12-12-2020

Patching The “Human” in Information Security: Using the Inoculation Defense to Confer Resistance Against Phishing Attacks

Dezhi Wu

Jun Zhang

Nicholas Brown

Paul Benjamin Lowry

Gregory D. Moody

Patching The "Human" in Information Security: Using the Inoculation Defense to Confer Resistance Against Phishing Attacks

Dezhi Wu*
University of South Carolina
dezhiwu@cec.sc.edu

Jun Zhang*
University of Science and Technology of China
jzhang90@ustc.edu.cn

Nicholas Brown*[1] &
Paul Benjamin Lowry[2]
Virginia Tech
[1]nichb15@vt.edu
[2]Paul.Lowry.PhD@gmail.com

Greg Moody
University of Nevada Las Vegas
greg.moody@unlv.edu

*Co-first authors

ABSTRACT

The COVID-19 pandemic has transformed the workspace, thrusting countless employees from organizational work settings to their homes, where they work virtually to access key organizational assets through their cyberinfrastructure. This large-scale virtual workforce imposes drastic cybersecurity issues, threats, and challenges to organizations. To onboard and train employees, companies are left with mainly virtual means to deliver SETA training, using two common training approaches: *rule-based* and *mindfulness*. Employees are also facing more challenges and distractions at home where practicing rules and mindfulness can become particularly difficult. Drawing on inoculation theory, this study proposes a new training approach to promote higher resiliency and "umbrella protection" against increasing phishing attacks. This study plans to conduct a mobile phishing SETA training field study at an organization to empirically examine the efficacy of the proposed inoculation-based security training method for work-from-home scenarios.

Keywords

Inoculation theory, resiliency ratio, SETA, ISec, security training, phishing, attack messages

INTRODUCTION

During the COVID-19 pandemic, with 70% of Americans in remote working environments (Hickman and Saad, 2020) and an alarming nearly 700% increase in phishing attacks (Shi, 2020), companies are faced with the daunting dilemma of protecting organizational assets while employees work remotely, many times using their own devices (i.e., bring your own device—BYOD). Companies offer security education training awareness (SETA) programs to train employees how to correctly identify and detect various cyber-attacks, 80% of which can be identified as phishing attacks. Many of these phishing training programs use a *rule-based approach* which requires the signal detection of cues indicating the presence of phishing attempts (Jensen, Dinger, Wright and Thatcher, 2017). Research shows, however, that a majority of individuals are overconfident in their ability, causing them to misinterpret information cues required for signal detection (Wang, Li and Rao, 2016). Conversely, exerting cognitive effort is shown to decrease overconfidence, which has inspired companies to employ new SETA training approaches involving mindfulness techniques (Jensen et al., 2017).

Mindfulness, a technique that promotes attentiveness and awareness to environments, both physical and digital, supplements rule-based instruction to "fill in the gap" in cases where explicit rules are not stated. It conditions individuals to forestall immediate judgment making, thus preventing those hasty outcomes typically accompanying imprudent decisions. With employees handling more roles at home, they are becoming more susceptible to *work-role overload*, a stressor of feeling overburdened and is positively linked to experiencing work distractions, also called *psychological preoccupations* (Cardenas, Major and Bernas, 2004). Added distractions reduce mindfulness (Chiang and Sumell, 2019), and we argue that a new SETA training approach is needed to more effectively inoculate remote-work employees, who must handle the two-prong effects of behavioral (external) and psychological (internal) distractions, from the persuasive attempts of predatory phishers.

Because phishing attacks draw on persuasive rhetoric and familiar cues to lure those who fail to detect signals or who are too distracted to be mindful, we propose a training program grounded on the *inoculation theory*, a theory of attitudinal resistance toward persuasive arguments (Compton, 2013). Inspired by the medical analogy of vaccination making the body resistant to viral threats,

McGuire (1964) proposed a similar notion that attitudes could be made resistant to highly persuasive attacks through preexposure to a weakened form of the attack. The ensuing inoculation would provide attitudinal resistance against rhetorical challenges to held beliefs and attitudes.

The combination of increased cyberattacks on remote workers during the pandemic and the effects of current employees who are psychologically preoccupied or who may use heuristic-systematic processing (Chaiken, 1987) as a cognitive shortcut (heuristic) to process information and make judgments (Goel, Williams and Dincelli, 2017), we propose an inoculation theory-inspired SETA training program. The purpose is to inoculate employees' heuristic-systematic processing against persuasive phishing attempts to buttress protective security behaviors among employees working remotely.

Accordingly, we propose the following two key research questions to guide our study:

RQ1: *How can the inoculation theory be contextualized to cybersecurity for use as a training tool to confer resistance against phishing attacks?*

RQ2: *Does the proposed inoculation defense training program promote greater resiliency against phishing attacks than the two commonly used SETA-trained defenses of mindfulness and rule-based approaches?*

CONCEPTUAL BACKGROUND

Is Phishing Still an Information Security Problem?

A recent study shows that a majority of employees in their new remote working conditions have not received updated security policies regarding data access and handling practices (Samra, 2020). The study further reveals that the lack of at-home IT support and policy guidance has created vast exploitation opportunities for threat actors. A lack of user education and security awareness knowledge contributes to possible victimization, specifically from phishing attacks (Huang, Tan and Liu, 2009). Phishers prey on prospective victims by using a variety of mediated channels (e.g., email, SMS message, telephone) to craft emails, websites, and messages that mimic content of legitimate media, to send personalized messages to *persuade* the recipients "to accept a falsehood and perform a specific action" (Wright, Jensen, Thatcher, Dinger and Marett, 2014, pg. 386).

Goel et al. (2017) describe psychological factors influencing an individual's ability to detect deception cues, but also describe message content attributes and message framing techniques (i.e., contextualized messages related to an individual's specific concerns) that make it increasingly difficult for individuals to detect phishing attempts. Equally, individual differences such as an individual's curiosity, risk propensity, internet anxiety, and general internet usage are known to contribute to phishing susceptibility (Moody, Galletta and Dunn, 2017).

While phishers enhance the specific designs and stimuli used in their phishing attacks, Chen, Gaia and Rao (2020, pg. 10) explain that "phishing susceptibility is not constant." Rather, they explain phishing susceptibility evolves with an individual's personal beliefs. That is, individuals' exposure to new phishing attempts forms their beliefs of possibly succumbing to phishing attacks. Thus, the evolutionary nature of phishing deception, mediums, and types, coupled with the evolving, subjective perception of how individuals view their susceptibility to persuasive phishing attacks create "evolving" opportunities for information security (ISec) researchers to broaden scope and consider novel approaches to phishing detection strategies (Chen et al., 2020).

What Is Inoculation Theory and How Can It Help?

Referred to as "psychological immunization," the inoculation theory is a biological metaphor used to illustrate how one may confer resistance to persuasive messages through the processing of weakened forms of messages that attack an individual's attitudes and beliefs toward one's own efficacy and abilities (Banas and Rains, 2010; Duryea, Ransom and English, 1990). Through this fortifying of beliefs and attitudes, an individual becomes "inoculated" to future persuasive attacks. Research has shown the successful application of the inoculation theory in disciplines such as health communication (Compton, Jackson and Dimmock, 2016), marketing (Lessne and Didow Jr, 1987), advertising (Burgoon, Pfau and Birk, 1995), family communication (Compton and Craig, 2019), tourism (Ivanov, Dillingham, Parker, Rains, Burchett and Geegan, 2018), and social media (Lim and Ki, 2007). It has also been applied to controversial topics such as animal testing (Nabi, 2003), genetically modified food (Wood, 2007), and marijuana legalization (Pfau, Tusing, Koerner, Lee, Godbold, Penaloza, Yang and Hong, 1997).

The process of inoculation begins with a *threat* that triggers an "underlying process of covert counterarguing" (Eagly and Chaiken, 1993, p. 564). Counterarguing, in turn, helps an individual to protect against forthcoming persuasion attacks (McGuire, 1964). Formally, a threat is a persuasive attack against one's attitudes and beliefs. Further, it forewarns of an impending attack and makes salient the "vulnerability of one's current beliefs to change" (Banas and Rains, 2010, p. 285). This threat triggers within a person a near instant defense mechanism to defend his or her beliefs being threatened. Once this process is activated, an individual becomes motivated to strengthen attitudes and will derive existing knowledge he or she possesses to refute the threat. This is called *refutational preemption* and it "provide[s] specific content that receivers can employ to strengthen attitudes against subsequent change" (Pfau et al., 1997, p. 188). More specifically, the individual conceives of counterarguments to counter the persuasion attempts. Thus, the more and diverse counterarguments an individual generates, the broader an umbrella of protection he or she can create against an attack (Pfau et al., 1997). This is seen commonly in politics, where individuals

psychologically guard against various political campaigns' persuasive attacks on their preferred candidates' images and positions (Pfau and Burgoon, 1988).

Refutational pretreatments can be administered to strengthen an individual's counterarguments and responses to threats and persuasion attacks. McGuire (1964) explains this as the primary approach to conferring resistance and further breaks down the types of refutational pretreatment messages. The first one is the *refutational-same message* and it is designed to raise an argument that may be seen in an attack message. It also provides the counterarguments against a similar message when it is later seen. For example, an individual receives an email phishing attempt regarding the urgent need to provide his or her personal information to prevent the deactivation of a 'Wells Fargo' bank account. Because an individual received a *refutational-same* pretreatment involving a 'Wells Fargo' spoofed email, when the individual receives an email from any banking institution, he or she generates near instant counterarguments to refute the claims in the email to urgently respond to any action steps, regardless of whether the email appears legitimate or spoofed.

The second pretreatment type is the *refutational-different message*, and the purpose of this is to present an individual with a completely novel (or different) type of attack he or she may receive. This treatment then provides refutational arguments against a class of attacks (Compton, 2013). For example, an individual inoculated using *refutational-different* messages may receive an entirely new phishing attack through *SMS messaging* or *telephone* regarding possible bank account deactivation. The conditioning received during *email* phishing attacks will immediately trigger counterarguments against novel attacks in different mediums or involving different scenarios.

Last, *perceived involvement* is known to influence the effects of inoculation. It is reasoned that if involvement levels toward a particular issue are high, then an individual can more readily perceive a threat and begin to develop counterarguments toward the persuasive attack. However, if involvement levels are too low, then an individual will not care enough to perceive a threat to his or her attitude (and thus will not generate the counterarguments necessary to resist against persuasive messages) (Compton, 2013). Those who are moderately involved are the most susceptible to a persuasion attack, as they care enough to perceive a threat. They are also the group most amenable to inoculation pretreatments to prevent any subsequent attitudinal changes as a result of a persuasion attack (Compton, 2013).

Figure 1 depicts the process of inoculation. An individual holds a pre-attitude toward an issue, and depending on the level of perceived involvement, may experience a threat which activates counterarguments leading to a response to the attack message and a post-adjustment to attitudes and beliefs (Pfau et al., 1997).

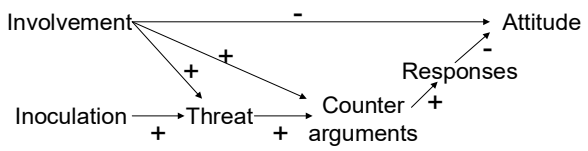

Figure 1. Pfau et al. (1997)'s Process of Inoculation

We reason an inoculation theory-inspired SETA training program is ideal for inoculating those who are only moderately involved in the issue of information security and for those who experience divided attention balancing work and personal roles under work-at-home conditions.

HYPOTHESIS DEVELOPMENT

We propose a set of hypotheses to empirically examine the efficacy of our proposed inoculation-based defense training in comparison to two other types of rule-based and mindfulness-based anti-phishing SETA training approaches. First, we make a distinction in the key dependent variable we will investigate—which deviates from phishing studies conducted in information security literature but aligns with metrics used by practitioners. Researchers (e.g., Goel et al., 2017; Jensen et al., 2017; Moody et al., 2017) when conducting phishing campaigns will follow ethical and technical guidelines such as those established by Finn and Jakobsson (2007) and Jagatic, Johnson, Jakobsson and Menczer (2007). Oftentimes, the performance measure used to evaluate the effectiveness of a campaign is *susceptibility to phishing* (or susceptibility rate) which is the binary measure of whether a subject clicks on a link in an email or performs the requested action of the phisher. Parsons, McCormac, Pattinson, Butavicius and Jerram (2015) state that extra care should be taken to evaluate the effectiveness of simulated phishing campaigns because the "open email," "susceptibility," and "completed survey" rates are usually in small numbers, sometimes leading to low counts in each experimental condition—too few to be analyzed for reporting significant differences (Goel et al. 2017).

Second, the teachable moments and the insights gleaned primarily come from those subjects who fall victim to phishing attacks, oftentimes leaving out inferences that could be made from the majority of subjects who were not susceptible to the attack. Thus, keeping in line with industry practices, our dependent variable will comprise both the *susceptibility rate* and a *reporting rate*, which is the rate at which subjects report phishing to incident responders, to calculate a *resiliency ratio* computed as the reporting rate divided by the susceptibility rate (Figure 2). Formally, r*esiliency ratio* is an industry term that indicates more subjects reported than fell victim to phishing attempts.

Anti-phishing training programs include a description of the types of phishing attacks, the appropriate response behaviors one should take to address the attacks, followed by an opportunity to practice the behavior in a simulated environment.

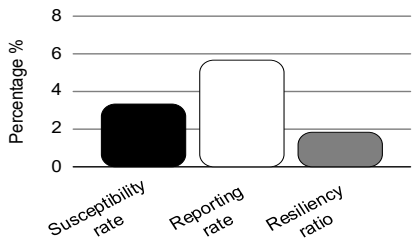

Figure 2. Three Phishing Performance Measures

Upon conclusion, feedback is provided on the results of the planned SETA training system. The training approach is similar in all three methods; however, the stimulus training materials differ. The expected efficacy of the three types of training programs are proposed in our hypotheses.

H1: Individuals without supplemental SETA training will have resiliency ratios significantly lower than those who receive rule-based SETA training.

H2: Individuals using mindfulness SETA training will report higher resiliency ratios than those who receive the rule-based SETA training.

H3: Individuals receiving the inoculation defense SETA training will report higher resiliency ratios than those (a) who received rule-based training only and (b) those who received mindfulness-based training only.

RESEARCH METHDOLOGY

To test our hypotheses, we will design and implement an inoculation-based SETA phishing training system within an organization. We plan to conduct a three-month longitudinal field study to engage work-from-home employees and measure the resiliency ratios (i.e., number of reported cases/number of victim cases) of the three supplemental SETA training approaches.

The study will entail three phases of activities. *Initial phase*: Upon receiving study IRB approval and following the organization's security policies and compliance, we will first recruit study participants in an organization, and then we plan to collect employees' demographic data and their responses to our control variable questions related to, for example, cybersecurity experience, issue involvement toward information security practices and compliance behaviors, distraction factors by work-from-home, and virtual work environment.

Second phase: Every employee participant will receive the same SETA anti-phishing training material, and then will be randomly assigned to one of four experimental conditions: control, rule-based, mindfulness, or inoculation defense. Members in each group will receive a text-only message containing unique recommendations depending on their assigned groups. Members in the control group will receive a message reinforcing the material from the training. Those in the rule-based training condition will receive a message containing a list of recommended actions to take to avoid phishing attacks. In the mindfulness training condition, members will receive a set of recommendations reminding them to *stop*, *think*, and *check* the cues before casting judgment. In the inoculation defense training condition, members will receive a message comprising a forewarning and refutational preemption. We will use Becker, Bavelas and Braden (1961)'s Index of Contingency to measure English sentences to ensure equivalence of all treatment and control messages.

Final phase: After a predetermined time-lag between the second and third phases, we will send the email phishing attack messages and assess employees' susceptibility and reporting rates to identify not only victims but also those practicing protective measures against the attack message by reporting it to the proper incident responders. To determine the efficacy of each SETA training approach, we will calculate the resiliency ratio and provide post-experiment questionnaires to all employees.

EXPECTED CONTRIBUTIONS

The proposed study will contribute to two research streams. First, to the HCI stream, we contribute a novel inoculation defense approach to the design, implementation and delivery of SETA training material to employees in remote work environments. Second, to the ISec literature, we propose a new form of SETA training to promote greater compliance behaviors to strengthen information security policies for those working remotely and using their own devices at home. We further provide a successful contextualization and extension of the inoculation theory, providing ISec researchers a novel lens from which to view compliance behaviors in today's dominant work-from-home environment.

ACKNOWLEDGEMENT

This project is generously funded by a research grant (grant # 80002838) at the University of South Carolina.

REFERENCES

1. Banas, J. A. and Rains, S. A. (2010). A meta-analysis of research on inoculation theory. *Communication Monographs, 77*(3), 281-311.
2. Becker, S. W., Bavelas, A. and Braden, M. (1961). An index to measure contingency of English sentences. *Language and Speech, 4*(3), 139-145.
3. Burgoon, M., Pfau, M. and Birk, T. S. (1995). An inoculation theory explanation for the effects of corporate issue/advocacy advertising campaigns. *Communication Research, 22*(4), 485-505.
4. Cardenas, R. A., Major, D. A. and Bernas, K. H. (2004). Exploring work and family distractions: Antecedents and outcomes. *International Journal of Stress Management, 11*(4), 346-365.
5. Chaiken, S. (1987). The Heuristic Model of Persuasion. In M. P. Zanna, J. M. Olson, & C. P. Herman (Eds.),

Social Influence: The Ontario Symposium (Vol. 5, pp. 3-39). Hillsdale, NJ: Lawrance Eribaum Associates.
6. Chen, R., Gaia, J. and Rao, H. R. (2020). An examination of the effect of recent phishing encounters on phishing susceptibility. *Decision Support Systems, 133*, 113287.
7. Chiang, E. P. and Sumell, A. J. (2019). Are your students absent, not absent, or present? Mindfulness and student performance. *Journal of Economic Education, 50*(1), 1-16.
8. Compton, J. (2013). Inoculation Theory. In J. P. Dillard & L. Shen (Eds.), *The SAGE Handbook of Persuasion: Developments in Theory and Practice* (2nd ed., pp. 220-236). Thousand Oaks, CA, US: Sage Publications, Inc.
9. Compton, J. and Craig, E. A. (2019). Family communication patterns, inoculation theory, and adolescent substance-abuse prevention: Harnessing post-inoculation talk and family communication environments to spread positive influence. *Journal of Family Theory & Review, 11*(2), 277-288.
10. Compton, J., Jackson, B. and Dimmock, J. A. (2016). Persuading others to avoid persuasion: Inoculation theory and resistant health attitudes. *Frontiers in Psychology, 7*, 122.
11. Duryea, E. J., Ransom, M. V. and English, G. (1990). Psychological immunization: Theory, research, and current health behavior applications. *Health Education Quarterly, 17*(2), 169-178.
12. Eagly, A. H. and Chaiken, S. (1993). *The Psychology of Attitudes*. Orlando, FL, US: Harcourt Brace Jovanovich College Publishers.
13. Finn, P. and Jakobsson, M. (2007). Designing ethical phishing experiments. *IEEE Technology and Society Magazine, 26*(1), 46-58.
14. Goel, S., Williams, K. and Dincelli, E. (2017). Got phished? Internet security and human vulnerability. *Journal of the Association for Information Systems, 18*(1), 22-44.
15. Hickman, A. and Saad, L. (2020). Reviewing remote work in the U.S. under COVID-19. Retrieved from https://news.gallup.com/poll/311375/reviewing-remote-work-covid.aspx
16. Huang, H., Tan, J. and Liu, L. (2009). *Countermeasure techniques for deceptive phishing attack*. Paper presented at the Proceedings of the 2009 International Conference on New Trends in Information and Service Science. https://doi.org/10.1109/NISS.2009.80
17. Ivanov, B., Dillingham, L. L., Parker, K. A., Rains, S. A., Burchett, M. and Geegan, S. (2018). Sustainable attitudes: Protecting tourism with inoculation messages. *Annals of Tourism Research, 73*, 26-34.
18. Jagatic, T. N., Johnson, N. A., Jakobsson, M. and Menczer, F. (2007). Social phishing. *Communications of the ACM, 50*(10), 94-100.
19. Jensen, M. L., Dinger, M., Wright, R. T. and Thatcher, J. B. (2017). Training to mitigate phishing attacks using mindfulness techniques. *Journal of Management Information Systems, 34*(2), 597-626.
20. Lessne, G. J. and Didow Jr, N. M. (1987). Inoculation theory and resistance to persuasion in marketing. *Psychology & Marketing, 4*(2), 157-165.
21. Lim, J. S. and Ki, E.-J. (2007). Resistance to ethically suspicious parody video on YouTube: A test of inoculation theory. *Journalism & Mass Communication Quarterly, 84*(4), 713-728.
22. McGuire, W. J. (1964). Inducing Resistance to Persuasion: Some Contemporary Approaches. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 1, pp. 191-229). New York: Academic Press.
23. Moody, G. D., Galletta, D. F. and Dunn, B. K. (2017). Which phish get caught? An exploratory study of individuals′ susceptibility to phishing. *European Journal of Information Systems, 26*(6), 564-584.
24. Nabi, R. L. (2003). "Feeling" resistance: Exploring the role of emotionally evocative visuals in inducing inoculation. *Media Psychology, 5*(2), 199-223.
25. Parsons, K., McCormac, A., Pattinson, M., Butavicius, M. and Jerram, C. (2015). The design of phishing studies: Challenges for researchers. *Computers & Security, 52*, 194-206.
26. Pfau, M. and Burgoon, M. (1988). Inoculation in political campaign communication. *Human Communication Research, 15*(1), 91-111.
27. Pfau, M., Tusing, K. J., Koerner, A. F., Lee, W., Godbold, L. C., Penaloza, L. J., . . . Hong, Y.-H. (1997). Enriching the inoculation construct: The role of critical components in the process of resistance. *Human Communication Research, 24*(2), 187-215.
28. Samra, K. (2020). IBM Security study finds employees new to working from home pose security risk. Retrieved from https://newsroom.ibm.com/2020-06-22-IBM-Security-Study-Finds-Employees-New-to-Working-from-Home-Pose-Security-Risk
29. Shi, F. (2020). Threat spotlight: Coronavirus-related phishing. Retrieved from https://blog.barracuda.com/2020/03/26/threat-spotlight-coronavirus-related-phishing/
30. Wang, J., Li, Y. and Rao, H. R. (2016). Overconfidence in phishing email detection. *Journal of the Association for Information Systems, 17*(11), 759-783.
31. Wood, M. L. M. (2007). Rethinking the inoculation analogy: Effects on subjects with differing preexisting attitudes. *Human Communication Research, 33*(3), 357-378.
32. Wright, R. T., Jensen, M. L., Thatcher, J. B., Dinger, M. and Marett, K. (2014). Influence techniques in phishing attacks: An examination of vulnerability and resistance. *Information Systems Research, 25*(2), 385-400.