable sensor that tracked their 'chew count' and were given the goal of reducing chew counts by 25%

regulator communications and (iv) costs for detection and escalation, such as forensic activities and crisis management.

https://doi.org/10.1017/bpp.2024.19 Published online by Cambridge University Press

(Farooq et al., 2017). They found that just-in-time feedback reduced the amount of food consumed, with no changes in perceived satiety.

The second paper is a review of the methods and results of 31 different just-in-time interventions targeting diet and physical activity. Conclusions from these collective findings were that the most successful interventions involved feedback that was continuously available, personalized, actionable and coupled with a well-defined behavioral objective (Schembre et al., 2018).

Papers addressing the modality and timing of feedback extend beyond health behaviors and management into other realms of human behavior. For instance, in the context of energy consumption, Zangheri et al. (2019) conducted a comprehensive analysis of over 70 studies and concluded that customizing feedback to energy users can result in changes in energy consumption behavior and drive investments in sustainable energy use. However, the success of feedback was found to be dependent on various factors including the medium of feedback delivery, the geographical context and the timeframe in which feedback was provided. In a different paper, Schwartz and Loewenstein (2017) documented the transient impact of emotions produced by affect-inducing stimuli on behaviors intended to combat climate change. While sadness-evoking messages initially led to increased pro-environmental behaviors, such as more engagement with an energy-footprint calculator and heightened donations, these behaviors diminished as the emotional arousal subsided. In contrast, non-emotion-evoking interventions produced much lower initial behavioral responses, but these persisted more over time. There was greater behavioral persistence in response to the emotion-inducing stimulus; however, in a treatment reported in a follow-up study (Study 4), in which participants could make non-binding behavioral commitments immediately following exposure to the emotion-evoking stimuli. These results point to the importance of feedback timing not only in relation to an initial behavior but also in relation to some desired action or commitment to change behavior in the future. We review relevant work on just-in-time feedback in phishing detection and cybersecurity in the next section.

Prior research on phishing detection by email recipients

Teaching

Perhaps the most common type of anti-phishing intervention involves teaching and training. Many organizations offer mandatory or voluntary programs to alert employees to the threat of phishing and instruct them about how to respond. One study examined the impact of a mandatory anti-phishing training program provided to 5,416 employees, at a U.S. healthcare institution, who were classified as "offenders" and "nonoffenders" based on their click rates in response to a series of 20 sham phishing attacks (Gordon et al., 2019). Click rates declined over the course of the study for both groups, most likely as a result of cumulative exposure to the sham attacks. However, there was no evidence that the training program produced a more rapid decline in desired (non)responses to phishing.

Gamified simulations

A specific form of teaching involves computer-based games intended to train individuals to spot phishing attacks (Arachchilage and Love, 2013). One group of

https://doi.org/10.1017/bpp.2024.19 Published online by Cambridge University Press

researchers describes the design and evaluation of an online game, "Anti-Phishing Phil," that teaches users to identify phishing sites and techniques to avoid phishing attacks (Sheng et al., 2007). Evaluating the game in a small study, and comparing it to other interventions, they found that users were better able to identify fraudulent websites immediately after spending 15 min playing the game. All of the interventions led to greater success in resisting phishing, but the game was no better than the alternative interventions that were tested, and the immediacy of the testing raises questions about the persistence, in real-world settings, of any of the interventions.

Warnings

A somewhat weaker and less-expensive intervention involves simply warning individuals about the threat of phishing. Not surprisingly, given the common failure of more elaborate training, such warnings have not proved very effective. In one experiment, for example, 290 visitors to a shopping district in the Netherlands were randomly assigned to receive a warning, a priming intervention or neither (Junger et al., 2017). They were then asked for their email address and nine digits from their 18-digit bank account number. The warning consisted of a leaflet which began with the bold text "Beware of Phishing!" and then briefly explained what a phishing attack is and what it is designed to do. It also included an admonition: "Never **share** personal or banking information **with anyone**!" (emphases from the source). In the priming condition, shoppers were asked, before the sensitive information was solicited, a series of questions gauging their familiarity with phishing and internet security. Relatively high rates of disclosing sensitive information were observed across all conditions: 79.1% of the subjects filled in their email address and 43.5% provided bank account information. However, despite the short time interval between the intervention and the measured behavior, neither the priming questions nor the warning influenced revelation rates.

Related work has investigated the optimal timing of warnings to encourage users to set stronger passwords. Qu et al. (2023) investigated the optimal timing of security warnings in online games and found that warnings presented after, rather than before, a profile set-up activity effectively shifted participants toward stronger security practices in one of two framing conditions tested.

Punishments

In some environments in which phishing is especially damaging, more heavy-handed interventions involving punishments for succumbing to a phishing attempt may be effective. In one study illustrating this potential (Kim et al., 2020), a sham phishing email was sent to employees at a medium-sized manufacturing company. Employees who succumbed to the phishing scam were then randomly assigned to either a punishment condition or a no-punishment control group. The punishment was heavy-handed: employees who succumbed to the phishing attack were visited by a member of the employer's security team, temporarily lost access to the intranet, had to submit a note explaining their misbehavior and were informed that they might receive a negative annual performance evaluation. This punitive intervention significantly reduced susceptibility to a second pseudo-phishing attack by an impressive 25 percentage points. However, even after this reduction, the rate of succumbing to

https://doi.org/10.1017/bpp.2024.19 Published online by Cambridge University Press

the second phishing scam by the punished group was still roughly twice as high as the succumbing rates of employees who had not succumbed to the initial phishing scam. Moreover, the penalizing nature of the intervention would seem unlikely to be implemented in a broad range of workplace settings.

Just-in-time feedback

Several papers report on the implementation of just-in-time feedback delivered to individuals who are exposed to, and fall prey to, simulated phishing attempts (Dodge et al., 2007; Jakobsson et al., 2008), although research to evaluate the efficacy of this approach is less common. Moreover, much of the research that has been done to evaluate the efficacy of feedback following a simulated phishing attempt suffers from methodological limitations. For example, Jansson and von Solms (2013) used a pre-post design[2] (with no control group), in which phishing emails were sent to 25,579 faculty users of an academic email system, followed by instant warning messages to the 1,304 users who succumbed to the attack, followed by a second set of sham emails to the original 25,579 recipients. Although there was a decline in adverse responses to the second test-phishing email, there was also a decline in active users of the email system (faculty who actually read their email), which made it difficult to interpret this result, a problem that could have been mitigated by the inclusion of a no-warning control group.

Two studies in the prior research on phishing prevention are closest in design to our work. In one study (Caputo et al., 2013), researchers sent three sham phishing emails to 1,359 employees at a large organization. Employees who succumbed to at least one of them were randomized to either a control group, which received no offer of training, or four treatment arms, each of which offered training that varied in the framing of the message offering the training. Specifically, the offer message varied gain vs loss frame language and whether it focused on the risk for the individual vs their coworkers. The intervention did not have a statistically significant effect; there were no statistically significant differences in reporting or click rates between the differently framed warning messages or, more importantly, between the average of the warning message groups and the no feedback control.

In a later, but not as methodologically strong study (in part, because it lacked a no-message control condition), Burns et al. (2019) sent a sham phishing email to 260 participants enrolled in a Master of Business Administration (MBA) program. Subsequently, to the 70% of students who succumbed to the attack, they provided links to one of five brief training modules. All of the experimental treatments informed participants that they had been spear phished, and four also provided messages about why it is bad to succumb to a spear phishing attack. Finally, a second phishing email was sent to all 260 individuals who received the first message plus a new no-randomly assigned sample of 140 individuals who had not received the original email. Although no significant differences were found between the randomized treatment groups, given the small sample and the lack of a comparable

[2]Pre-post designs are notoriously subject to threats to internal validity, such as chance events occurring simultaneously ('history'), as well as regression to the mean (Campbell and Stanley, 1963).

https://doi.org/10.1017/bpp.2024.19 Published online by Cambridge University Press

no-intervention control group, the lack of positive results should be treated with caution.

We build upon this existing work by testing the efficacy of an intervention that provides just-in-time feedback at a teachable moment on phishing susceptibility with a large sample and a no-intervention control condition. Although our primary contribution is to test the efficacy of this intervention in the realm of online security, in so doing we also provide additional evidence on the impact of behavioral interventions that leverage just-in-time feedback and teachable moments in a field setting where the impact of such an intervention can be clearly evaluated.

Methods

Participants

We randomly selected 11,802 employees from a large health solutions organization in Florida to participate in our study. Demographic data were not available for all participants but for participants with available data. The mean age was 41.1 years (SD: 13.1) and 67% were female.[3]

Procedures

The study was conducted in three waves of data collection. Wave 1, which was conducted in July 2021, included 3,000 employees; Wave 2, which was conducted in October 2021, included a different set of 3,999 employees and Wave 3, which was conducted in March 2022, included yet another new set of 4,803 employees. The first wave was pre-registered with AsPredicted (https://aspredicted.org/R16_ZNQ).[4] Due to an error in implementation, however, 957 individuals were included in multiple waves.[5] To deal with this problem, only the first instance an individual appears in the study is included in our analyses, resulting in a total sample size by the wave of 3,000, 3,874 and 3,970 for Waves 1, 2 and 3, respectively.

Figure 1 presents the experimental design, which did not differ across experimental waves. The only variation in design across waves was the content of the pseudo-phishing emails that were sent to participants. All study participants were sent two emails by the research team, an initial pseudo-phishing email (*Baseline Email*) and a second pseudo-phishing email (*Assessment Email*). These phishing emails were curated to emulate phishing emails that participants might receive in

[3]Means and standard deviations for age were calculated, omitting 772 unknown values, and gender was not identifiable for 43% of respondents.

[4]The pre-registration was not repeated for Waves 2 and 3 as they were identical to Wave 1 in structure. A sample size of 4,000 was pre-registered, but it was only possible to recruit 3,000 participants for Wave 1. We pre-registered randomizing Reporters into Feedback and No Feedback conditions. However, due to unforeseen limitations in the email reporting software, this aspect of the study could not be implemented as planned. Unfortunately, constraints in the software used to report emails made that infeasible. Additionally, we intended to repeat our analysis incorporating demographic covariates for enhanced robustness, but the absence of demographic data for a majority of participants rendered this aspect of the analysis infeasible.

[5]Specifically, 124 individuals were included in Waves 1 and 2, 716 individuals were included in Waves 1 and 3, 116 individuals were included in Waves 2 and 3, and 1 individual was included in all waves.

https://doi.org/10.1017/bpp.2024.19 Published online by Cambridge University Press

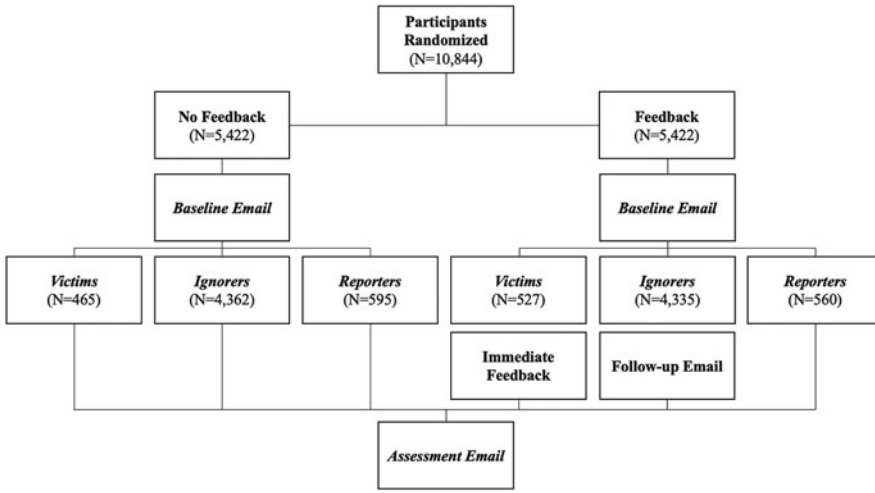


Figure 1. Experimental design. *Note:* Sample sizes are pooled across experimental waves.

the workplace. The content of the phishing emails used in the three waves is provided in Supplementary Appendix Figures A1–A6. Phishing emails offered either a link for the participant to click on or an attachment for the participant to open. Had they been real phishing emails, clicking the link or opening the attachment would have posed risks to the user and the system. In Wave 1, both the *Baseline Email* and the *Assessment Email* included links, and in Waves 2 and 3, the *Baseline Email* included a link and the *Assessment Email* included an attachment.

Participants were randomly assigned to one of two conditions, *Feedback* or *No Feedback*. We then used responses to the *Baseline Email* to divide participants into three groups: (i) Victims – participants who succumbed to the pseudo-phishing attack by clicking on the link in the email or opening the attachment, (ii) Ignorers – participants who did not fall victim to the phishing email but did not report it and (iii) Reporters – participants who did not fall victim to the phishing email did report it. The percent of participants falling into each of these categories varies by the experimental wave as follows: Experimental Wave 1 ($N = 3{,}000$): 7% *Victims*, 78% *Ignorers*, 14% *Reporters*; Experimental Wave 2 ($N = 3{,}874$): 3% *Victims*, 87% *Ignorers*, 9% *Reporters*; Experimental Wave 3 ($N = 3{,}970$): 16% *Victims*, 75% *Ignorers*, 9% *Reporters*.

Participants labeled *Reporters* did not receive different treatments depending on their treatment randomization. All received an email congratulating them on reporting the phishing attempt. *Victims*, in the *Feedback* condition, received a feedback message immediately after they succumbed to the phishing email by clicking on a link or opening an attachment. The feedback message, contained in a webpage pop-up (see Supplementary Appendix Figure A7), informed participants that they had fallen victim to a test-phishing email. The webpage also provided tips for avoiding falling victim to another phishing email and advice to report any subsequent suspicious emails they received. *Victims* in the *No Feedback* condition received no such feedback.

https://doi.org/10.1017/bpp.2024.19 Published online by Cambridge University Press

Ignorers in the *Feedback* condition received a follow-up email from the organization's Information Security team one week after the *Baseline Email*. The follow-up email, which is shown in Supplementary Appendix Figure A8, congratulated the participant for not falling victim to the *Baseline Email* and underscored the importance of reporting suspicious emails. Participants in the *No Feedback* condition received no such feedback. All employees received a second pseudo-email, *Assessment Email*, two weeks after the *Baseline Email* to measure the efficacy of our interventions.

Results

We look at two different outcomes to evaluate response behavior to the *Assessment Email*: (i) a binary variable indicating if a participant fell victim to the *Assessment Email* and (ii) a binary variable indicating if a participant reported the *Assessment Email*. As the intervention differed for *Victims* and *Ignorers*, we consider results separately for these two groups. For all analyses, we pool data across the three experimental waves, including controls for the experimental wave in our specifications.[6]

First, we consider the impact of the feedback treatment on the likelihood that study participants fell victim to the second pseudo-phishing email. These results are presented in Figure 2a with corresponding results from logistic regression analyses in Table 1, Columns 1 and 2.[7] Across all three waves, the proportion of *Baseline Email Victims* falling victim to the *Assessment Email* is lower among those receiving feedback than those who did not. The likelihood that a participant fell victim to the second pseudo-email was 10 percentage points lower for those who received feedback than those who did not – 50% of the control group fell victim to the *Assessment Email* compared to 40% in the treatment group. For *Ignorers*, feedback led to a 2-percentage-point reduction in victim rates (22% in the control group compared to 20% in the treatment group). The improvements in victim rates for *Victims* and *Ignorers* are both statistically significant at conventional levels.

Next, we consider the impact of our treatment on the likelihood that study participants reported the second pseudo-phishing email in Figure 2b, and Table 1, Columns 3 and 4. Directionally there is an improvement in reporting rates among *Victims* (15% in the control group compared to 19% in the treatment group), but this improvement is not statistically significant. For *Ignorers*, there is a statistically significant increase in report rates. The likelihood that *Ignorers* reported the *Assessment Email* was 3 percentage points higher for those who received feedback compared to those who did not (10% in the feedback group compared to 7% in the control group).

[6]For the results of each wave, see Supplementary Appendix Tables A3–A5. For each wave, the treatment effects are directionally similar but not always statistically significant.

[7]Table 1 presents results from a logistic regression of each outcome variable (fell victim and reported) on an indicator for treatment status and indicators for the experimental wave. Supplementary Appendix Table A2 is the same but without wave-fixed effects. The results in the two tables are qualitatively similar and the pattern of statistical significance is the same.

https://doi.org/10.1017/bpp.2024.19 Published online by Cambridge University Press

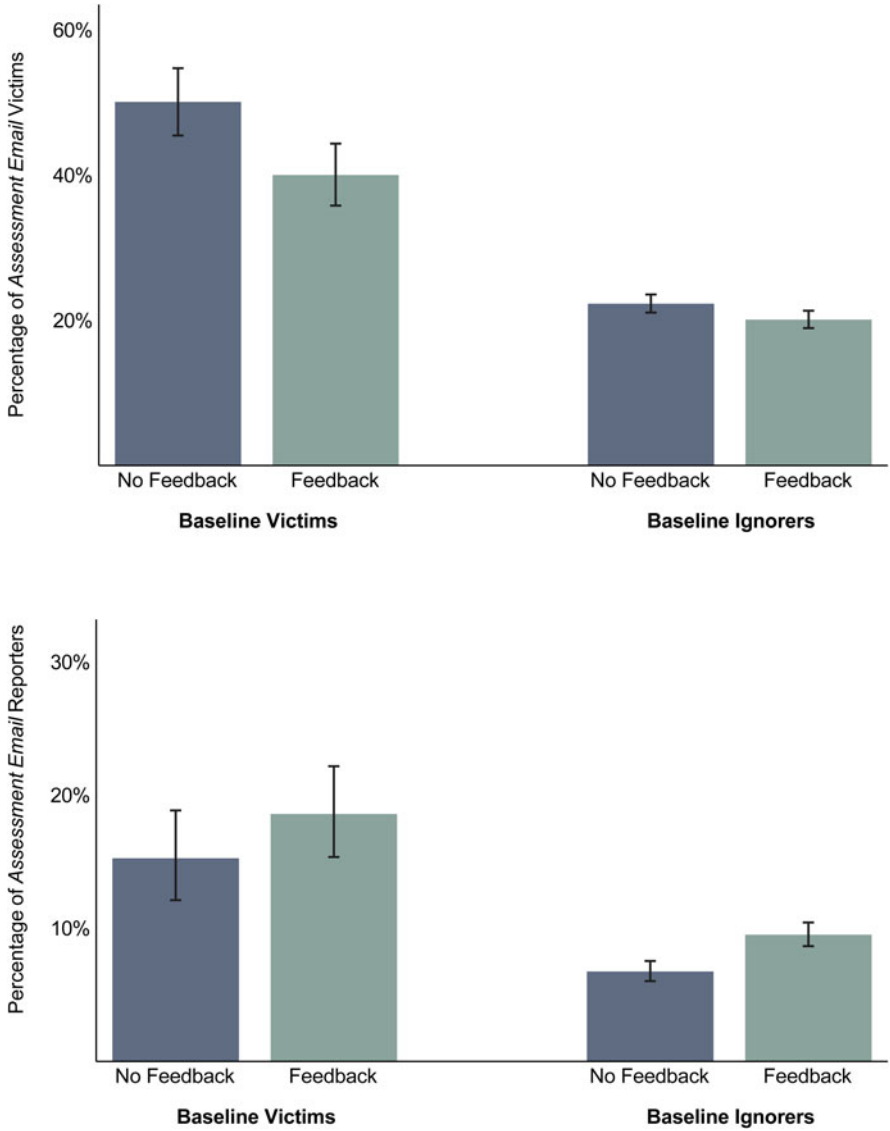


Figure 2. Assessment of email behaviors by the baseline group and the study condition.

Taken together, these results show that just-in-time feedback at a teachable moment reduces vulnerability to phishing attacks among those who are most likely to fall victim to phishing attacks but, for this group, does not significantly increase the likelihood of reporting a phishing attempt. In addition, the delayed feedback experienced by the group who ignored the first pseudo-phishing email reduces the likelihood that participants succumb to a subsequent phishing email and increases their likelihood of reporting a subsequent phishing attack.

https://doi.org/10.1017/bpp.2024.19 Published online by Cambridge University Press

https://doi.org/10.1017/bpp.2024.19 Published online by Cambridge University Press

Table 1. Logistic regression of treatment effects

	(1)	(2)	(3)	(4)
Dependent variable	Fell victim to phishing attempt		Reported phishing attempt	
Sample restriction	Victims	Ignorers	Victims	Ignorers
Received feedback (1/0)	0.660** (0.090)	0.870* (0.048)	1.268 (0.219)	1.463*** (0.118)
Experimental wave 2 (1/0)	5.910*** (1.568)	16.768*** (2.508)	0.155*** (0.065)	0.177*** (0.022)
Experimental wave 3 (1/0)	7.658*** (1.651)	20.779*** (3.112)	0.560** (0.106)	0.717*** (0.062)
Constant	0.194*** (0.041)	0.023*** (0.003)	0.309*** (0.057)	0.127*** (0.010)
N	992	8,697	992	8,697

Notes: Coefficients are odds ratios. Standard error in parentheses.
*p < 0.05, **p < 0.01, and ***p < 0.001.

Conclusion

In this paper, we report results from a large field experiment testing the impact of feedback messages delivered to employees who succumbed or failed to report a pseudo-phishing email. This feedback improved subsequent responses to phishing attempts: employees who fell victim to the first pseudo-phishing email and received feedback were less likely to fall victim to the second pseudo-phishing email relative to those who did not receive feedback. Additionally, employees who ignored the first pseudo-phishing email and received feedback were less likely to fall victim and more likely to report the second pseudo-phishing email relative to those who did not receive feedback.

While our study provides evidence of the potential efficacy of feedback in altering cybersecurity behaviors, some limitations are worth highlighting. First, our experimental design allows us to evaluate the effect of just-in-time feedback against a control condition. However, as the timing of feedback is not varied across conditions, it does not test whether just-in-time feedback outperforms feedback provided at a later time.

Second, and somewhat relatedly, our experimental design does not shed light on the mechanism underlying the impact of feedback on responses to phishing emails. It is unclear if the treatment effect is driven by changes in knowledge, fear of future punishment or emotional responses. Speculatively, we think that an emotional pathway is unlikely, as the time between the feedback and the second phishing email was likely sufficient to dampen emotional response (Gneezy and Imas, 2014); but, given our design, we are unable to conclusively rule out any particular pathway.

Third, an important consideration in our study is the potential influence of the Hawthorne effect, where participants alter their behavior simply because they are aware they are being studied. This awareness could have made participants more vigilant and responsive to phishing attempts during the experiment, potentially impacting the efficacy of the feedback intervention. Similar phenomena have been observed in other field experiments, such as one on residential consumers' electricity use (Schwartz et al., 2013), where mere awareness of participation in a study led to reduced electricity usage.

Fourth, we only test the efficacy of our feedback intervention on a single pseudo-phishing episode. Future work that follows individuals over a series of pseudo-phishing attempts could explore whether individuals exposed to feedback exhibit learning and sustained improvement or revert to pre-intervention behaviors over time.

In conclusion, our findings provide supportive evidence of the role of feedback in improving cybersecurity behaviors. With the increased sophistication of phishing attempts amid the proliferation of artificial intelligence generated text and images, improving individual responses to phishing attempts is increasingly important. Understanding the underlying mechanism and optimal timing of just-in-time phishing feedback are important directions for future work that can, hopefully, inform refinements to enhance the impact of feedback on subsequent behavior.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/bpp.2024.19.

https://doi.org/10.1017/bpp.2024.19 Published online by Cambridge University Press

Acknowledgements. We thank the organization's Information Security Team for their collaborative efforts in the design and execution of phishing experiments including the development of educational material. The organization prioritizes the protection of data and believes training employees is an integral part of a comprehensive security program. S.H. was supported by a Center for Machine Learning and Health (CMLH) Fellowship while contributing to this research.

Author contributions. S.B., S.H., G.L. and O.R. designed the research. O.R. analyzed the data. S.B., S.H., G.L. and O.R. wrote the paper.

Competing interests. The authors declare none.

References

Abbasi, A., R. Y. Lau and D. E. Brown (2015), 'Predicting behavior', *IEEE Intelligent Systems*, **30**(3): 35–43.

Arachchilage, N. A. G. and S. Love (2013), 'A game design framework for avoiding phishing attacks', *Computers in Human Behavior*, **29**(3): 706–714.

Burns, A., M. E. Johnson and D. D. Caputo (2019), 'Spear phishing in a barrel: insights from a targeted phishing campaign', *Journal of Organizational Computing and Electronic Commerce*, **29**(1): 24–39.

Campbell, D. and J. Stanley (1963), 'Experimental and Quasi-Experimental Designs for Research on Teaching', in N. Gage (eds), *Handbook of Research on Teaching*, Chicago, IL: Rand McNally, 171–246.

Caputo, D. D., S. L. Pfleeger, J. D. Freeman and M. E. Johnson (2013), 'Going spear phishing: exploring embedded training and awareness', *IEEE Security & Privacy*, **12**(1): 28–38.

Dodge, R. C. Jr, C. Carver and A. J. Ferguson (2007), 'Phishing for user security awareness', *Computers & Security*, **26**(1): 73–80.

Emmons, K. M. and M. G. Goldstein (1992), 'Smokers who are hospitalized: a window of opportunity for cessation interventions', *Preventive Medicine*, **21**(2): 262–269.

Farooq, M., M. A. McCrory and E. Sazonov (2017), 'Reduction of energy intake using just-in-time feedback from a wearable sensor system', *Obesity*, **25**(4): 676–681.

Gentilello, L. M., A. Villaveces, R. R. Ries, K. S. Nason, E. Daranciang, D. M. Donovan, M. Copass, G. J. Jurkovich and F. P. Rivara (1999), 'Detection of acute alcohol intoxication and chronic alcohol dependence by trauma center staff', *Journal of Trauma and Acute Care Surgery*, **47**(6): 1131.

Gneezy, U. and A. Imas (2014), 'Materazzi effect and the strategic use of anger in competitive interactions', *Proceedings of the National Academy of Sciences*, **111**(4): 1334–1337.

Gordon, W. J., A. Wright, R. J. Glynn, J. Kadakia, C. Mazzone, E. Leinbach and A. Landman (2019), 'Evaluation of a mandatory phishing training program for high-risk employees at a us healthcare system', *Journal of the American Medical Informatics Association*, **26**(6): 547–552.

IBM (2023), IBM Security: Cost of a Data Breach Report 2023. https://www.ibm.com/reports/data-breach [12 October 2023].

Jakobsson, M., N. Johnson and P. Finn (2008), 'Why and how to perform fraud experiments', *IEEE Security & Privacy*, **6**(2): 66–68.

Jansson, K. and R. von Solms (2013), 'Phishing for phishing awareness', *Behaviour & Information Technology*, **32**(6): 584–593.

Junger, M., L. Montoya and F.-J. Overink (2017), 'Priming and warnings are not effective to prevent social engineering attacks', *Computers in Human Behavior*, **66**: 75–87.

Kim, B., D.-Y. Lee and B. Kim (2020), 'Deterrent effects of punishment and training on insider security threats: a field experiment on phishing attacks', *Behaviour & Information Technology*, **39**(11): 1156–1175.

Lawson, P. J. and S. A. Flocke (2009), 'Teachable moments for health behavior change: a concept analysis', *Patient Education and Counseling*, **76**(1): 25–30.

McBride, C. M., K. M. Emmons and I. M. Lipkus (2003), 'Understanding the potential of teachable moments: the case of smoking cessation', *Health Education Research*, **18**(2): 156–170.

O'Connor, S. S., M. M. Mcclay, S. Choudhry, A. D. Shields, R. Carlson, Y. Alonso, K. Lavin, L. Venanzi, K. A. Comtois, J. E. Wilson and S. E. Nicolson (2020), 'Pilot randomized clinical trial of the teachable moment brief intervention for hospitalized suicide attempt survivors', *General Hospital Psychiatry*, **63**: 111–118.

Qu, L., R. Xiao and W. Shi (2023), 'Interactions of framing and timing in nudging online game security', *Computers & Security*, **124**: 102962.

https://doi.org/10.1017/bpp.2024.19 Published online by Cambridge University Press

Sasse, M. A., S. Brostoff and D. Weirich (2001), 'Transforming the "weakest link"—a human/computer interaction approach to usable and effective security', *BT Technology Journal*, **19**(3): 122–131.

Schembre, S. M., Y. Liao, M. C. Robertson, G. F. Dunton, J. Kerr, M. E. Haffey, T. Burnett, K. Basen-Engquist and R. S. Hicklen (2018), 'Just-in-time feedback in diet and physical activity interventions: systematic review and practical design framework', *Journal of Medical Internet Research*, **20**(3): e106.

Schwartz, D. and G. Loewenstein (2017), 'The chill of the moment: emotions and proenvironmental behavior', *Journal of Public Policy & Marketing*, **36**(2): 255–268.

Schwartz, D., B. Fischhoff, T. Krishnamurti and F. Sowell (2013), 'The Hawthorne effect and energy awareness', *Proceedings of the National Academy of Sciences*, **110**(38): 15242–15246.

Sheng, S., B. Magnien, P. Kumaraguru, A. Acquisti, L. F. Cranor, J. Hong and E. Nunge (2007). 'Anti-Phishing Phil: The Design and Evaluation of a Game that Teaches People Not to Fall for Phish'. In *Proceedings of the 3rd symposium on Usable privacy and security*, 88–99.

Williams, S., A. Brown, R. Patton, M. J. Crawford and R. Touquet (2005), 'The half-life of the "teachable moment" for alcohol misusing patients in the emergency department', *Drug and Alcohol Dependence*, **77**(2): 205–208.

Zangheri, P., T. Serrenho and P. Bertoldi (2019), 'Energy savings from feedback systems: a meta-studies' review', *Energies*, **12**(19): 3788.

Cite this article: Bender, S., S. Horn, G. Loewenstein and O. Roberts (2026), 'Phishing feedback: just-in-time intervention improves online security', *Behavioural Public Policy* **10**(3): 527–539. https://doi.org/10.1017/bpp.2024.19

https://doi.org/10.1017/bpp.2024.19 Published online by Cambridge University Press