Using fMRI to Measure Stimulus Generalization of Software Notification to Security Warnings

Brock Kirwan[1], Bonnie Anderson[1], David Eargle[2], Jeffrey Jenkins[1], and Anthony Vance[3]

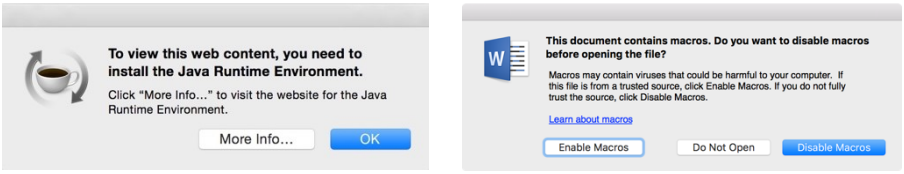
Brigham Young University, Provo, Utah, USA, University of Colorado, Boulder, Colorado, USA, Temple University, Philadelphia, Pennsylvania, USA
[1]{kirwan, bonnie_anderson, jeffrey_jenkins}@byu.edu,
[2]dave@daveeargle.com, [3]anthony@vance.name

Abstract. This paper examines how habituation to frequent software notifications may carry over to infrequent security warnings. This general process—known as stimulus generalization or simply generalization—is a well-established phenomenon in neurobiology that has clear implications for information security. Because software user interface guidelines call for visual consistency, software notifications and security warnings have a similar look and feel. Consequently, through frequent exposure to notifications, people may become habituated to security warnings they have never seen before. The objective of this paper to propose an fMRI experimental design to measure the extent to which this occurs. We also propose testing security warning designs that are resistant to generalization of habituation effects.

Keywords: Security warnings · habituation · generalization · fMRI · mouse cursor tracking · NeuroIS.

1 Introduction

In neurobiology's habituation theory, *stimulus generalization*—or simply *generalization*—occurs when the effects of habituation to one stimulus generalize, or carry over, to other novel stimuli that are similar in appearance [1, 2]. Applied to the domain of human–computer interaction, generalization suggests that users not only habituate to individual security warnings, but also to whole classes of user interface dialogs (e.g., notifications, alerts, confirmations, etc.—hereafter referred to collectively as "notifications" for brevity) that share a similar look and feel (see Figure 1). If true, then the threat and potential impact of habituation is much broader than previous work has suggested [3-5], as users may already be deeply habituated to a security warning that they have never seen before.



System-generated notification Security warning

Fig. 1. A notification and security warning. Note the similarities in UI and mode of interaction.

Building on prior research [6], we outline an experiment using fMRI and, mouse cursor tracking to (1) measure the extent to which a non-clicking mode of interaction for security warning designs can reduce the occurrence of generalization and (2) which mode of interaction is the most effective in reducing the occurrence of generalization.

2 Literature Review

2.1 Habituation and Generalization to Security Warnings

Although habituation to security warnings is well known and has been examined in a number of studies [4, 7-9], the phenomenon of generalization is less well recognized. West noted that "Security messages often resemble other messages dialogs. As a result, security messages may not stand out in importance and users often learn to disregard them" [10, p. 39]. Böhme and Köpsell observed that users' automatic response to notifications "seems to spill over from moderately relevant topics (e. g., EULAs) to more critical ones (online safety and privacy)" [11, p. 2406]. However, neither of these studies empirically examined this effect.

Similarly, researchers have observed that habituation to a single warning in one context can carry over to a different context. For example, Sunshine et al. [12] observed that users who correctly identified the risks of an SSL warning in a library context inappropriately identified these same risks in a banking context. Likewise, Amer et al. [13] found that users who habituated to exception notifications in one context were habituated to a different through visually identical exception notification in a different context. However, in each of these cases, users habituated to the same type of security warning or notification. As a result, it is unclear to what extent software notifications generalize to security warnings.

2.2 Hypotheses

Extending a pilot study that examined the generalization of habituation using Amazon Mechanical Turk, we hypothesize that users' habituation to security messages will generalize to security warning messages. When users repeatedly see software notifications, the brain creates a mental model of these notifications. Rather than giving attention to future exposures to the notifications or similar looking warnings, the brain increasingly relies on this mental model. As a result, users' responses to future warnings

decrease (i.e., habituate) in response to repeated exposures of notifications [14]. In summary, we predict:

H1: Security warnings which are designed with a distinctive look will be more resistant to generalization (as measured by both fMRI activation and mouse cursor movements), and the greater the difference in the look, the lower the amount of generalization.

In addition to habituating to the visual features of notifications, participants may also form high-level memory representations (or schemas) for how to interact with notifications and warnings. According to schema theory [15], schemas can represent general knowledge about objects, situations, or sequences of actions. Similar to habituation, which conserves attentional resources to increase efficiency, the development of schemas for sequences of actions improves behavioral efficiency. Unfortunately, behavioral schemas developed in one situation may generalize to another situation, leading to inappropriate responses. In order to test variations of interaction models, we will develop alternatives to what we call the "click to dismiss" model for interacting with warnings, such as a swipe or slider bar (Figure 2).

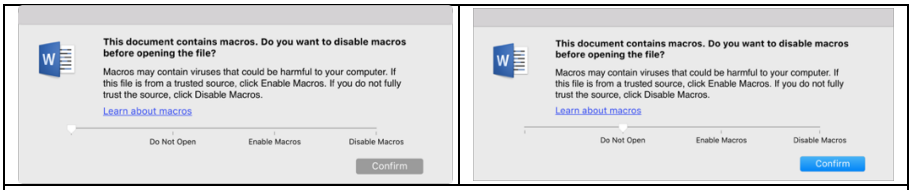


Fig 2. Warning with slider to evaluate an alternative mode of interaction.

By changing the mode of interaction, we predict that users will be able to break out of their schemas and make more considered responses to warnings. Again, we will examine both mouse cursor movements and fMRI activation in response to notifications, compared to warnings that have a distinctive interaction paradigm. In summary, we predict:

H2: Security warning messages which are designed with a change the mode of interaction will be more resistant to generalization (as measured by fMRI activation and mouse cursor movements).

3 Experimental Design

3.1 Methodology

We plan to use fMRI and mouse cursor tracking tools simultaneously while participants receive repeated exposures to notifications and occasional exposures to warnings. Following our pilot work, we will use mouse cursor tracking to behaviorally demonstrate generalization of notification habituation to warnings. We will use fMRI to confirm habituation and generalization in neural activity. FMRI gives us a sensitive

measure of neurocognitive processes that are otherwise difficult to directly observe. MRI data will be collected with a Siemens 3T Tim-Trio scanner and mouse cursor-tracking data will be collected with a custom-built MRI-compatible touchpad.

3.2 Task

We will expose participants to (1) repeated notifications while they perform a simple classification task in the MRI scanner. Additionally, participants will encounter (2) standard security warnings, (3) security warning designs that vary visually from the look and feel of the notifications as in our pilot study, and (4) security warning designs that vary in the mode of interaction from notifications (see Figure 2). Finally, (5) novel software images will be displayed to rule out fatigue.

All stimuli will be presented on an MRI-compatible LCD monitor while fMRI data are collected. Participants will perform an image classification task in a naturalistic manner and interact by means of an MRI-compatible trackpad. Participants will interact with frequent notifications as part of the image classification task. Participants will also interact with occasional security warnings in each treatment condition after repeated exposures to the notifications. The exact timing of stimulus presentation will be based on pilot testing and will vary as a function of participant performance. Timing information for the analysis of fMRI time course data will be determined by both stimulus presentation time and mouse cursor movement onset times and latencies. Standard structural and functional MRI scanning parameters will be used.

In addition to exploratory whole-brain analyses, *a priori* anatomical regions of interest will be examined, including the visual cortex and ventral visual pathway, medial temporal lobe, and motor control regions such as the prefrontal cortex and basal ganglia. The analysis of each of these regions allows for an examination of generalization at different levels of processing. First, the visual cortex and ventral visual pathway are involved in object perception [see 17 for review]. Second, the medial temporal lobe is involved memory specificity, or the detection of differences between similar stimuli [18]. Lastly, motor control regions will allow us to examine the change in motoric scripts and schemas.

3.3 Analysis

We will examine established behavioral and neural indices of habituation, including faster response times and decreased neural activation to repeated stimuli. If generalization occurs, we would expect these same responses to security warnings as well. To isolate these effects from effects due to fatigue, we will compare behavioral and neural responses to warnings and notifications against a novel stimulus that should result in full recovery of responses if the participant is not fatigued.

4 Anticipated Contributions

We anticipate that our findings of this study will complement and extend previous work that examined habituation to individual warnings. With the proposed experimental design, we intend to examine how to mitigate the effects of the generalization of habituation to frequent software notifications. Specifically, our anticipated contributions are:

1. Determine how the effects of habituation to frequent notifications and warnings generalize to novel warnings.
2. Measure the if changing the mode of interaction can reduce generalization.
3. Test warning designs with distinctive modes of interaction to determine the highest resistance to generalization.

5 Future Research

In this paper, we theoretically explain how changing the mode of interaction will contradict existing metal schemas and thereby make a warning more resistant to generalization. We propose running the above experiment to empirically test our hypotheses related to decreasing generalization.

In addition, the mode of interaction can improve security behaviors through several other mechanisms that can be examined in future research. First, the mode of interaction can be used to improve comprehension. For example, Bravo-Lillo et al. [4] had people highlight key text-components in the warning before they were allowed to reject it, improving comprehension of the security risk. The mode of interaction can also help users understand the consequence of the action. For instance, one could have a person perform an action that imitates the potential danger of ignoring the warning before allowing them to reject it (e.g., dragging a password to a hacker icon, disabling a lock that makes a computer public, etc.).

Second, the mode of interaction can influence the amount of work required to behave non-securely, and thereby make alternative, more secure behaviors, more appealing. It is often easier to engage in risky behavior than secure behavior, resulting in poor security decisions. For example, ignoring an SSL warning often requires less effort than try to find an alternative website to address the user's need that is secure [19]. While much research has focused on making security more usable [20], less research has focused on making non-secure behaviors less usable. However, the mode of interaction can be used to accomplish this objective and thereby improve secure behaviors. For example, for SSL warnings on Chrome, you must click on 'Advanced" before finding an option to dismiss the warning. This extra work can help deter non-secure behavior and promote easier secure behaviors.

Acknowledgements. This research was funded by NSF Grant #CNS-1931108.

References

1. Rankin, C.H., et al., *Habituation revisited: An updated and revised description of the behavioral characteristics of habituation.* Neurobiology of Learning and Memory, 2009. **92**(2): p. 135-138.
2. Thompson, R.F. and W.A. Spencer, *Habituation: A Model Phenomenon for the Study of Neuronal Substrates of Behavior.* Psychological Review, 1966. **73**(1): p. 16-43.
3. Anderson, B.B., et al., *How Polymorphic Warnings Reduce Habituation in the Brain: Insights from an fMRI Study*, in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2015, ACM: Seoul, Republic of Korea. p. 2883-2892.
4. Bravo-Lillo, C., et al., *Your Attention Please: Designing Security-decision UIs to Make Genuine Risks Harder to Ignore*, in *Proceedings of the Ninth Symposium on Usable Privacy and Security* 2013, ACM: Newcastle, United Kingdom. p. 1-12.
5. Egelman, S., L.F. Cranor, and J. Hong, *You've Been Warned: An Empirical Study of the Effectiveness of Web Browser Phishing Warnings*, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2008, ACM: Florence, Italy. p. 1065-1074.
6. Anderson, B., et al. *It All Blurs Together: How the Effects of Habituation Generalize across System Notifications and Security warnings*. in *NeuroIS Retreat 2016*. 2016. Gmunden, Austria: Springer International.
7. Bravo-Lillo, C., et al. *Harder to Ignore? Revisiting Pop-Up Fatigue and Approaches to Prevent It*. in *10th Symposium On Usable Privacy and Security (SOUPS 2014)*. 2014. USENIX Association.
8. Brustoloni, J.C. and R. Villamarín-Salomón. *Improving Security Decisions with Polymorphic and Audited Dialogs*. in *Proceedings of the Third Symposium on Usable Privacy and Security (SOUPS 2007)*. 2007. New York, NY, USA: ACM.
9. Vance, A., et al., *Tuning Out Security Warnings: A Longitudinal Examination of Habituation through fMRI, Eye Tracking, and Field Experiments.* MIS Quarterly, 2018. **42**(2): p. 355-380.
10. West, R., *The Psychology of Security.* Communications of the ACM, 2008. **51**(4): p. 34-40.
11. Böhme, R. and S. Köpsell. *Trained to Accept?: A Field Experiment on Consent Dialogs*. in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)* 2010. Atlanta: ACM.
12. Sunshine, J., et al. *Crying Wolf: An Empirical Study of SSL Warning Effectiveness*. in *SSYM'09 Proceedings of the 18th Conference on USENIX Security Symposium*. 2009. Montreal, Canada.
13. Amer, T.S. and J.-M.B. Maris, *Signal words and signal icons in application control and information technology exception messages—Hazard matching and habituation effects.* Journal of Information Systems, 2007. **21**(2): p. 1-25.
14. Groves, P.M. and R.F. Thompson, *Habituation: A Dual-Process Theory.* Psychological Review, 1970. **77**: p. 419-450.

15. Rumelhart, D.E., *Schemata: the building blocks of cognition*, in *Theoretical Issues in Reading Comprehension*, R.J. Spiro, Editor. 1980, Lawrence Erlbaum: Hillsdale, NJ.
16. Nosek, B.A. and M.R. Banaji, *The go/no-go association task.* Social cognition, 2001. **19**(6): p. 625-666.
17. Grill-Spector, K., *The neural basis of object perception.* Current Opinion in Neurobiology, 2003. **13**(2): p. 159-166.
18. Kirwan, C.B. and C.E.L. Stark, *Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe.* Learning & Memory, 2007. **14**(9): p. 625-633.
19. Adams, A. and M.A. Sasse, *Users are not the enemy.* Communications of the ACM, 1999. **42**(12): p. 40-46.
20. Balfanz, D., et al., *In Search of Us-able security: Five Lessons from the Field. IEEE.* IEEE Security & Privacy, 2004. **2**(5): p. 19-24.