The Fog of Warnings: How Non-essential Notifications Blur with Security Warnings

Anthony Vance, *Temple University* David Eargle, *University of Colorado Boulder*
Jeffrey L. Jenkins, *Brigham Young University* C. Brock Kirwan, *Brigham Young University*
Bonnie Brinton Anderson, *Brigham Young University*

Abstract

Adherence to security warnings continues to be an important problem in information security. Although users may fail to heed a security warning for a variety of reasons, a major contributor is habituation, which is decreased response to repeated stimulation. However, the scope of this problem may actually be much broader than previously thought because of the neurobiological phenomenon of generalization. Whereas habituation describes a diminished response with repetitions of the same stimulus, generalization occurs when habituation to one stimulus carries over to other novel stimuli that are similar in appearance.

Generalization has important implications for the domains of usable security and human–computer interaction. Because a basic principle of user interface design is visual consistency, generalization suggests that through exposure to frequent non-security-related notifications (e.g., dialogs, alerts, confirmations, etc.) that share a similar look and feel, users may become deeply habituated to critical security warnings that they have never seen before. Further, with the increasing number of notifications in our lives across a range of mobile, Internet of Things, and computing devices, the accumulated effect of generalization may be substantial. However, this problem has not been empirically examined before.

This paper contributes by measuring the impacts of generalization in terms of (1) diminished attention via mouse cursor tracking and (2) users' ability to behaviorally adhere to security warnings. Through an online experiment, we find that:

- Habituation to a frequent non-security-related notification does carry over to a one-time security warning.
- Generalization of habituation is manifest both in (1) decreased attention to warnings and (2) lower warning adherence behavior.
- The carry-over effect, most importantly, is due to generalization, and not fatigue.
- The degree that generalization occurs depends on the similarity in look and feel between a notification and warning.

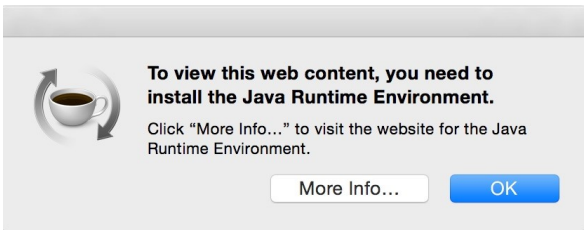
These findings open new avenues of research and provide guidance to software developers for creating warnings that are more resistant to the effects of generalization of habituation, thereby improving users' security warning adherence.

USENIX Symposium on Usable Privacy and Security (SOUPS) 2019.
August 11 -- 13, 2019, Santa Clara, CA, USA.

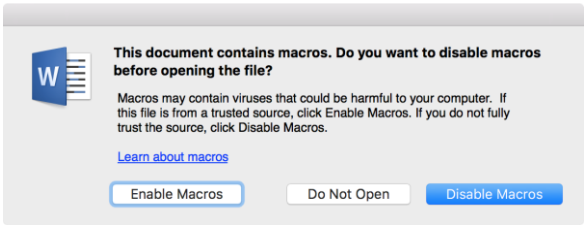
1. Introduction

Users' adherence to security warnings continues to be an important problem in information security because warnings are often the last defense standing between a user and compromise [1, 36]. Although users may fail to heed a warning for a variety of reasons [24], an important contributor is habituation, which is defined as decreased response to repeated stimulation [9, 11, 18, 19, 25]. This phenomenon is fundamentally neurobiological in nature [23], and past work has shown how the brain habituates to security warnings over time [5, 34].

However, there is a key aspect of neurobiology's habituation theory that has not been examined but that has critical implications for security warnings. *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 [23, 31]. 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 (UI) notifications (e.g., dialogs, alerts, confirmations, etc.—hereafter referred to collectively as "notifications" for brevity) that share a similar look and feel (see Figure 1).



System-generated notification



Security warning

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

Consistency of look and feel is a foundational principle in UI design [14, 21] and is reinforced by major software companies, such as Apple and Microsoft, which provide development libraries and guidelines to ensure consistency across software applications [8, 22]. As a result, users may already be deeply habituated to a security warning that they have never seen before.

With the increasing number of notifications in the lives of users across a range of mobile, Internet of Things, and computing devices, the accumulated effect of generalization may be substantial, lessening the effectiveness of comparatively rare security warnings that are truly critical. For example, an analysis of 40,191 Android users showed that they received an average of 26 notifications per day on their mobile devices, not including apps that "flood" users with notifications, such as Skype, Viber, and DropSync [26]. In such a saturated environment, it is crucial that habituation to notifications not generalize to security warnings; the latter are to have protective value.

Although the problem of the blurring of security warnings and notifications has previously been recognized (e.g., [9, 33]), it has not been empirically studied. Consequently, the scope and severity of generalization, as well as the conditions under which it occurs, are not known. By measuring these things, we can better understand how generalization occurs and mitigate its influence.

The objective of this research is to measure and explain how habituation to a frequent non-security-related notification generalizes or carries over to security warnings. In doing so, we answer the following research questions:

> RQ1: Does habituation to non-security-related notifications generalize to security warnings?
>
> RQ2: Does the degree of look-and-feel similarity influence the amount of generalization of habituation?

Using mouse cursor tracking and other behavioral responses in an online experiment, we show that:

- Habituation to a frequent non-security-related notification does carry over to a one-time security warning.
- Generalization of habituation is manifest both in (1) decreased attention to warnings and (2) lower warning adherence behavior.
- Importantly, we show that this carry-over effect is due to generalization, and not fatigue.
- The degree that generalization occurs depends on the similarity in look and feel between a notification and warning.

These findings help form a foundation for developing warning designs that are resistant to the influence of generalization.

2. Related Work

Generalization in Useable Security Research

Although habituation to security warnings is well known and has been examined in a number of studies [10-12, 38], the phenomenon of generalization is less well recognized. West noted that "security messages often resemble other message dialogs. As a result, security messages may not stand out in importance and users often learn to disregard them" [37, p. 39]. Böhme and Köpsell observed that a user's automatic response to notifications "seems to spill over from moderately relevant topics (e.g., EULAs) to more critical ones (online safety and privacy)" [9, 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, Egelman et al. [15] observed that some lab participants disregarded a phishing warning because they confused it with a previous warning they had seen. However, this was an incidental observation and not the focus of their study. They speculated that warning visual similarity caused the confusion, but they did not test this supposition. Similarly, Sunshine et al. [29] 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. [3] found that users who habituated to exception notifications in one context were habituated to a different though visually identical exception notification in a different context. However, in each of these cases, the 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.

Generalization in Neuroscience Research

As users respond repeatedly to notifications, they are likely to devote fewer neural resources toward those stimuli, either through habituation or through perceptual learning [4, 6, 7, 34]. Perceptual learning occurs when there is a structural change in visual processing structures of the brain to support performance on a perceptual task as a result of previous visual experience [16]. The neuroscience literature has long shown that this increased efficiency of the neural response comes at the price of generalizing from one stimulus set to another similar set of stimuli [31].

Generalization has been demonstrated in the neuroscience literature at a number of different levels [31], including decreased neural responses to stimuli similar to habituated stimuli [23], the transfer of perceptual learning to novel tasks [13], and the retrieval of long-term memory representations to similar memory cues [20]. Habituation is typically short-lived, as neural responses typically return to baseline after a delay. Conversely, perceptual learning can be long-lasting, can occur without overt attention [13], and is more likely to be involved in more complex tasks (such as using complex software) [17].

3. Methods

In order to examine generalization, we designed an online experiment to measure habituation (a prerequisite condition for generalization), generalization, and warning adherence behavior. Research shows that people are not very accurate in self-reporting security behavior [35], so we instead captured direct behavioral measures. First, we measured habituation in terms of the mousing speed of users' responses to notifications and warnings as measured via mouse cursor tracking. Previous research has demonstrated this to be a robust measure of habituation to security warnings [5, 33, 34]. Similarly, we also measured habituation in terms of the time between the display of a notification or warning and when a user responded to it. Finally, we also measured users' adherence to the security warning, "the rates at which users do not proceed through a warning, i.e., the rate at which they choose the safer option" [24, p.7].

Participants

We recruited 600 participants via Amazon Mechanical Turk (mTurk). Following Steelman et al. [28], all participants were required to be from the United States. The average age of participants was 36 years old (min: 18, max 76); 53% were male. Participants were ultimately paid $1.50 ($1.00 up front, with a $0.50 bonus) for an approximately five-minute task. Table 1 shows the participant breakdown per condition.

Ethics

The university Institutional Review Board (IRB) approved the protocol used. In an informed consent statement, participants were told that the study objective was to determine how people visually evaluate and cognitively process computer software messages. They were also told that in the experimental task they would be browsing websites and perform simple tasks such as comparing images. However, participants were not told that we were specifically interested in their response to security warnings. At the end of the experiment, participants were debriefed about the specific objectives of the experiment.

Experimental Task

We followed a previously established experimental protocol in which participants classified images on the web as either animated or photographic versions of Batman [32]. Participants from Amazon Mechanical Turk were required to use the Firefox browser and were directed to a server on which we hosted our experiment. A dashboard allowed participants to classify each loaded image (Figure 2).

In pre-task instructions, participants were told that random webpages containing images of Batman would be loaded into a central frame on the task dashboard. Using the following language, participants were told that because the sites that would be loaded were random and external, some risk to their devices was involved:

"Warning: The researchers are not responsible for the content of the webpages loaded into the center frame. By participating in this task, you understand that despite the pages being in a center frame, the risks are the same as if you were visiting the pages directly. You assume all risks associated with visiting these websites."

Participants went through a task warm-up "internet connectivity" test where two actual live external pages were loaded into the central frame, which participants were instructed to interact with and peruse. However, in reality, the main Batman classification task loaded *static* screenshot images of websites with photos of Batman into the central dashboard frame. This allowed us to control what participants saw during the task.

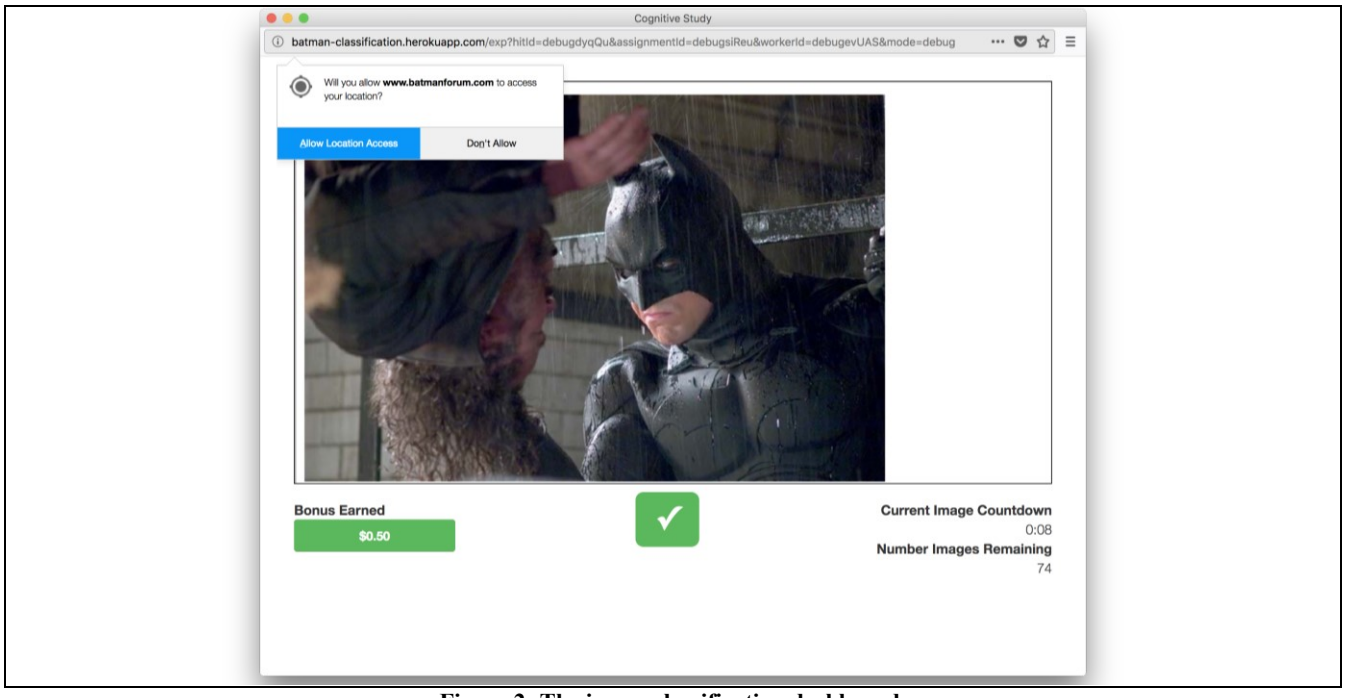


Figure 2: The image classification dashboard.

We reasoned that if participants thought that the task was loading real external websites, then they would be more likely to believe that the appearance of a popup security notification was triggered by the loaded external Batman website, as opposed to by the experiment dashboard. The source URLs that we put int the text of some of the security warnings reinforced the perception that the external sites were triggering the warnings to appear. We also encouraged a belief that the task loaded unregulated external websites in a bid to dampen the likelihood of lab experiment bias [27], wherein participants may feel an invincibility against threats because they feel secure within the walled confines of an artificial experiment approved by an ethics board. Our analysis suggested that participants believed security popups were real (see section 4.1).

Participants were under time pressure to complete the task. For each website, participants had ten seconds in which to classify the image. Failure to classify the image was counted as an incorrect answer. A performance bar in the bottom-left corner of the screen provided participants with live feedback of their current bonus standing. Initially, the bar was green, but an incorrect classification decreased a participant's bonus by 5 cents, updating the bonus bar with a depressing red slider animation from the right side to represent the loss. We had the bonus be dependent on performance in order to encourage continued participant engagement with the task. In reality, however, all participants received the full bonus regardless of their performance. They were informed of their full reward as part of the post-task debrief.

After the internet connectivity test and instructions, participants first completed a warm-up round of four Batman image classifications, during which no popups or security warnings appeared, before beginning what they thought would be 75 total image classifications. After each classification in the non-warmup 75-set, a HTML5-styled notification styled after the Firefox location permission request reported the participant's current classification performance (see Figure 3). Importantly, participants had to click a "continue" button on this performance notification before going on to the next image, thus forcing them to interact with each notification. Each participants encountered a single randomly-assigned security warning during their task after a randomly-assigned number of interactions with Batman image classifications and performance notifications. Once participants saw their security warning, the main classification experimental task abruptly terminated. Javascript recorded all participant interactions during the task, including mouse cursor movements, reaction times, and security warning choice click-behavior. Following the main task, participants were directed to a short post-task survey and debrief, after which the experiment was complete.

In summary, we chose the Batman protocol because it provided an excuse to show participants, who were using their own computers, multiple ostensibly-real browser task-related notifications within a short timeframe, one of which was a security notification supposedly triggered by a non-experimenter-controlled external website, in a closely-web-observable (through javascript) environment.

Experimental Treatments

To answer our research questions, we randomly assigned participants to 1 of 10 experimental conditions in a 2

(manipulating generalization) × 5 (manipulating the similarity of the look-and-feel) factorial experimental design (Table 1). First, we manipulate generalization by either displaying the warning first or after a series of notifications. Second, we manipulate how similar the look-and-feel is between the notification and the target stimulus (using four security warnings in Firefox with varying look-and-feel similarity to the notification, and a novel stimulus). We describe these manipulations in more detail below.

Table 1: Experimental Design (2x5, fully-crossed) with cell *n*'s.

	Appeared After Classification	
Security Warning Type	Position 1	Position 15
Permission warning	$n = 59$	$n = 60$
Extension warning	$n = 60$	$n = 61$
Save executable	$n = 60$	$n = 60$
Open macro	$n = 60$	$n = 60$
Novel stimulus	$n = 60$	$n = 60$

In order to assess whether habituation to notifications generalized to security warnings, we first manipulated whether participants were habituated to notifications by assigning them to view a security warning either after the *first* Batman image classification or after the *fifteenth* image classification. By measuring responses to warnings at both positions, we could measure and control for differences within each security warning type between its two appearance positions, as well as calculate differences across security warning types for a given position. Participants who were in one the "position 15" treatment groups classified 15 Batman images, with a performance notification being shown after each of the first 14 Batman classifications, and their assigned security warning being shown after the 15th Batman image classification instead of a performance notification, followed by an abrupt task termination. Participants who were in one of the "position 1" treatment groups only classified one Batman image, after which they saw and interacted with their assigned security warning, followed by an abrupt task termination. This means that participants who saw a security warning position 1 did not see *any* performance notifications – they only saw one Batman and one security notification.

We also manipulated the type and look of the security warning. Participants were randomly assigned to view either one of four different simulated Firefox security warnings, or a visually novel stimulus (described in section 3.5). The Firefox security warnings were chosen because they had varying levels of look-and-feel similarity to the task-performance notification, which helps address our second research question (see Figure 3). The most visually similar security warning to the performance notification was the location permission warning ("permission warning"; Figure 4); the second-most visually-similar was a Firefox add-on installation permission warning ("extension warning";

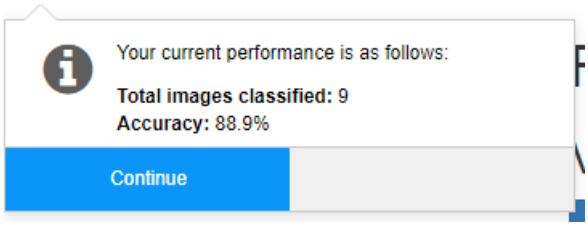


Figure 3: HTML5 performance notification.

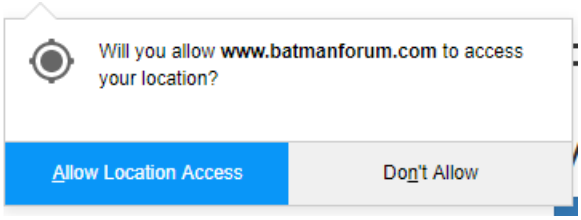


Figure 4: HTML5 permission warning.

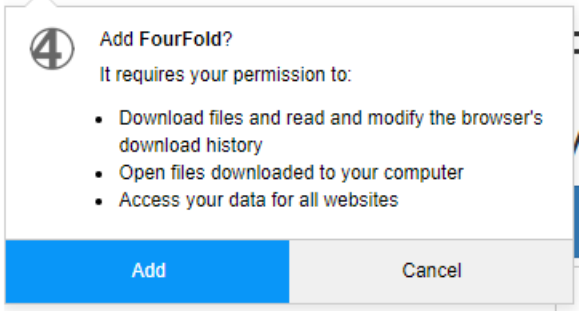


Figure 5: Firefox add-on (extension) warning.

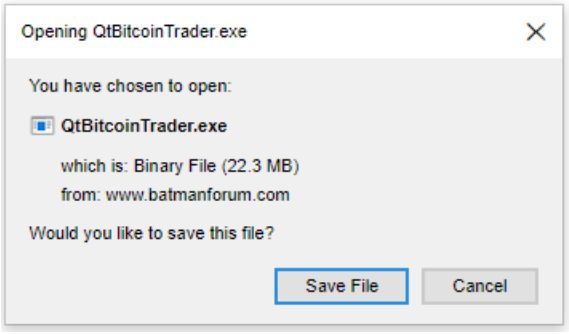


Figure 6: Firefox save executable message.

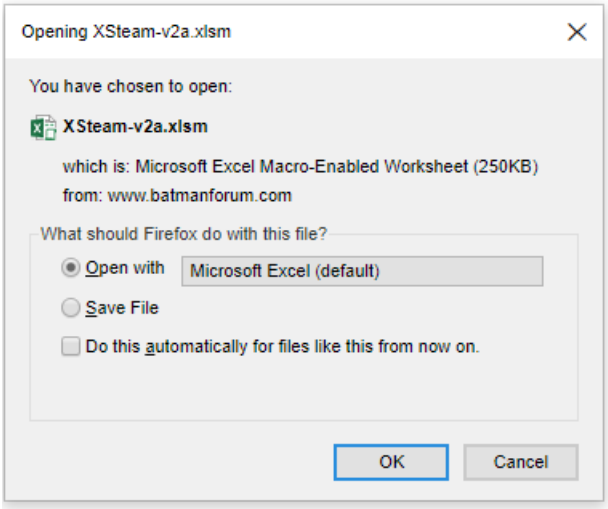


Figure 7: Firefox open macro-enabled spreadsheet message.

Figure 5); and the most visually discrepant security warnings compared to the performance notification were Firefox save-executable message ("executable save"; Figure 6), and a Firefox 'open a macro-enabled spreadsheet' message ("open macro"; Figure 7). Each of these four fake security notifications were designed in HTML5 and javascript to look just as would their legitimate Firefox warning counterparts.

We recognize that the save executable and open macro messages are not security warnings, strictly speaking, because they do not actually warn the user of anything. However, these messages do have strong security implications. In particular, opening documents with malicious macros are a longtime and increasingly popular avenue of attack [30]. For simplicity, we refer to all of our security message treatments as warnings.

Ruling out the effect of fatigue

To rule out the effect of fatigue, we designed a treatment that was visually novel compared to the other notification and security warnings (Figure 8). Following the neurobiological literature [23], generalization of habituation is measured by showing that once a participant habituates to a stimulus, a neural or behavioral response shows little increase when a novel stimulus is presented that is similar to the original stimulus. However, when a novel stimulus—an image of a yellow duck—is presented that is very different from the original stimulus, the response recovers to where it was before any stimuli were displayed, thus demonstrating that fatigue was not the reason for the diminished response to similar stimuli [23]. Participants assigned to the novel-stimulus condition saw it at either position 1 and position 15, which allowed us to test for differences between positions. Any slower reaction times between participants who saw the duck at position 15 versus position 1 would be indicative of fatigue or of general task dismiss-the-notification familiarity for the former group. If there was evidence of such fatigue within the duck position-treatments, then we could control for that magnitude of fatigue in our other security warning tests.

4. Analysis

Realism check

The real-website ruse worked—participants were successfully led to believe that security warnings were triggered by the loaded websites they automatically visited. Both quantitative and qualitative (after the debrief) responses from participants supported that they held this belief. For instances, in a free-response field on the post-task survey, one participant said "The pop up was unexpected and I thought I might have clicked on something wrong. I did pause for a second and panic," and another said "That was incredible deception. I am a software engineer with a background in cybersecurity and you fooled [me]." A third stated, "I got bamboozled." When asked in the survey about their perceived realism of the security messages that they

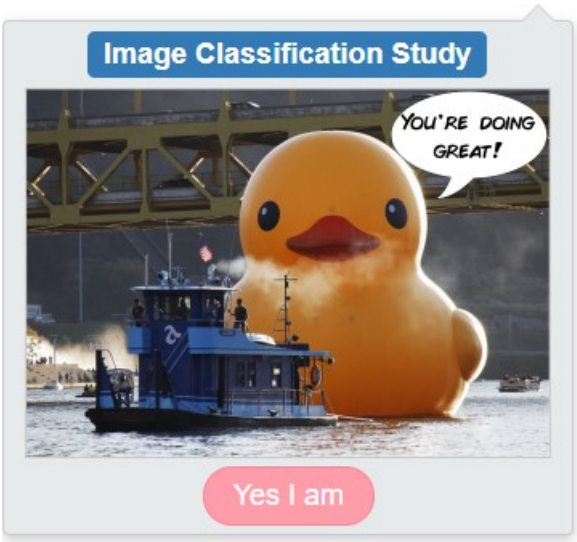


Figure 8: Novel stimulus for assessing fatigue.

saw, participants rated the security message mockups well above 5 out of 10 (see Figure 9).

Adherence Behavior

We measured whether participants who saw a security warning clicked through it (e.g., taking the "accept" or "proceed" action for one of Figures 5–8). By comparing click-through rates for each security warning between its two appearance positions, we can test whether generalization had an impact on an actual security behavior — which it indeed did.

We built a logistic regression model including only those who received warnings and not the novel stimulus ($N = 487$, Nagelkerke's $R^2 = .546$), which predicted whether participants clicked through their security warning. A click-through was coded as a 1, and any action that dismissed the warning without clicking through was coded as a 0. Independent variables were the security warning type (permission warning, extension warning, save executable,

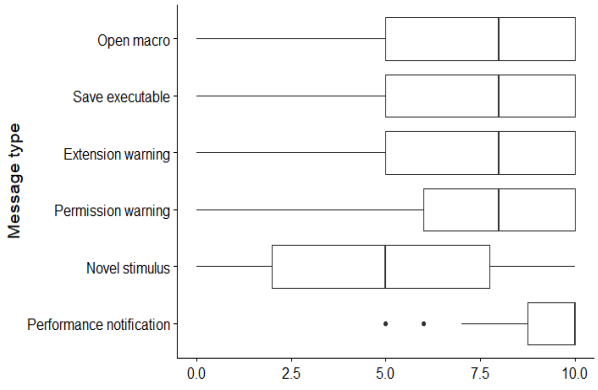


Figure 9. Realism of message (self-reported, scale of 0 to 10). Note that perceived realism was not required for the novel stimulus.

open macro), crossed with the position or order in which the warning was displayed (position 1, position 15). The model fit is shown in Table 2

The permission warning was more likely to be clicked through if seen at position 15 than at position 1 (OR = 2.60, $p = 0.008$), as was the extension request (OR = 1.95, *one-tailed* $p = .047$). No differences in click-through behaviors between positions were observed for either the open-macro (OR=0.59, $p = .192$) or the save-executable warnings (logOdds = 1.00, $p = 1.00$) (see Figure 10 and Table 2). As the permission request and extension request are more

visually similar to the performance notification than the open-macro and save-executable warnings, these findings support that the similar look-and-feel of security warnings to other notifications may be magnifying generalization.

Mouse cursor movement speed

As an indicator of habituation, we used mouse cursor movement speed as a dependent variable to test whether habituation to non-security notifications generalizes to security warnings. Movement speed refers to how fast a user moves over the warning to dismiss or adhere to the warning (in pixels traversed per millisecond). Faster movement speed indicates that the user is paying less attention to the content of the warning, and that the user is providing a habituated response to the warning. Slower movement speed indicates that the user is paying more attention to the warning and providing a non-habituated response to the warning [33].

Table 2. Click-through predicted by interaction of warning type and appearance position, 0-intercept for ease of interpreting within-type slopes.

did_click_through ~ 0 + security_message + security_message:showSecurityMessageAt

	Clicked-through		
Predictors	*Odds Ratios (OR)*	*CI*	*P (one-tailed)*
Permission warning	0.33	0.18 – 0.58	**<0.001**
Extension warning	0.33	0.19 – 0.60	**<0.001**
Save-Executable warning	0.03	0.01 – 0.14	**<0.001**
Open-macro warning	0.15	0.07 – 0.32	**<0.001**
Permission warning × position 15	2.60	1.20 – 5.62	**0.008**
Extension warning × position 15	1.95	0.89 – 4.24	**0.047**
Save-executable × position 15	1.00	0.14 – 7.34	1.000
Open-macro × position 15	0.59	0.18 – 1.92	0.192
Observations	487		
Cox & Snell's R^2 / Nagelkerke's R^2	0.409 / 0.546		

We conducted several analyses to examine how generalization influences movement speed. First, we limited the data just to the warnings, and examined whether the position of the warning (1 or 15) influences movement speed. If the position of the warning influences movement speed, this indicates that habituation to the non-security

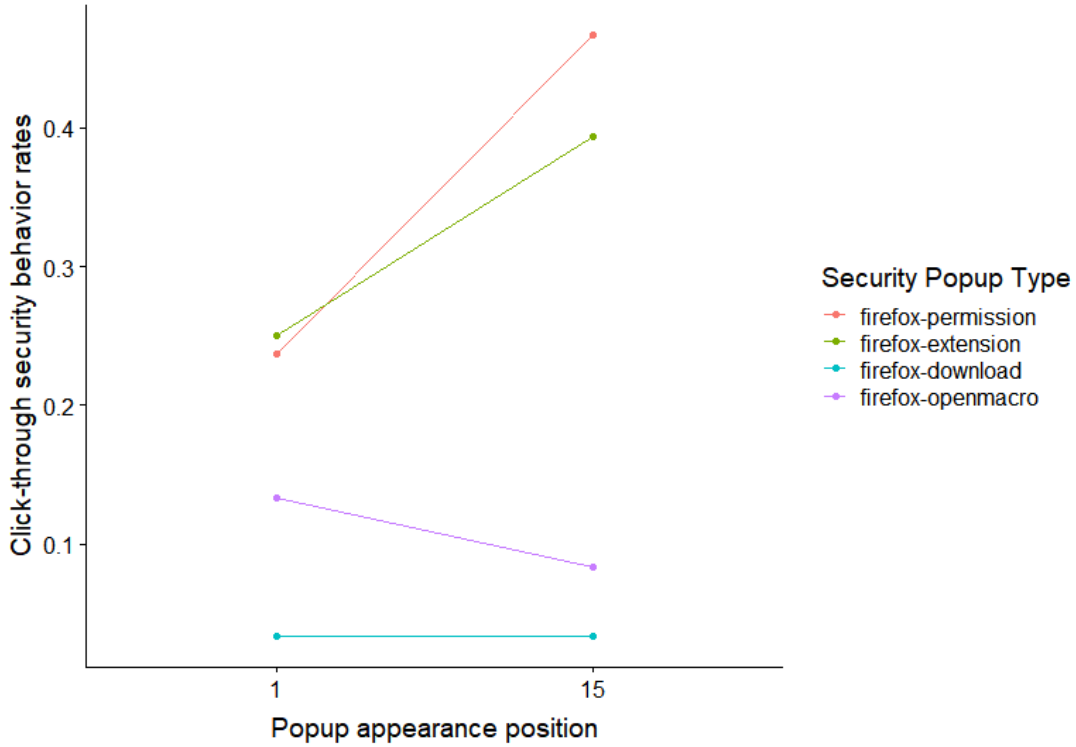


Figure 10: Adherence behavior at positions 1 and 15.

notifications is generalizing to the security notifications. Otherwise, there should be no significant difference. We specified a linear mixed model predicting movement speed by position. The type of warning was treated as a random effect. Position was treated as a fixed effect and was coded as 0 if the security notification was first, or 1 if the security warning occurred in position fifteen. The position significantly predicted speed: $t(449.004) = 5.471$, $p < .001$, conditional R^2: 0.231, supporting that generalization occurs (see Table 3).

To help ensure that the differences observed are due to generalization and not to fatigue, we specified a general linear model examining the influence of position on movement speed for the novel stimulus. In this analysis, position (1 vs. 15) did not influence how fast someone responded to the notification (see Table 4). This suggests that generalization rather than fatigue influenced movement speed.

Table 3: Mixed linear regression predicting speed (px/ms) by position.

	Estimate	Std. Error	*df*	*t*-value	*Pr*(>|*t*|)
(Intercept)	0.482	0.087	3.213	5.551	0.010
position	0.178	0.032	449.004	5.471	< 0.001

Table 4: Linear regression predicting speed (px/ms) based on position for novel stimulus.

	Estimate	Std. Error	*t*-value	*Pr*(>|*t*|)
(Intercept)	0.574	0.050	11.535	< 0.001
position	-0.100	0.071	-1.412	0.161

Finally, we examined whether the type of notification influenced the amount of generalization. To do this, we conducted a general linear model examining the interactions between the security warning types and position. Each warning type was coded as a dummy variable, leaving the performance notification as the baseline condition. Again, order was coded as a 0 if the notification was the first one shown. Otherwise, it was coded as a 1 if it was the fifteenth notification shown. The results are shown in Table 5. Although the main effects of warning type were significant, only the interactions (slope modifiers) for the extension warning and the permission warning with order were significant. These two types of warnings generalized less when compared to the non-security notification. The trends in speed for each notification type are shown in Figure 11. Again, the permission request and extension request are more visually similar to the performance notification than the macro and save executable warnings, these findings support that the similar look-and-feel of security warnings to other notifications may be magnifying generalization.

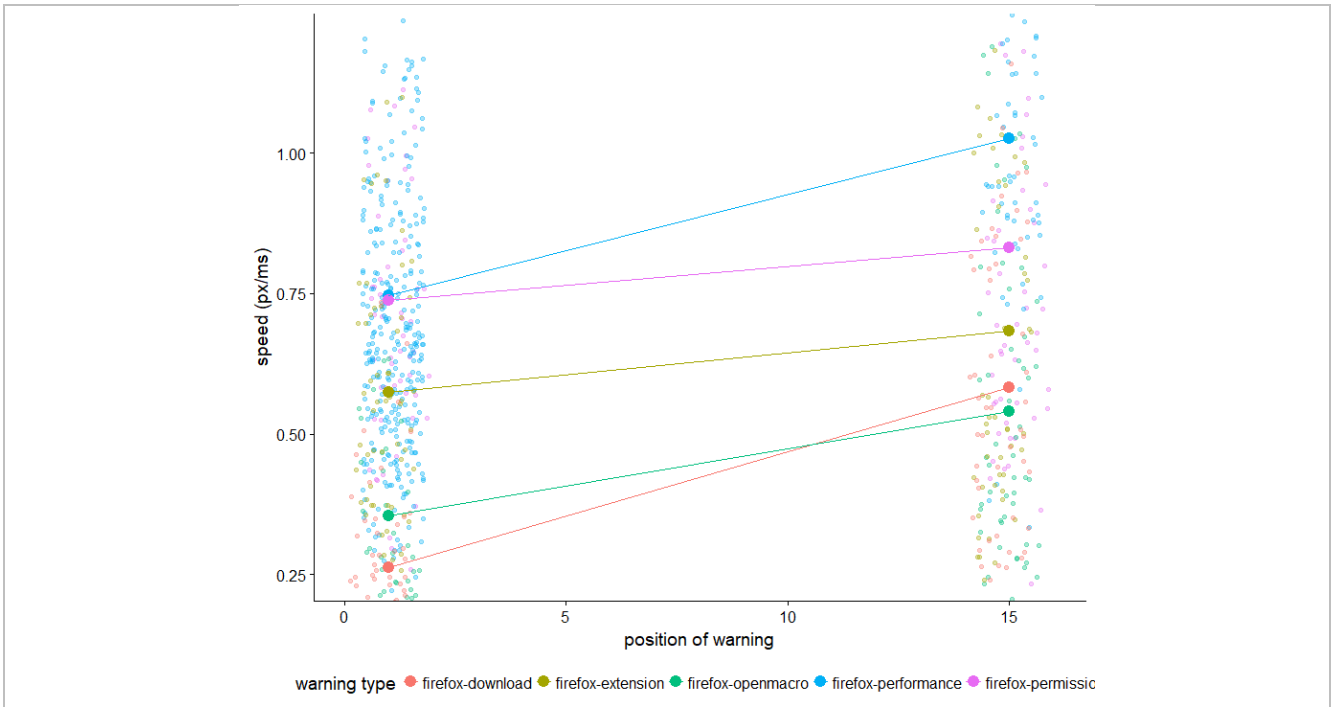


Figure 11: Movement speed (px/ms) trend for each warning / notification type.

Table 5: Linear regression predicting speed (pixel per millisecond) by interaction of security warning type by appearance position.

	Estimate	Std.Error	t-value	Pr(>|t|)
(Intercept)	0.727	0.020	36.795	0.000
Position	0.020	0.003	5.824	0.000
Extension warning	(0.161)	0.052	(3.097)	0.002
Save executable	(0.488)	0.051	(9.479)	0.000
Open macro	(0.387)	0.051	(7.511)	0.000
Permission warning	0.004	0.051	0.070	0.944
Position × extension	**(0.012)**	**0.006**	**(2.128)**	**0.034**
Position × executable	0.003	0.006	0.531	0.595
Position × open macro	(0.007)	0.006	(1.174)	0.241
Position × permission	**(0.013)**	**0.006**	**(2.338)**	**0.020**

Reaction Times

We induced a linear model to examine the impact of warning type and appearance position on user reaction times. All reaction times greater than 2.5 standard deviations from the median (median = 1,447 ms, SD = 2,732 ms) were flagged as outliers and were summarily ousted. The remaining reaction times were subjected to a linear regression model, wherein they were predicted by the interaction of modal position and modal types (dummy-coded) (see Table 6 and Figure 12). The slope for the novel stimulus between positions one and fifteen was not significantly different from 0 (β = -13.2, SE = 16.76, t = -0.79, p = 0.431). This supports the notion that fatigue was not at play over the course of the experimental task.

The slope for the performance notification was precipitous (see Figure 12), flattening out around four exposures, as would be expected given that this warning appeared often in the classification task. Interestingly, the drops in reaction time

Between positions one and fifteen for the permission and extension warnings were also negative, and statistically significantly so; (β = -80.27, SE = 16.93, t = -4.74, p = <0.001) and (β = -50.15, SE = 18.13, t = -2.77, p = 0.006) respectively. Because we have ruled out fatigue, we can infer that the negative slopes of the permission and extension warnings are indicative of generalization carrying over from the performance warning. However, these two warnings' slopes did not differ from one another β = 421.5763, SE = 491.7897 df = 532, t = 0.857, p = 0.3917, indicating that the rate of generalization was, while constant, nondiscriminatory. In contrast, the slopes for the save-executable and open-macro warnings were not different from zero; (β = -37.81, SE = 16.38, t = -2.31, p = 0.021) and (β = 4.29, SE = 16.38, t = 0.26, p = 0.793) respectively. This is consistent with the mouse cursor tracking results. Because the last two warnings which were were quite visually discrepant from the performance notification did not have statistically different reaction times between positions 15 and 1, and because the first two warnings which were quite visually similar to the

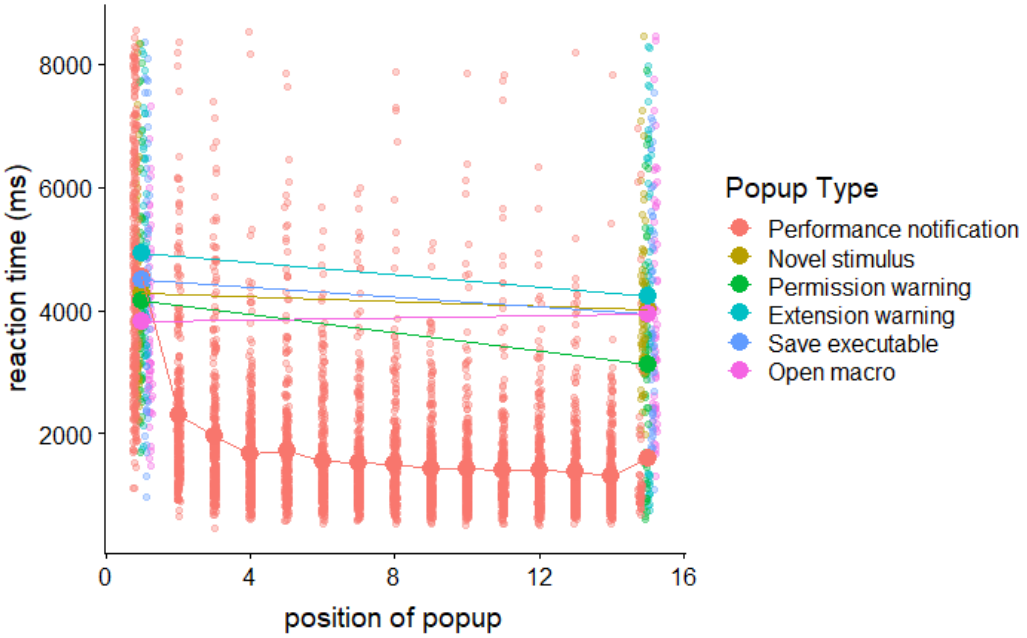


Figure 12. Reaction times for each warning at various positional appearances.

Table 6. Predicting reaction time by interaction of modal position and modal type. 0-intercept for ease of interpreting the slopes. Practical effects of slopes (ms reaction speeds at position 15) are obtained by multiplying the estimate by 15 and adding to the corresponding main effect.

	reaction time			
Predictors	*Estimates*	*std. Error*	*Statistic*	*p*
Performance	2720.52	36.87	73.79	<0.001
Novel stimulus	4240.49	181.31	23.39	<0.001
Permission warning	4257.72	184.81	23.04	<0.001
Extension warning	4970.84	196.76	25.26	<0.001
Save executable	4536.18	174.85	25.94	<0.001
Open macro	3850.46	174.85	22.02	<0.001
Performance × warning position	-124.13	4.26	-29.14	<0.001
novel stimulus × position	-13.20	16.76	-0.79	0.431
Permission warning × position	-80.27	16.93	-4.74	<0.001
Extension × position	-50.15	18.13	-2.77	0.006
Executable × position	-37.81	16.38	-2.31	0.021
Open macro × position	4.29	16.38	0.26	0.793
Observations	5586			
R^2 / adjusted R^2	0.757 / 0.756			

performance warning had statistically faster response times at position 15 than 1, these findings support the hypothesis that similar look-and-feel of security warnings to other notifications may be trigger generalization.

Survey responses

In a post-task survey (included in the appendix), participants reported the concern they felt when they encountered their assigned security warning. On the whole, participants reported anticipated levels of concern for the messages. Higher levels of concern were reported for security warnings, including the open-macro warning, permission warning, and save-executable warnings, whereas low concern was reported when seeing the novel stimulus or the performance notification. This pattern held for participants who saw the messages at either the first or the fifteenth position (see Figure 13).

We also asked participants for their preferred operating system, preferred web browser, whether they noticed seeing their assigned security message (a manipulation check), their general risk perceptions, and their information security threat severity and susceptibility perceptions. By and large, our participants preferred Windows (82.4%, *n*=551) over Mac (14.6%, *n*=98) or "other" (0.03%, *n*=20). Participants were neatly split between preferring Firefox and Chrome (48.7%, n=326 and 46.8%, n=313 respectively), with a sprinkling of other participants preferring Edge (n=7), Safari (n=13), Opera (n=6), or "other" (n=4). Participants in general reported above-average risk-taking attitudes (mean=5.61, SD=1.41), above-average perceptions of severity of a personal information security attack (mean=5.38, SD=1.47), yet lower perceptions of susceptibility to information security attacks (mean=4.16, SD=1.46) (each reported mean is an aggregate of three 7-

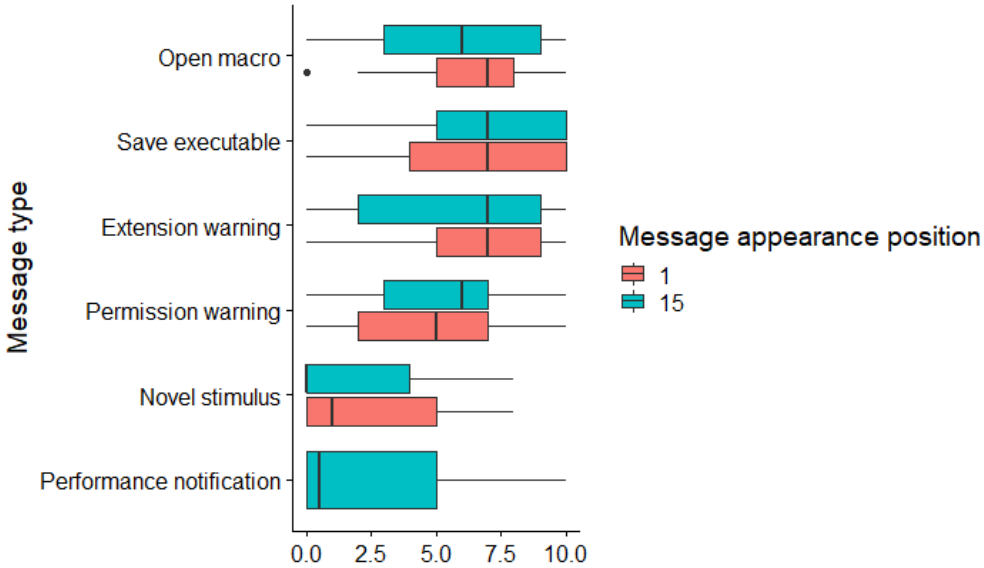


Figure 13. Concern for message (self-reported, scale of 0 to 10).

point Likert-scale agree-disagree survey items for each construct).

ANOVAs were performed for each survey construct individually to test whether responses were predictive of whether a security warning was clicked-through. The only significant overall ANOVA F-statistic was for the manipulation check ($F=28.997$, $p < .001$). A follow-up pairwise analysis with Tukey adjustment for each security warning grouped by appearance position suggested that participants who saw the extension install security warning at position 15 were more 16.28 times more likely (SD = 2.89) to have not noticed it than were participants who saw either the Open Macro security warning or the Save File security warning message at position 15 Also at position 15, participants approached being statistically more likely to have failed the manipulation check for the Extension security warning than for the Location Permission one (two-tailed $p = 0.064$). No pairwise comparison at position 1 was statistically significant. These findings provide some support for the notion that participants were less likely to notice (were more likely to have generalized habituation) to security warnings more visually similar to the performance notification after 14 exposures to the latter, than to less visually similar ones.

5. Discussion

This study contributes by showing the conditions under which generalization of habituation from routine notifications to security warnings occurs. Our paper does not claim to be the first to report the confusing of one warning with another [3, 15]. In contrast, our study specifically measures and tests the occurrence of generalization, and shows under what conditions it occurs.

Similarly, although our previous work has studied habituation in depth, we have not examined how habituation to one warning generalizes to another. Further, we know of no study besides the present study that investigates how habituation to a non-security-related notification can generalize to security warnings.

This paper (1) specifically examine how visual similarity leads to generalization, (2) test how habituation to a notification can generalize to different types of warnings, and (3) rule out the rival explanation of fatigue.

Specifically, we contribute by showing the following:

1. We provide empirical evidence that habituation to a frequent non-security-related notification does generalize to a one-time security warning.
2. We measure generalization in terms of (a) decreased attention to warnings, both in mouse cursor speed and response time; and (b) lower warning adherence behavior.
3. We show that this carry-over effect is due to generalization, and not fatigue. In past habituation literature, habituation and fatigue have been considered to be more or less synonymous (e.g., [1, 2]), but they are distinct phenomena with different implications. We show that participants ignored warnings not because they were tired, but because they had previously habituated to the performance notifications.
4. Finally, our results demonstrate that not all security warnings are equal in terms of the amount of generalization of habituation. Our results indicate that the more similar the security warning is to the non-security warnings in terms of "look and feel", the greater the degree of generalization. This finding questions whether corporate efforts to create a consistent UI look and feel is promoting better security or inhibiting security.

These insights open new avenues of research, pointing the way for researchers and practitioners to develop and test security warning designs that are resistant to generalization by distinguishing the appearance of security warnings from common notifications.

6. Limitations and Future Research

Our research was subject to several limitations. First, this research examines how similarity of appearance between notifications and security warnings can lead to the occurrence of generalization. Future research can additionally examine whether changing the mode of interaction for security warnings from the common "click to dismiss" paradigm can also reduce generalization.

Second, our experiment was designed to expose participants to notifications at a higher rate than is normally encountered in the same amount of time during usual computer usage. In future research, it would be interesting to explore if generalization of habituation occurs with the same amount of exposures distributed across a longer time window. However, participants' exposure to up to 15 notifications during the experimental session is not that far off from the number of notifications reported in observational studies [26]. Similarly, although the warning messages were meant to appear as if they were triggered by the website for each image, some messages (e.g., the save executable message) may have appeared incongruent for the experimental task. Consequently, some users may have been more dismissive than if the warning message better matched the task context.

Finally, while we explicitly controlled for fatigue in our experimental design, there are other factors that could have affected the speed and accuracy of participants' responses in our task. For example, participants could have become more engrossed in the task over time and therefore been quicker to dismiss notifications and less accurate at responding to warnings. Alternatively, faster responding may have been

due to participants learning about the task (e.g., which locations to click and when). For this reason, future work will be needed to tease out these alternative explanations. While habituation is a type of learning, it involves different low-level neural mechanisms than higher-order skill learning processes. Because habituation is fundamentally a neurobiological phenomenon, neurophysiological tools such as EEG or fMRI, may be especially useful to tease out these alternative explanations.

7. Conclusion

Generalization of habituation is a serious problem because it may cause users to tune out important security notifications, even if it is the first time any particular notification is displayed. However, an awareness of this problem can encourage software developers to create visually novel notifications that will receive the requisite attention to facilitate users' adherence to security warnings.

ACKNOWLEDGMENTS

This was supported by the National Science Foundation under Grant CNS-1931108.

REFERENCES

[1] M.E. Acer, E. Stark, A.P. Felt, S. Fahl, R. Bhargava, B. Dev, M. Braithwaite, R. Sleevi and P. Tabriz. 2017. Where the Wild Warnings Are: Root Causes of Chrome HTTPS Certificate Errors. in *ACM Conference on Computer and Communications Security (CCS)*, Dallas, TX.

[2] D. Akhawe and A.P. Felt. 2013. Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness *Proceedings of the 22nd USENIX Conference on Security*, USENIX Association, Washington, D.C., 257-272.

[3] T.S. Amer and J.-M.B. Maris. (2007). Signal words and signal icons in application control and information technology exception messages—Hazard matching and habituation effects. *Journal of Information Systems*, 21 (2). 1-25.

[4] B.B. Anderson, J. Jenkins, A. Vance, C.B. Kirwan and D. Eargle. (2016). Your Memory is Working Against You: How Eye Tracking and Memory Explain Susceptibility to Phishing. *Decision Support Systems*, 92. 3-13.

[5] B.B. Anderson, C.B. Kirwan, J.L. Jenkins, D. Eargle, S. Howard and A. Vance. 2015. How Polymorphic Warnings Reduce Habituation in the Brain: Insights from an fMRI Study *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, Seoul, Republic of Korea, 2883-2892.

[6] B.B. Anderson, C.B. Kirwan, J.L. Jenkins, D. Eargle, S. Howard and A. Vance. 2015. How Polymorphic Warnings Reduce Habituation in the Brain: Insights from an fMRI Study. in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, Seoul, South Korea, ACM.

[7] B.B. Anderson, A. Vance, C.B. Kirwan, J. Jenkins and D. Eargle. (2016). From Warnings to Wallpaper: Why the Brain Habituates to Security Warnings and What Can Be Done about It. *Journal of Management Information Systems*, 33 (3). 713-743.

[8] Apple.com. 2017. https://developer.apple.com/design/.

[9] R. Böhme and S. Köpsell. 2010. Trained to Accept?: A Field Experiment on Consent Dialogs. in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)* Atlanta, ACM, 2403-2406. 10.1145/1753326.1753689

[10] C. Bravo-Lillo, L. Cranor, S. Komanduri, S. Schechter and M. Sleeper. 2014. Harder to Ignore? Revisiting Pop-Up Fatigue and Approaches to Prevent It. in *10th Symposium On Usable Privacy and Security (SOUPS 2014)*, USENIX Association, 105-111.

[11] C. Bravo-Lillo, S. Komanduri, L.F. Cranor, R.W. Reeder, M. Sleeper, J. Downs and S. Schechter. 2013. Your Attention Please: Designing Security-decision UIs to Make Genuine Risks Harder to Ignore *Proceedings of the Ninth Symposium on Usable Privacy and Security* ACM, Newcastle, United Kingdom, 1-12.

[12] J.C. Brustoloni and R. Villamarín-Salomón. 2007. Improving Security Decisions with Polymorphic and Audited Dialogs. in *Proceedings of the Third Symposium on Usable Privacy and Security (SOUPS 2007)*, New York, NY, USA, ACM, 76-85.

[13] A. Byers and J.T. Serences. (2012). Exploring the relationship between perceptual learning and top-down attentional contro. *Vision Research*, 74. 30-39.

[14] A. Cooper, R. Reinmann and D. Cronin. 2007. *About Face 3: The Essentials of Interaction Design*. Wiley, Indianapolis, IN.

[15] S. Egelman, L.F. Cranor and J. Hong. 2008. You've Been Warned: An Empirical Study of the Effectiveness of Web Browser Phishing Warnings *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Florence, Italy, 1065-1074.

[16] R.L. Goldstone. (1998). Perceptual Learning. *Annual Review of Psychology*, 49 (1). 585-612.

[17] C. Green and D. Bavelier. (2003). Action video game modifies visual selective attention. *Nature*, 423. 534-537.

[18] M. Harbach, M. Hettig, S. Weber and M. Smith. 2014. Using personal examples to improve risk communication for security & privacy decisions *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM, Toronto, Ontario, Canada, 2647-2656.

[19] M.J. Kalsher and K.J. Williams. 2006. Behavioral Compliance: Theory, Methodology, and Results in Wogalter, M. ed. *The Handbook of Warnings*, CRC Press, 313.

[20] C.B. Kirwan and C.E.L. Stark. (2007). Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learning & Memory*, 14 (9). 625-633.
[21] S. Krug. 2014. *Don't Make Me Think, Revisited: A Common Sense Approach to Web and Mobile Usability*. New Riders.
[22] Microsoft. 2017. https://www.microsoft.com/en-us/design.
[23] C.H. Rankin, T. Abrams, R.J. Barry, S. Bhatnagar, D.F. Clayton, J. Colombo, G. Coppola, M.A. Geyer, D.L. Glanzman, S. Marsland, F.K. McSweeney, D.A. Wilson, C.-F. Wu and R.F. Thompson. (2009). Habituation revisited: An updated and revised description of the behavioral characteristics of habituation. *Neurobiology of Learning and Memory*, 92 (2). 135-138. http://dx.doi.org/10.1016/j.nlm.2008.09.012
[24] R.W. Reeder, A.P. Felt, S. Consolvo, N. Malkin, C. Thompson and S. Egelman. 2018. An Experience Sampling Study of User Reactions to Browser Warnings in the Field *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal QC, Canada, 1-13.
[25] F. Schaub, R. Balebako, A.L. Durity and L.F. Cranor. (2015). A Design Space for Effective Privacy Notices. *To appear in the*.
[26] A.S. Shirazi, N. Henze, T. Dingler, M. Pielot, D. Weber and A. Schmidt. 2014. Large-scale assessment of mobile notifications. . in *SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*, Toronto, Ontario, Canada, ACM, 3055-3064. 10.1145/2556288.2557189
[27] A. Sotirakopoulos, K. Hawkey and K. Beznosov. 2011. On the Challenges in Usable Security Lab Studies: Lessons Learned From Replicating a Study on SSL Warnings *Proceedings of the Seventh Symposium on Usable Privacy and Security (SOUPS)*, ACM, Menlo Park, CA, 3:1-3:18.
[28] Z.R. Steelman, B.I. Hammer and M. Limayem. (2014). Data collection in the digital age: innovative alternatives to student samples. *MIS Quarterly*, 38 (2). 355-378.
[29] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri and L.F. Cranor. 2009. Crying Wolf: An Empirical Study of SSL Warning Effectiveness. in *SSYM'09 Proceedings of the 18th Conference on USENIX Security Symposium*, Montreal, Canada, 399-416.
[30] Symantec. 2017. Internet Security Threat Report.
[31] R.F. Thompson and W.A. Spencer. (1966). Habituation: A Model Phenomenon for the Study of Neuronal Substrates of Behavior. *Psychological Review*, 73 (1). 16-43.
[32] A. Vance, B. Brinton Anderson, C. Brock Kirwan and D. Eargle. (2014). Using Measures of Risk Perception to Predict Information Security Behavior: Insights from Electroencephalography (EEG). *Journal of the Association for Information Systems*, 15 (10). 679-722.
[33] A. Vance, J. Jenkins, B. Anderson, D. Bjornn and B. Kirwan. (2018). Tuning Out Security Warnings: A Longitudinal Examination of Habituation through fMRI, Eye Tracking, and Field Experiments. *MIS Quarterly*, 42 (2). 355-380.
[34] A. Vance, C.B. Kirwan, D. Bjornn, J.L. Jenkins and B.B. Anderson. 2017. What Do We Really Know about How Habituation to Warnings Occurs Over Time? A Longitudinal fMRI Study of Habituation and Polymorphic Warnings *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, Denver, CO.
[35] R. Wash, E. Rader and C. Fennell. 2017. Can People Self-Report Security Accurately?: Agreement Between Self-Report and Behavioral Measures *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver, CO, USA, http://dx.doi.org/10.1145/3025453.3025911.
[36] J. Weinberger and A.P. Felt. 2016. A week to remember: The impact of browser warning storage policies. in *Symposium on Usable Privacy and Security (SOUPS)*, Denver.
[37] R. West. (2008). The Psychology of Security. *Communications of the ACM*, 51 (4). 34-40.
[38] M.S. Wogalter. 2006. Communication-Human Information Processing (C-HIP) Model. in Wogalter, M.S. ed. *Handbook of Warnings*, Lawrence Erlbaum Associates, Mahwah, NJ, 51-61.

Appendix A – Post-task Survey

Please select your gender:

- Female
- Male
- Other

Please enter your age: ____

Please select your preferred OS:

- Mac
- Windows
- Other

Please select your preferred browser:

- Chrome
- Edge
- Firefox
- Opera
- Safari
- Other

Presentation order for the following items was randomized. All items in this section allowed respondents to choose from the following Likert-scale options:

- *1-Strongly disagree (1)*
- *2-Moderately disagree (2)*
- *3-Mildly disagree (3)*
- *4-Neutral (4)*
- *5-Mildly agree (5)*
- *6-Moderately agree (6)*
- *7-Strongly agree (7)*

[RISK1] Ignoring malware warning screens can cause damages to computer security.

[TSUS1] My computer is at risk for becoming infected with malware.

[RISK2] Ignoring malware warning screens can put important data at risk.

[TSUS2] It is likely that my computer will become infected with malware.

[TSUS3] It is possible that my computer will become infected with malware.

[RISK3] Ignoring malware warning screens will most likely cause security breaches.

[TSEV1] If my computer were infected by malware, it would be severe.

[TSEV2] If my computer were infected by malware, it would be serious.

[TSEV3] If my computer were infected by malware, it would be significant.

[attention check] Select "3-mildly disagree" for this answer (attention).

The following questions appeared at the end of the survey:

[manipulation_check] Did you notice the following popup during the Batman image classification task?

[Yes / No/ I'm not sure]

[realism] On a scale of 0 to 10, how realistic do you think the following message is? [participants were shown a screenshot of the security notification for their treatment group]

[0-Not realistic (1) ... 10-100% realistic (11)]

[concern] On a scale of 0 to 10, how concerned did the following screen make you feel during the Batman image classification task? [participants were shown a screenshot of the security notification for their treatment group]

[0-Not concerned at all (1) ... 10-Extremely concerned (11)]

[debrief] The primary objective of this study was to observe how you responded to browser popups. You were randomly assigned to a condition in which you saw a variant of a browser popup. Additionally, the browser popups you saw were simulated. Your response to them will have no impact on your browser or computer.

[free_response] Any feedback for the research team?

[free response]