The Fog of Warnings:
How Non-essential Notifications Diminish Security Warning Efficacy

Anthony Vance
Business Information Technology
Pamplin College of Business
Virginia Tech
anthony@vance.name

David Eargle
Carve Systems
dave@daveeargle.com

C. Brock Kirwan
Department of Psychology and Neuroscience Center
Brigham Young University
kirwan@byu.edu

Bonnie Brinton Anderson
Information Systems Department
Marriott School of Business
Brigham Young University
bonnie_anderson@byu.edu

Jeffrey L. Jenkins
Information Systems Department
Marriott School of Business
Brigham Young University
jeffrey_jenkins@byu.edu

The Fog of Warnings:
How Non-essential Notifications Diminish Security Warning Efficacy

ABSTRACT

Internet users' disregard of security warnings remains an important problem in cybersecurity. Previous research has shown that a key contributor of warning disregard is habituation, which is the decreased response to a repeated stimulus. However, this problem is broader than previously recognized because of the neurobiological phenomenon of generalization of habituation. Generalization occurs when habituation to one stimulus extends to other novel stimuli that share similar characteristics. Therefore, through frequent exposure to non-security-related notifications, users may become habituated to critical security warnings that they have never encountered before. However, the extent of this problem is unknown, and how to mitigate it is unclear.

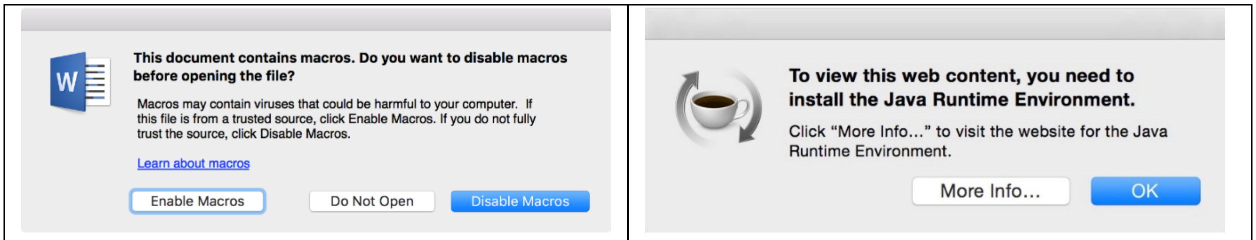
In this study, we addressed this problem in three experiments. First, we conducted a field experiment to demonstrate how generalization of habituation occurs and can be mitigated by differentiating the visual appearance of the warning. Second, we used functional magnetic resonance imaging (fMRI) to rule out rival hypotheses and demonstrate that changing the mode of interaction of the warning also mitigates generalization. Finally, we conducted an additional field experiment to demonstrate that changing the mode of interaction while holding visual appearance constant reduces generalization. These findings provide guidance for creating warnings that are resistant to generalization of habituation and promote greater warning adherence.

Keywords: security warning, habituation, generalization, fMRI, field experiment, NeuroIS

The Fog of Warnings:
How Non-essential Notifications Diminish Security Warning Efficacy

INTRODUCTION

Internet users' disregard of security warnings continues to be an important problem in cybersecurity because warnings are often the last defense before a person or organization is compromised (Acer et al., 2017; Weinberger & Felt, 2016). For example, numerous significant breaches and ransomware incidents can be traced back to a user opening a malicious macro-enabled Microsoft Office email attachment and disregarding the security warning about the danger of macros (see Figure 1a; CISA, 2022; Mandiant, 2021; Verizon, 2022).[1] Previous research has shown that a key contributor to security warning disregard is habituation (Anderson, Vance, et al. 2016), which is "decreased response to repeated stimulation" (Groves & Thompson, 1970, p. 419). As a user is repeatedly exposed to a warning, the brain neurobiologically pays increasingly less attention to it, which leads to warning disregard behavior (Weinert et al., 2022; Vance et al., 2018).



This document contains macros. Do you want to disable macros before opening the file?

Macros may contain viruses that could be harmful to your computer. If this file is from a trusted source, click Enable Macros. If you do not fully trust the source, click Disable Macros.

To view this web content, you need to install the Java Runtime Environment.

Click "More Info…" to visit the website for the Java Runtime Environment.

(a) Microsoft Office macro security warning

(b) System-generated notification

Figure 1. A software notification (left) and security warning (right) showing similarities in visual appearance and mode of interaction

However, the problem of habituation to security warnings is likely much larger in scope than previously recognized. Habituation theory from neurobiology explains that organisms

[1] In recognition of the persistent problem of security warning disregard, in 2022, Microsoft announced plans to gradually disable macros in Office applications by default (Eickmeyer, 2022).

habituate not only to an individual stimulus but also collectively to stimuli that share similar characteristics (Rankin et al., 2009; Thompson & Spencer, 1966). This phenomenon, known as *stimulus generalization* or simply *generalization*, has important implications for security warnings. This is because a foundational principle of user interface (UI) design is consistency (Cooper et al., 2007; Krug, 2014), which is reinforced by major software companies that provide development libraries and guidelines (Apple, 2022; Google, 2022; Microsoft, 2022). Generalization suggests that users habituate not only to individual UI notifications (e.g., dialogs, alerts, and confirmations; hereinafter referred to collectively as "notifications" for brevity) but also to whole classes of notifications that share similar characteristics. As a result, users may already be deeply habituated to a security warning that they have never seen before simply because it is similar to frequently seen notifications (compare Figure 1a and 1b).

Furthermore, the generalization problem will likely worsen as users are increasingly exposed to more notifications. For example, mobile phone users receive an average of 218 notifications per day (Turner et al., 2019). As notifications multiply 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. In such a saturated environment, it is crucial that habituation to notifications does not generalize to security warnings in order to protect individuals and organizations.

Although the problem of the blurring of security warnings and non-security-related notifications has previously been speculated (e.g., Böhme & Köpsell, 2010; Vance et al., 2018; West, 2008), it has not been empirically examined. Therefore, the scope and severity of generalization of habituation to security warnings and the conditions under which it occurs are unknown. Furthermore, solutions to habituation to warnings offered by previous research are insufficient to address the problem of generalization. For example, the common recommendation to reduce habituation is to substantially limit the appearance of warnings (Acer et al., 2017; Akhawe & Felt, 2013; Krol et al., 2012; Zeng et al., 2019). However, this approach will not work if habituation to pervasive non-security-related notifications carries over to rare security warnings. Therefore, theoretically and empirically grounded mitigations that take into account the problem of generalization are needed.

Given this need, the objectives of this study were (1) to explain how habituation generalizes from frequent notifications to security warnings and (2) to evaluate ways to mitigate this problem. In pursuing these objectives, we answer the following research questions:

RQ1: How does habituation to frequent non-security-related notifications generalize to infrequent security warnings?

RQ2: How can generalization to security warnings be reduced?

We addressed these questions in three experiments. First, we conducted a field experiment to demonstrate that habituation to a frequent non-security-related notification generalizes to a novel security warning in terms of both decreased attention and higher warning disregard behavior. Furthermore, we showed that changing the visual appearance of the warning to make it less similar to the non-security-related notification made generalization substantially less likely to occur.

Second, we performed a functional magnetic resonance imaging (fMRI) experiment to demonstrate that generalization occurs neurobiologically in the brain, allowing us to rule out rival explanations of fatigue or cognitive engagement. We also showed that generalization can be reduced by changing the mode of interaction of the warning. We define mode of interaction as the way by which a person interacts with a user interface (Reeves et al., 2004), such as using a mouse to click one of several options in a notification message (e.g., “OK” and “Cancel”).

Warnings with a distinctive mode of interaction (e.g., using the mouse to drag a slider widget to select an option) showed significantly less generalization.

Third, we performed an additional field experiment to disentangle the effects of changing the mode of interaction and visual appearance. While holding visual appearance constant, we found that changing the mode of interaction alone was effective in reducing generalization. The triangulated findings from these three experiments open new avenues of research and provide guidance to software developers for creating warnings that are resistant to the effects of generalization, thereby reducing users' security warning disregard behavior.

LITERATURE REVIEW, THEORY, AND HYPOTHESES

Generalization and Security Warnings

Habituation to security warnings is well known and has been observed or inferred in several studies (see Amran et al. 2018 and Vance et al. 2018 for reviews). However, the phenomenon of generalization is less well recognized. West (2008) noted that "security messages often resemble other messages dialogs. As a result, security warnings may not stand out in importance and users often learn to disregard them" (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)" (2010, p. 2406). However, this effect was not empirically examined in previous studies.

Empirically examining generalization is needed because the recommendations of previous research that examined security warning habituation may not apply in contexts where generalization occurs. Chief among these is the consensus of previous studies to limit the appearance of warnings such as "Reducing the onslaught of interrupting security warning dialogs might help reduce the strain on users' attention" (Bravo-Lillo et al., 2013, p. 1; Acer et al., 2017;

Akhawe & Felt, 2013; Brustoloni & Villamarín-Salomón, 2007; Krol et al., 2012; Weinberger & Felt, 2016; Zeng et al., 2019). Though reasonable, this guidance does not consider the impact of non-security-related notifications that share similar characteristics with warnings and can appear literally hundreds of times throughout a user's day (Pielot et al., 2018; Shirazi et al., 2014; Turner et al., 2019). Consequently, limiting the appearance of warnings in such contexts will have little effect in the presence of generalization.

Another recommendation of previous studies is to update the appearance of a warning with each presentation through a "polymorphic" design, which can substantially reduce habituation to a single warning (Anderson, Jenkins, et al., 2016; Anderson, Vance, et al., 2016; Vance et al., 2018). However, it is unclear whether a polymorphic warning is robust to generalization of habituation caused by frequent notifications. While some variations of the polymorphic design may be sufficiently different in appearance from common notifications, others may be too subtle to resist the effect of generalization (Anderson, Vance, et al., 2016; Vance et al., 2018). More fundamentally, the conditions under which generalization occurs in the context of notifications and warnings are not understood; therefore, the interventions that are most effective in reducing generalization are unknown. For these reasons, research to investigate how "other facets of habituation may influence the users' responses to security warnings, such as […] generalization" (Vance et al., 2018, p. 376) is called for. We address these gaps in the literature in this research.

Generalization Theory and Hypotheses

Generalization can be understood as the failure to distinguish different stimuli (Shepard, 1987). According to the dual-process theory of habituation (Groves & Thompson, 1970), when a person is initially exposed to a stimulus, a mental model of the stimulus is created in memory. When the same stimulus is encountered again, a person will automatically and unconsciously compare the stimulus to the mental model. If the stimulus and mental model match, behavioral responses to the stimulus are inhibited in favor of reliance on the model in memory instead (Thompson, 2009). However, when a novel stimulus that does not match the mental model is encountered, inhibited behavioral responses will recover. This process of habituation allows the brain to efficiently conserve resources to filter out irrelevant stimuli and respond to new stimuli encountered.

However, neuroscience literature has shown that this increased efficiency of the neural response comes at the price of failing to distinguish stimuli that share similar characteristics, so an inhibited response to one stimulus carries over or generalizes to another novel but similar stimulus (Thompson & Spencer, 1966). The degree to which the brain relies on an existing mental model, instead of giving a novel stimulus full attention, depends on how frequently the person has been exposed to the stimulus in the past and how similar the novel stimulus is to existing models (Rankin et al., 2009). Therefore, the stronger the habituation and the similarity of two stimuli, the more likely that habituation will generalize from one stimulus to another.

Generalization has been studied in many domains outside of neuroscience, including image recognition in toddlers (Anderson et al., 2022); attention and vigilance responses after watching missing person alert videos across several days (Lampinen & Moore, 2016); cues to start and stop eating (Epstein et al., 2022), emotional desensitization to witnessing or participating in video game violence (Bushman & Anderson, 2009; Mrug et al., 2008; Mrug et al., 2016; Mrug & Windle, 2010; Rosenfield et al., 2014); dampened arousal, blood pressure, and heart rate (Grizzard et al., 2015); and guilt responses from playing unethical video games (Grizzard et al., 2017). However, little research has examined generalization in the context of UI design, and few if any studies have investigated generalization in the context of notifications and security warnings.

Following the dual-process theory of habituation, we hypothesized that habituation to notifications will generalize to security warnings that share a similar visual appearance. Maintaining visual consistency has long been considered a fundamental principle of UI design (Curtis, 1989). Hence, security warnings are typically visually similar to non-security-related notifications (see Figure 1). The greater the visual similarity between a security warning and a notification, the greater will be the reliance on the mental model of the previously seen notification and the less will be the response to the security warning, both in terms of attention and behavioral response. When evaluating a stimulus, people often notice the big-picture elements (e.g., visual appearance) before diving into small details (e.g., text; Djamasbi et al., 2011). Thus, even though some information may differ across security warnings and notifications (e.g., text and icons), visual similarity and other high-level information will increase reliance on model in memory. In summary, we predict:

H1: Habituation to non-security-related notifications will generalize to security warnings with a similar visual appearance.

The foregoing theory and hypothesis also allow for a corollary prediction. According to dual-process theory, generalization will not occur for a novel stimulus that is sufficiently distinct from an existing model in memory (Thompson, 2009). In such cases, the brain will recognize the new stimulus as novel and will process it with full attentional resources (Rankin et al., 2009). Therefore, if a novel security warning is dissimilar to the mental model of a notification, people will be more likely to pay attention to the warning, and generalization will occur less. Accordingly, we hypothesized as follows:

H2: Habituation to non-security-related notifications will generalize less to novel security warnings with a distinctive visual appearance.

We also theorized that differentiating the mode of interaction of a warning from common notifications will reduce generalization. According to schema theory (Rumelhart, 1980), schemas can represent general knowledge about objects, situations, or sequences of actions. This phenomenon is often explained by the concept of muscle memory. Muscle memory refers to a type of procedural memory, that is, a type of unconscious memory that aids in performing repeated tasks efficiently, that occurs when a person performs a specific motor (e.g., muscle movement) task repeatedly. Through repeated performance of a given motor task, long-term muscle memory is created for that task, which can then be performed more exactly and efficiently in the future (Krakauer & Shadmehr, 2006). For example, many types of athletes repeatedly practice an activity so that they can perform it well without conscious effort (Adkins & Boychuck, 2006).

Likewise, in computing tasks, by far the dominant UI paradigm for responding to a notification is the use of a mouse to click one of several labeled buttons (e.g., "OK" and "Cancel"; West, 2008), a paradigm we call "click to dismiss." As a result of this paradigm dominance, people often habitually respond to a message due to muscle memory, even though it sometimes contradicts their conscious intentions (Bravo-Lillo et al., 2013). Thus, similar to visual habituation, people may also form high-level memory representations (or schemas) for *how* to interact with notifications and warnings.

However, schema theory also explains that changing the schema forces conscious engagement with a task (Rumelhart, 1980). Thus, a person who can kick a soccer ball proficiently without conscious effort may not necessarily be able to hit a golf ball with the same level of automaticity. Rather, this would require attention and purposeful movements (James et al., 2013). Similarly, a person who can play a scale on a violin without conscious thought of finger placement may not necessarily be able to automatically do the same on a cello, given the differences between the instruments and the way they are held. Applied to the context of security warnings, changing the standard mode of interaction from “click to dismiss” to something distinctive (e.g., using the mouse to drag a slider widget to select an option) will break the schema of past responses to notifications and force people to respond consciously without relying on muscle memory (Rumelhart, 1980). Thus, people will not rely on the existing mental model, avoid automatically responding, and pay the warning more attention. Therefore, we hypothesized as follows:

H3: *Habituation to non-security-related notifications will generalize less to novel security warnings with a distinctive mode of interaction.*

METHODS

We tested our hypotheses through three complementary experiments. In Experiment 1, we conducted an online field experiment to test H1 and H2 by observing participants’ behaviors. In Experiment 2, we performed an fMRI laboratory experiment to test H1 via neural activation and to rule out rival explanations of fatigue and increasing cognitive engagement. We also tested H3 by showing that warnings with different modes of interaction from a notification exhibit less generalization of habituation. Finally, in Experiment 3, we performed an additional online field experiment to test H3 again, this time testing the effect of changing the mode of interaction of a warning while holding visual appearance constant. Table 1 summarizes each experiment.

Table 1. Overview of the Experiments

Experiment #	Purposes (hypotheses tested)	Method (design)	Participant pool (*n*)	Dependent variables
1	1. Establish whether generalization occurs (H1) 2. Test whether changing the visual appearance of a warning can reduce generalization (H2)	Online field experiment (between subjects)	Amazon Mechanical Turk (609)	Warning disregard and reaction time
2	1. Rule out rival explanations to generalization (H1) 2. Test whether changing the mode of interaction of a warning reduces generalization (H3)	fMRI laboratory experiment (within subjects)	University subject pool (25)	Neural activation in response to notifications and warnings
3	1. Isolate the effect of changing the mode of interaction of a warning by holding visual appearance constant (H3)	Online field experiment (between subjects)	Amazon Mechanical Turk (234)	Warning disregard and reaction times

EXPERIMENT 1: ONLINE FIELD EXPERIMENT

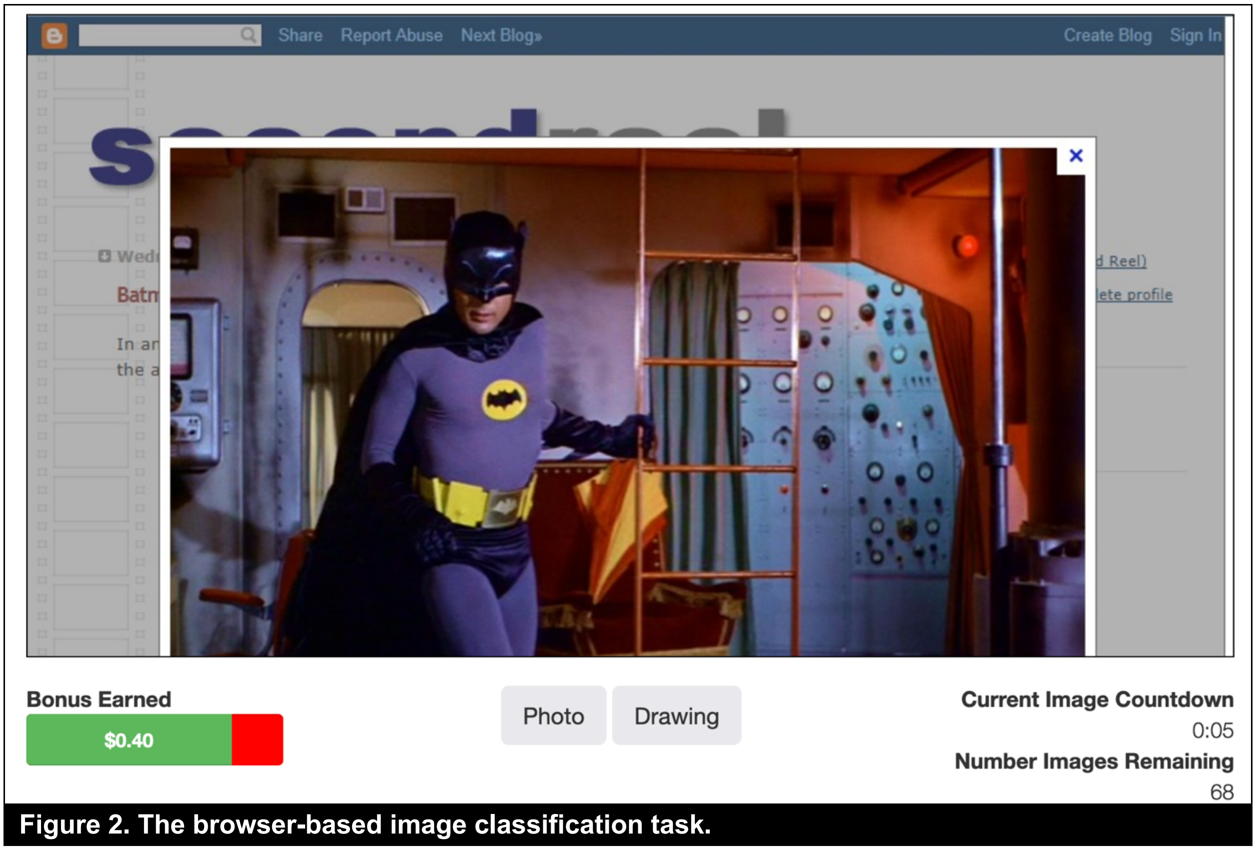
We designed an online experiment to establish whether habituation generalizes from notifications to security warnings (H1) and whether distinguishing the visual appearance of the warning can reduce this generalization (H2). We tested these hypotheses in the context of a web browser-based task. This had the advantage of allowing us to observe the effects of repeated browser notifications on various actual browser warnings. It also had the advantage of observing participants in their natural environment and on their own computers, making possible perceptions of risk to their devices and data.

Experiment 1: Participants

We recruited 609 participants via Amazon Mechanical Turk (mTurk). Following Steelman et al. (2014), all participants were required to be based in the United States. The average age of participants was 36 years. Of the participants, 53% were male. The participants were ultimately paid $1.50 ($1.00 up front and $0.50 bonus) USD for an approximately 5-minute task. To standardize the notification and warning messages, the participants were required to use the Firefox browser.[2]

Experiment 1: Experimental Task

We followed a previously established experimental protocol designed to measure participants' disregard of security warnings in a web browser (Vance et al., 2014). However, with institutional review board approval, deception was used to conceal this purpose from the participants. Instead, the participants were instructed that the purpose of the experiment was to train a machine learning algorithm by manually classifying images from the web, a task common on Amazon Mechanical Turk (MTurk, 2017). Participants were required to view images of Batman and classify them as either photographs or drawings (Figure 2).



Bonus Earned
$0.40

Current Image Countdown
0:05

Number Images Remaining
68

Figure 2. The browser-based image classification task.

[2] This was enforced by verifying that "Firefox" was listed in the user agent string of the HTTP header. Participants could not proceed with the mTurk human intelligence task (HIT) until they used the Firefox browser.

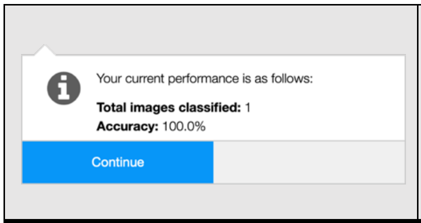
Typically, because security warnings and notifications must compete with time-sensitive work tasks in the real world (Jenkins et al., 2016), we simulated time pressure by giving participants 10 seconds to classify each website image. Failure to classify the image was counted as an incorrect answer. Participants started with a bonus of $0.50, which decreased by $0.05 with each incorrect classification. A status bar in the bottom-left corner of the screen showed participants their current bonus (see Figure 2). However, in reality, all participants received the full bonus regardless of their performance at the end of the task.

After a short practice round, participants began classifying what appeared to be a total of 75 images. In actuality, they classified at most 14 images to avoid participants biasing their behavior by anticipating the end of the experiment. Each classification was followed by the display of an HTML5-style browser notification in the upper-left side of the screen, which reported the participant's current classification performance (see Figure 3a). Importantly, the participants had to click a "Continue" button on the performance notification before going on to the next image. In this way, participants were naturally and repeatedly exposed to the performance notification throughout the course of the experimental task.

Experiment 1: Experimental Design

Warning Treatments

We selected four existing Firefox browser warnings, from which one was randomly selected to be shown to a participant during the experimental task (Figure 3). The permission request warning (Figure 3b) and browser extension install warning (Figure 3c) were selected because of their similarity to Firefox's HTML notification (Figure 3a). Conversely, we also selected two other warnings, a save executable warning (Figure 3d) and an open macro for a macro-enabled spreadsheet (Figure 3e) warning, because of their visual differences from the performance notification.[3]

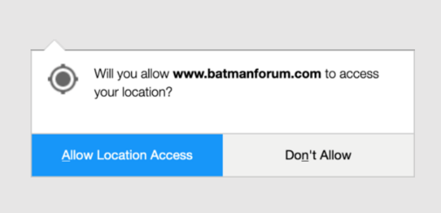


Your current performance is as follows:

Total images classified: 1
Accuracy: 100.0%

Continue

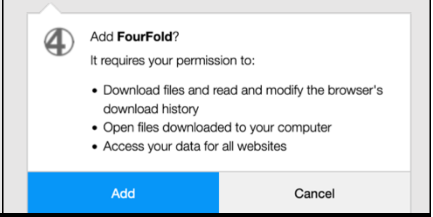
Figure 3a. Performance notification.



Will you allow **www.batmanforum.com** to access your location?

Allow Location Access | Don't Allow

Figure 3b. Permission warning and location request.



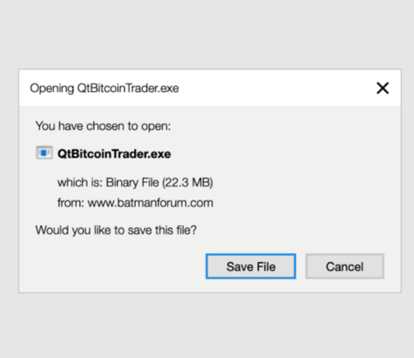
Add **FourFold**?

It requires your permission to:

- Download files and read and modify the browser's download history
- Open files downloaded to your computer
- Access your data for all websites

Add | Cancel

Figure 3c. Firefox add-on permission warning.



Opening QtBitcoinTrader.exe ✕

You have chosen to open:

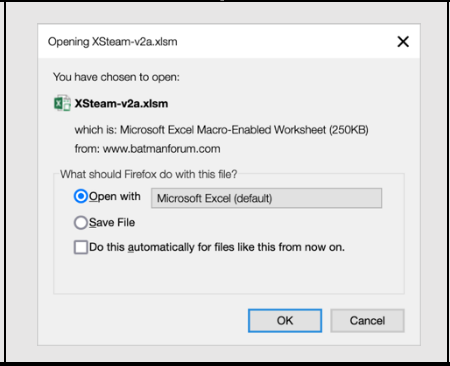
QtBitcoinTrader.exe

which is: Binary File (22.3 MB)
from: www.batmanforum.com

Would you like to save this file?

Save File | Cancel

Figure 3d. Firefox executable download notification



Opening XSteam-v2a.xlsm ✕

You have chosen to open:

XSteam-v2a.xlsm

which is: Microsoft Excel Macro-Enabled Worksheet (250KB)
from: www.batmanforum.com

What should Firefox do with this file?

● Open with Microsoft Excel (default)
○ Save File
□ Do this automatically for files like this from now on.

OK | Cancel

Figure 3e. Firefox file-open dialog for macro-enabled spreadsheet.

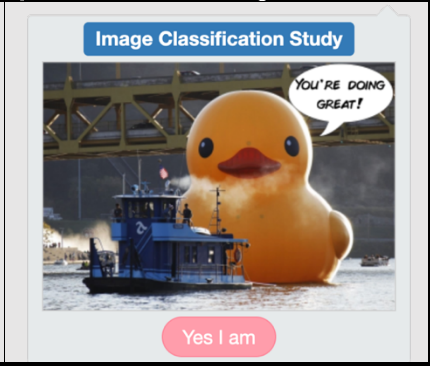


Image Classification Study

Yes I am

Figure 3f. Novel foil popup stimulus for assessing fatigue.

The experimental website loaded static images into a central frame, which participants were told contained live external websites:

> Warning: The researchers are not responsible for the content of the web pages 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.

In this way, participants were led to believe that security warnings were triggered by actual websites automatically visited in the center frame of the browser.

Both quantitative and qualitative responses from participants after the experiment supported that they believed that the warning was triggered by the website visited. For example,

[3] Although these warnings do not explicitly state the cybersecurity risk, nonetheless, they are security warnings because they prompt to confirm a potentially risky operation.

in a post-task survey, participants were shown a screenshot of the warning they received in the experiment and asked, “On a scale of 0 to 10, how realistic do you think the following message is?” Participants rated each warning an average of 7.5 or higher. Similarly, 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.” Another stated, “That was incredible deception. I am a software engineer with a background in cybersecurity and you fooled [me].” A third said, “I got bamboozled.”

Operationalizations of Habituation

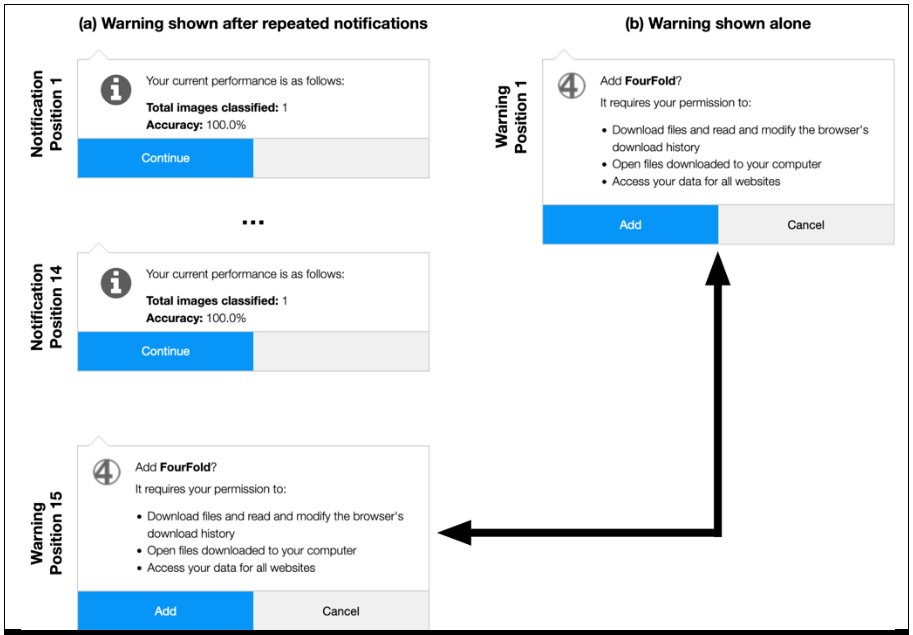
To demonstrate that habituation generalizes from notifications to security warnings (H1), it is first necessary to establish that habituation has occurred (Rankin, 2009). We operationalized habituation in two ways to serve as dependent variables. First, we measured habituation in terms of participants’ reaction time, measured as the time between the display of a notification or warning and when a user responds to it. As individuals become habituated, responses become more automatic with faster reaction times. This is because as individuals become habituated, they “rely more on the mental model of the warning as opposed to the actual warning presented, resulting in a decreased response” (Anderson, Vance, et al., 2016, p. 722). Reaction time has previously been demonstrated to accurately measure habituation both in IS (Anderson, Vance, et al., 2016; Eargle, 2017) and neurobiology (Mackworth, 1968).

Second, because our goal in this research was to ultimately improve security behavior (Crossler et al., 2013), our second dependent variable measured habituation in terms of its behavioral effect on users’ disregard of security warnings. Following habituation theory articulated by Vance et al. (2018), as the brain becomes more habituated to warnings, warning disregard behavior will likewise increase.[4] We operationalized security warning disregard as a binary variable that measures whether a participant disregarded the warning by clicking the less-safe option (Jenkins et al., 2016; Vance et al., 2014). For the warnings in this experiment, this means selecting “Don’t allow” (Figure 3b) or “Cancel” (Figure 3c–e).

Operationalizations of Generalization of Habituation

Because generalization is fundamental to habituation (Thompson & Spencer, 1966), they are often assessed together (Rankin et al., 2009). We operationalized generalization as the difference between responses in two separate experimental conditions: (1) the warning is shown *after* a person has been habituated to the performance notification through repeated exposures and (2) the same warning is shown *without* seeing a similar notification (Figure 4). We measured differences in responses both in terms of response time and warning disregard, as defined previously for our operationalization of habituation. If the response to the warning is lower after having become habituated to the performance notification compared with the response when the warning is seen on its own, then this is partial evidence that habituation has generalized from the notification to the warning (Rankin et al., 2009).

[4] Vance et al. (2018) measured “warning adherence,” which is the opposite of warning disregard.



Your current performance is as follows:

Total images classified: 1
Accuracy: 100.0%

Continue

Your current performance is as follows:

Total images classified: 1
Accuracy: 100.0%

Continue

Add **FourFold**?

It requires your permission to:

- Download files and read and modify the browser's download history
- Open files downloaded to your computer
- Access your data for all websites

Add Cancel

Add **FourFold**?

It requires your permission to:

- Download files and read and modify the browser's download history
- Open files downloaded to your computer
- Access your data for all websites

Add Cancel

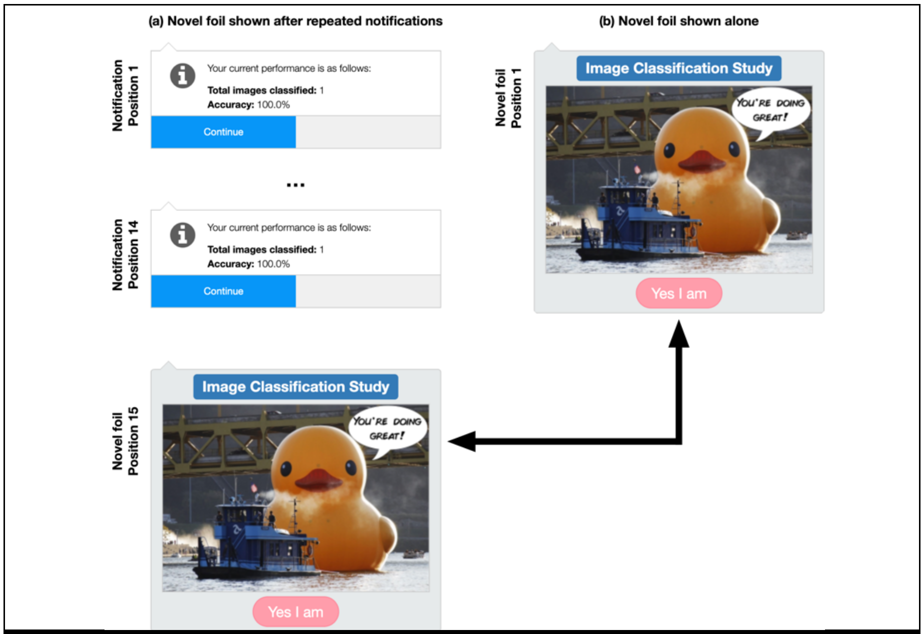
Figure 4. Test for generalization of habituation by comparing the response to a warning (a) after being habituated to a similar notification and (b) alone. Generalization of habituation is the degree to which habituation to the repeated notification carries over to the novel warning.

Ruling Out Fatigue

A rival explanation to habituation and generalization is fatigue (Thompson & Spencer, 1966); that is, that people have a lower response because they are tired after repeated exposures to a stimulus. Following the neurobiology literature, fatigue can be distinguished from habituation by displaying a novel stimulus that is greatly different from the stimulus to which a participant has become habituated (Rankin et al., 2009), which we refer to as a *novel foil*. First, a person becomes habituated to stimulus *A* through repeated exposure. If the response to a novel foil *F* recovers to the level of the response of *A* when it was first presented, then this demonstrates that habituation and not fatigue was the reason for the diminished response to *A*.

This approach can also be used to rule out fatigue in the case of generalization. If subjects show a diminished response to *A* through repeated exposure and exhibit a weak response to a novel-though-similar stimulus *B*, but a separate condition shows a strong response to *F*, then this is strong evidence that the diminished response to *B* is due to the generalization of habituation, not to fatigue (Rankin et al., 2009).

We used a novel foil in the form of a dialog window with a duck image designed to appear different from the performance notification to make generalization of habituation unlikely (Figure 5). We then compared the reaction times to the novel foil message in two separate experimental conditions: (1) the novel foil is seen for the first time before seeing a notification, and (2) the same novel foil is seen for the first time *after* a person has habituated to a notification through repeated exposures. If the response to the novel foil message is no different before seeing notifications compared with after having habituated to repeated notifications, then this suggests that habituation and not fatigue is the cause for the diminished response to the notification and to the generalization of this effect to warnings (Rankin et al., 2009).



Your current performance is as follows:

Total images classified: 1

Accuracy: 100.0%

Figure 5. Rule out fatigue by comparing the response to the novel foil (a) after being habituated to the performance notification and (b) alone. If there is no difference in response, this indicates that fatigue is not a factor.

Our full experimental design is shown in Table 2. We randomly assigned participants to 1 of 10 experimental conditions. These included the four Firefox warnings (Figure 3b–e) and the novel foil, both as the first exposure and 15th exposures (after 14 exposures to the performance notification).

Table 2. Experimental Design for Experiment 1

	Exposure position	
Message	**1**	**15**
Novel foil	*n* = 61	*n* = 61
Permission request	*n* = 61	*n* = 61
Extension install	*n* = 60	*n* = 61
Save executable	*n* = 61	*n* = 61
Open macro	*n* = 61	*n* = 61

Experiment 1: Analysis

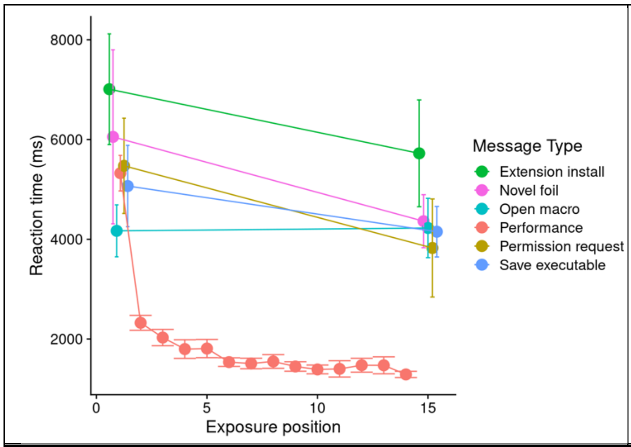
Testing for Habituation and Manipulation Check

We first performed a manipulation check to verify that habituation occurred for the participants in the treatment that received 14 performance notifications before seeing the warning. Figure 6 shows a clear pattern of habituation to the performance notification in terms of reaction time, with participants responding 24% faster at exposure 14 compared with exposure 1 ($p < 0.001$). This demonstrates that the manipulation of habituation was successful.

Testing H1: Whether Generalization Occurs

We tested H1 by examining two dependent variables: (1) the natural log of reaction time to the security warning[5] and (2) whether participants disregarded the warning. The test examined whether participants who received the warning after 14 exposures to the performance notification were more likely to respond faster to the warning and disregard it compared with participants who responded to the warning first without exposure to the performance notification.

[5] Taking the natural log of reaction time resolved issues of non-normality.



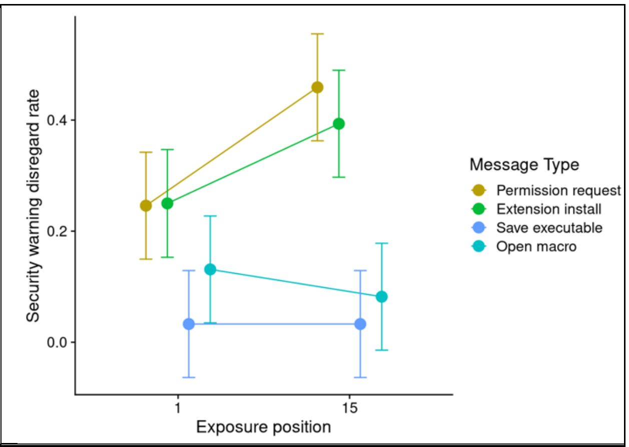


Figure 6. Comparison of reaction times to the security warnings and the performance notification across exposure positions.

Figure 7. Comparison of the rates of disregarding security warnings at exposure positions 1 and 15.

Note: Only the slopes for "Permission request" and "Extension install" are significant, indicating generalization to these warnings, which are visually similar to the performance notification. Standard error bars are shown at each point.

For this analysis, we fitted two models, one logistic regression for disregard behavior ("disregard") and one linear regression for the natural log of reaction time (see Table 3). Each had just one predictor: the exposure position at which the warning was displayed and specified as a dummy-coded fixed effect ("0" for warning alone and "1" for after 14 exposures to performance notifications).[6] The estimate of the exposure position at which the warning was displayed is statistically significant in both models in the hypothesized direction. Thus, the participants who viewed the warning after repeated notifications were 1.62 times more likely to disregard the warning, and they responded approximately 22% faster (4.6 seconds vs. 3.6 seconds). This suggests generalization of habituation and supports H1.

[6] These models were only fit against security warnings, not against participants in the novel-foil conditions, as the novel foil had to be dismissed and could not be disregarded.

Table 3. Logistic regression analysis of the response to the warning after exposure to notifications

	Warning Disregard		ln(Reaction time)	
Predictor	*Odds Ratio*	*Standard Error*	*Estimate*	*Standard Error*
(Intercept)	0.20***	0.03	8.44***	0.03
After notifications	1.62*	0.37	−0.24***	0.05
Observations	487†		609	
Tjur's (pseudo) R^2	0.009		0.037/0.036	
****p* < 0.001; **p* < 0.05. All tests were one-tailed. †The total number of participants less those who received the novel foil because the novel foil was not a warning that could be disregarded (see Table 2).				

Testing H2: Whether Distinctive Visual Appearance Mitigates Generalization

We also tested whether habituation to non-security-related notifications would generalize less to novel security warnings with a distinctive visual appearance (H2). To do this, we fitted regression models for both dependent variables, with predictors representing the different warning types. Both models specified the main effect of the warning type and the interaction of each warning type with whether it appeared at exposure 15. To facilitate model interpretation, neither the main effect of exposure position nor the model intercept was specified. In Table 4, the estimates for each interaction term indicate whether evidence of generalization existed for that warning type.

The results show that warnings most visually similar to the performance notification, namely the permission request and extension install warnings, were significantly more likely to be disregarded at exposure 15 after repeated exposures to notifications compared with exposure 1 (2.6 times vs. 1.95 times more likely). Participants also responded 39% and 30% faster to these warnings when seen at exposures 15 and 1, respectively (approximately 1.8 seconds faster for both warning types).

Table 4. Experiment 1 regression models predicting security warning disregard and reaction time

	DV: Warning disregard (Logistical regression)		**DV: ln(Reaction time) (Linear regression)**	
Predictor	*Odds Ratio*	*Standard Error*	*Estimate*	*Standard Error*
Permission request (at exposure 1)	0.33***	0.10	8.44***	0.08
Extension install (at exposure 1)	0.33***	0.10	8.69***	0.08
Save executable (at exposure 1)	0.03***	0.02	8.38***	0.08
Open macro (at exposure 1)	0.15***	0.06	8.24***	0.08
Permission request × after notifications	2.60**	1.02	−0.49***	0.11
Extension install × after notifications	1.95*	0.77	−0.36***	0.11
Save executable × after notifications	1.00ns	1.02	−0.16ns	0.11
Open macro × after notifications	0.59ns	0.36	−0.02ns	0.11
Novel foil			8.46***	0.08
Foil × after notifications			−0.16ns	0.11
Observations	487†		609	
Tjur’s (pseudo) R^2	0.142		0.995	
All tests were one-tailed. †The total number of participants less those who received the novel foil because the novel foil was not a warning that could be disregarded (see Table 2).				

By contrast, the warnings least visually similar to the performance notification, namely the save executable and open macro warnings, were no more likely to be ignored at exposure 1 than at exposure 15 (neither odds ratio was significantly different from 1; see Figure 7). Likewise, participants responded no faster to these warnings at exposure 15 than at exposure 1 (see Figure 6). These results strongly indicate that habituation generalized less to warnings with a distinctive visual appearance, supporting H2.

Ruling Out Fatigue

Finally, we tested whether participants' reaction times were caused by fatigue instead of generalization of habituation. To do this, we compared reaction times to the novel foil when it was shown alone with those after repeated exposure to the performance notification. This is tested by the "Foil × after notifications" row of the reaction time model information in Table 4. While the reaction time trended toward being faster when shown after notifications than when shown alone, it was not statistically significant. Therefore, these reaction time results do not support the rival explanation of fatigue for H1.

Experiment 1: Discussion

Experiment 1 supported H1 by showing that habituation to non-security-related notifications generalized to security warnings with a similar visual appearance in terms of both higher warning disregard and shorter reaction times. Experiment 1 also supported H2, demonstrating that generalization of habituation occurred less for warnings with a distinctive visual appearance. Furthermore, the results of experiment 1 also failed to support the rival explanation of fatigue by showing no change in reaction time in response to the novel foil when shown alone or after repeated exposure to notifications.

EXPERIMENT 2: FMRI EXPERIMENT

The results of experiment 1 showed a clear overall pattern of generalization of habituation as participants responded more quickly to the security warnings after being repeatedly exposed to the performance notification over time (supporting H1). However, while the rival hypothesis of fatigue was not supported by comparing reaction times to the novel foil of experiment 1, another rival explanation to H1 is that participants' faster reaction times could have been due to their becoming more cognitively engaged in the task over time and therefore quicker to dismiss notifications and less accurate at responding to warnings.

We conducted a follow-up fMRI experiment to (1) further rule out the rival explanations, (2) test H3 to determine whether varying the mode of interaction could mitigate generalization of habituation, and (3) examine the effect of generalization by decreased neural activity rather than response time. With regard to this last purpose, from the perspective of the philosophy of science, integrating findings across the behavioral and neural levels of analysis has value in providing a more complete view of phenomena (MacDougall-Shackleton, 2011; Marr, 1982).

Experiment 2: Participants

We scanned 25 participants (mean age: 24 years, 14 women) who underwent fMRI while they received repeated exposures to notifications and occasional exposures to warnings to confirm habituation and generalization in neural activity.[7] All participants were right-handed, were native English speakers, had normal or corrected visual acuity, and self-reported free of previous neurological and psychiatric disorders in addition to no current use of any psychoactive medications. Participants were recruited through a flier from the university community and provided written informed consent prior to participation in the experiment. Participants were compensated with either $20 or a one-fourth-scale three-dimensionally printed model of their brains. All procedures were approved by the local institutional review board.

[7] We conducted a pilot study that revealed a large estimate of effect size for the repetition effect (partial $\eta^2 = 0.7$). Using this estimated effect size, an a priori power analysis indicated that we would need four subjects to achieve a power greater than 0.8. This low sample size required for adequate power was due to the large effect size and is not typical of NeuroIS research, where samples of 15 to 25 participants are common. For example, Dimoka (2010), n = 15; Riedl et al. (2010), n = 20; Riedl et al. (2014), n = 18; Anderson, Vance, et al. (2016), n = 25; Jenkins et al. (2016), n = 24; Warkentin et al. (2016), n = 17, Vance et al. (2018), n = 15; Walden et al. (2018), experiment 1: n = 20, experiment 2: n = 20; Meservy et al. (2019), n = 29; Fadel et al. (2022), n = 29.

Experiment 2: fMRI Protocol

We followed the guidelines provided by Dimoka (2012) for conducting an fMRI study. Upon arrival, participants were screened to ensure MRI compatibility. They were verbally briefed about the MRI procedures and the task and were then positioned lying on their backs in the scanner. Once positioned in the scanner, participants completed a brief tutorial explaining the image classification task. Participants were not briefed regarding the security warnings. All stimuli were presented on an MRI-compatible liquid crystal display (LCD) monitor placed at the head of the bore of the MRI scanner and viewed through a mirror mounted on the head coil. Behavioral responses were collected with a fiber optic button box because mousing causes detrimental motion artifacts in the fMRI data.

We first performed a 10-second localizer scan, followed by a 7-minute structural scan. Following these scans, we started the experimental task, which was completed in one functional run. We used psiTurk (Eargle et al., 2022; Gureckis et al., 2016) to record participant keypresses and to display and record time stamps for stimuli in a web browser visible to participants on the LCD monitor. Time stamps were later synchronized with time stamps recorded by the fMRI scanner software. All ex post tests revealed that none of the subjects needed to be excluded (e.g., due to abnormalities or excessive movement).

Experiment 2: Task and Design

Participants performed the same Batman image classification task from experiment 1, albeit adapted as a within-subject, repeated-measures design, which is commonly used in fMRI studies to maximize the number of measurements taken from each subject to obtain reliable estimates of the subject's hemodynamic response (Dimoka, 2012). In this design, the participants were presented sets of 2–10 images to classify,[8] with each classification choice by the participant followed by a performance notification (Dale, 1999). At the end of each set, participants received one more image to classify, followed by either a randomly selected security warning or a random novel foil (see Figure 8 for the experimental design).

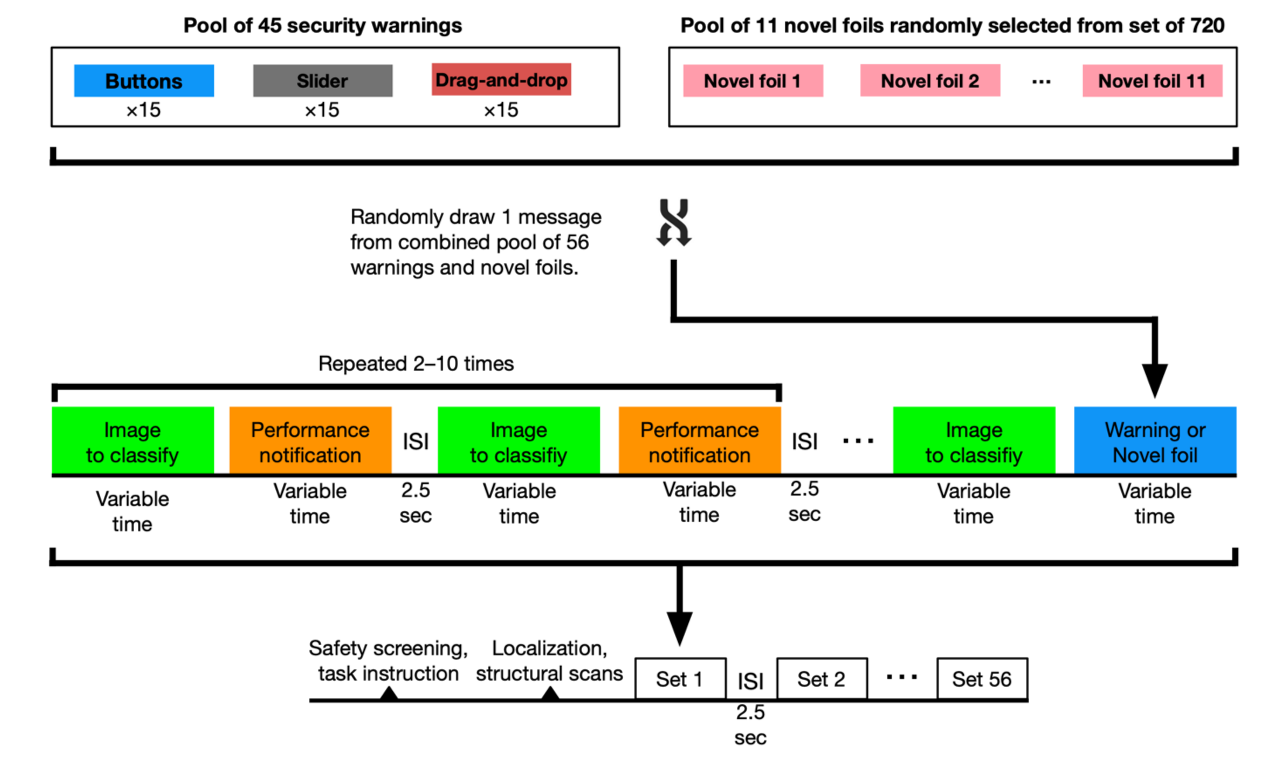
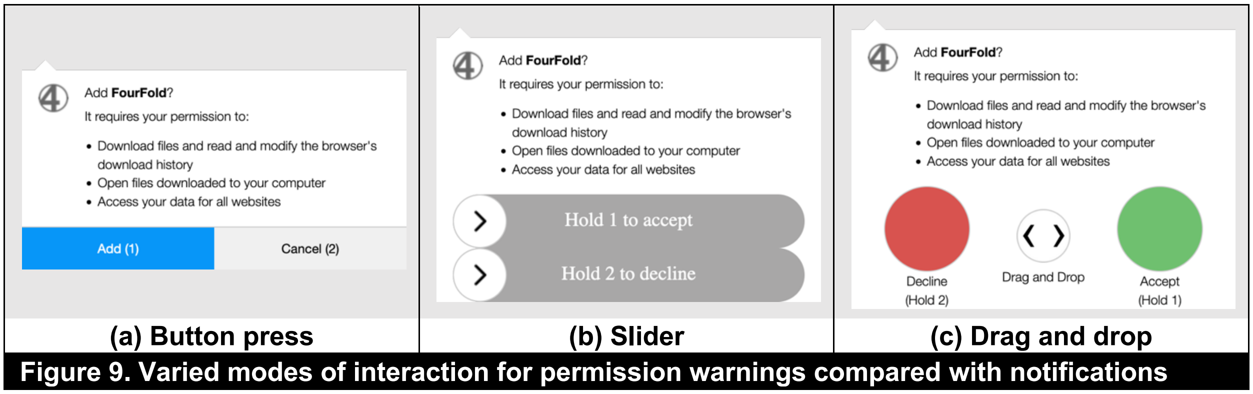


Figure 8. Within-subject experimental design for fMRI Experiment 2

For experiment 2, we used three security warning designs, which were all variations of the “Firefox add-on permission warning” from experiment 1 (see Figure 3c), which differed in their mode of interaction. The first mode-of-interaction variation was the standard button press, matching the mode of interaction of the performance notification (see Figure 9a). The second variation required participants to press and hold one of two buttons to move a slider (Figure 9b),

[8] The number of images in a set was randomly generated, which ensured that any effects observed would be due to the warning treatment, rather than the sequence of the presentation of the warnings or participants’ anticipation of the warnings. Furthermore, this randomization ensured an effective jitter between stimuli of similar types and avoided autocorrelation in the experimental design (see Dale, 1999).

and the third featured a drag-to-drop interaction by again requiring participants to press and hold one of two buttons to move the slider (Figure 9c). Each of these three warnings were shown 15 times, for a total of 45 sets of these types.



Add **FourFold**?

It requires your permission to:

- Download files and read and modify the browser's download history
- Open files downloaded to your computer
- Access your data for all websites

(a) Button press **(b) Slider** **(c) Drag and drop**

Figure 9. Varied modes of interaction for permission warnings compared with notifications

As with experiment 1, participants first completed a brief classification practice round that had no performance notifications. To avoid introducing bias, participants were not told about the warning treatments interspersed with performance notifications in the image classification task, nor were they instructed how to interact with them. Regardless, an analysis of participants' reaction times to the alternative modes of interaction treatments (Figure 9b–c) showed that the participants quickly learned how to interact with them: participants spent substantially longer to respond to the initial warnings than to all subsequent displays of those warning types (median seconds: slider, 1st = 8.4, 2nd–15th = 1.7; drag-and-drop, 1st = 9.7, 2nd–15th = 1.7), which suggests that their first impressions were sufficient to learn how to interact with these notification types.

The novel foils were random pictures of everyday objects (Kirwan and Stark 2007) and required participants to press a button to dismiss (Figure 10). A total of 11 sets with foils were shown, giving an approximate stimuli of interest to foil ratio of approximately 4:1 (a total of 56 sets). The 11 foils were randomly chosen without replacement for each participant from a set of 720 foils, ensuring that no participant saw the same foil twice.

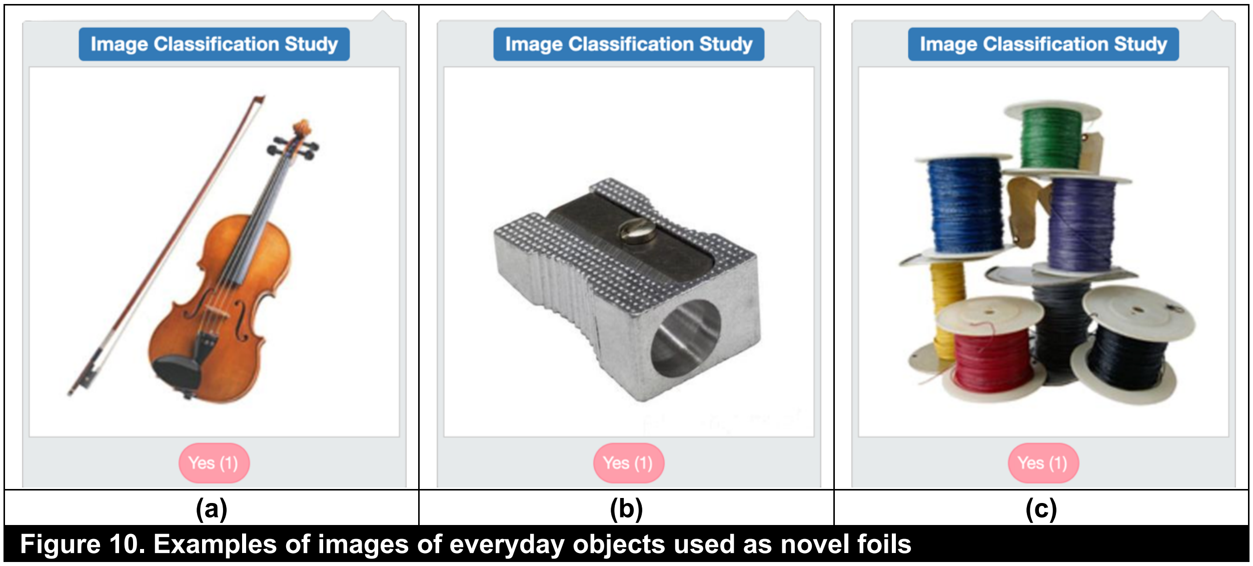


Image Classification Study
Yes (1)
Image Classification Study
Yes (1)
Image Classification Study
Yes (1)
(a)
(b)
(c)

Figure 10. Examples of images of everyday objects used as novel foils

Performance notifications, security warnings, and foils were displayed until dismissed by the participant. Owing to the self-paced nature of the task, the time to complete the task varied between the participants, with a minimum time of 20 minutes 20 seconds and a maximum time of 29 minutes 8 seconds (mean 23 minutes 37 seconds). Timing information for the analysis of the fMRI time course data was determined on the basis of both the stimulus presentation time and participant keypresses. See Appendix A for scan acquisition parameters and detailed MRI data analysis description.

Operationalization of Habituation and Generalization

We operationalized habituation as decreased neural activity in response to a warning, as measured indirectly by tracking changes in blood oxygenation level, which are driven by changes in the metabolic demands of active neural populations. This phenomenon is known as the blood oxygen level dependent (BOLD) effect, and its magnitude is proportional to the degree of underlying neural activation (Logothetis et al., 2001). By measuring the BOLD effect, researchers can both identify distinct regions of the brain where activity correlates with specific cognitive processes (e.g., perception or memory retrieval) and evaluate the degree of activation in these regions. Habituation has been widely observed in terms of decreased neural response to a stimulus in all biological sciences (Rankin et al., 2009). Previous work in IS has also demonstrated habituation in terms of decreased neural activity with repeated exposure to security warnings (Anderson, Vance, et al., 2016; Vance et al., 2018).

Therefore, if we observe that (1) decreased neural response to a repeated performance notification (2) carries over to a novel security warning that is similar in appearance to the notification but (3) the neural response to a novel foil is still high after repeated exposure to the performance notification, then this is strong evidence that generalization of habituation has occurred (Rankin et al., 2009).

Following our within-subject design, a common test of generalization is to assert whether the neural response is weaker for the first exposure to the novel stimuli than for the *first* exposure to the initial stimuli (Grizzard et al., 2015). Comparing instead with the *immediately preceding exposure* of the initial stimuli is therefore a more stringent version of that test.

Experiment 2: Analysis

Testing for Habituation and Manipulation Check

As all our hypotheses involved generalization of habituation, we first identified brain regions that demonstrated habituation to the performance notifications. This was accomplished by comparing changes in neural activity in the brain (i.e., a linear contrast) across the first 15 presentations of the notification (thus matching the number of presentations of each type of security warning). We identified four clusters in the brain that demonstrated a clear pattern of habituation in the form of significant linear decreases in response to repeated presentations of the performance notification. These regions were the bilateral ventral visual stream, left dorsal visual stream, bilateral dorsomedial prefrontal cortex, and right inferior frontal gyrus (see Table B1 in Appendix B for full cluster details). The bilateral ventral visual stream has been identified as a site of habituation using similar fMRI paradigms in previous studies (e.g., Anderson, Vance, et al., 2016; Kirwan et al., 2020; Vance et al., 2018). As this is a well-established site of habituation to visual stimuli, we focused our subsequent tests for generalization on this cluster (Figure 11; Berry et al., 2017). This also demonstrated that our experimental design successfully manipulated habituation.

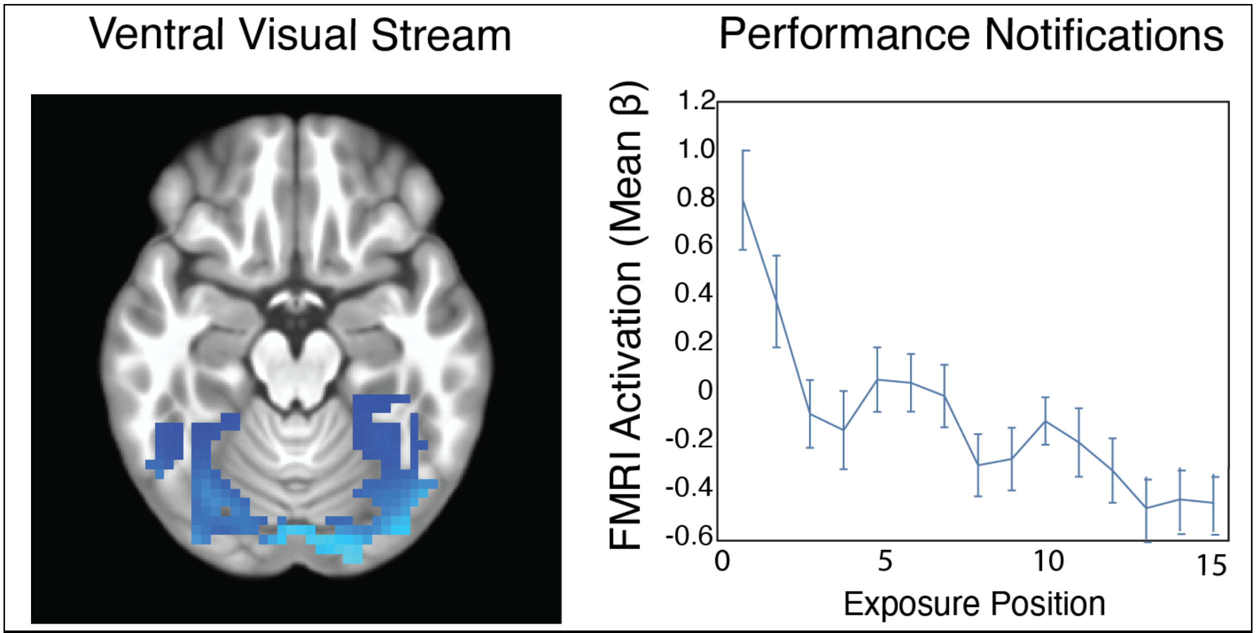


Figure 11. Location (left) of the ventral visual stream cluster where there was a significant linear decrease in fMRI activation over 15 presentations of the performance notification (right). Error bars = ± standard error of the mean. Baseline (y-axis 0) is the image classification task activation.

Testing H1: Whether Generalization Occurs

Having established habituation in the ventral visual processing stream in response to performance notifications, we next examined whether habituation to the performance notification will generalize to security warnings with a similar visual appearance (H1). To test this, we examined the fMRI activation for each participant's first presentation of the button-press warning (Figure 9a) compared with its immediately preceding performance notification activation. We observed significantly lower fMRI activation for the button-press warning relative to its immediately preceding performance notification, which suggests that habituation generalized to the button-press warning [$t(23) = -3.071$; $p = 0.005$ two-sided; 95% confidence interval [CI]; −0.62 to −0.12; Cohen's d=.059], supporting H1.

Ruling out cognitive engagement and fatigue

We first found that the linear activation *decrease* discovered in the visual ventral processing stream region fails to support a rival hypothesis of cognitive engagement from experiment 1. As experiments 1 and 2 used highly similar tasks and because the window of the first 15 notification exposures from the experiment 2 task aligned with the maximum number of exposures in the experiment 1 task, the evidence of decreased neural activity over repeated exposures in experiment 2 fails to support the rival hypothesis of increasing cognitive engagement that explains the faster security warning reaction times in experiment 1. If that were the case, cognitive engagement would have shown an *increase* in activations for the region tested in this experiment (Berry et al., 2017).

We then tested whether the observed decrease in fMRI activation was due to fatigue, instead of habituation. We compared the mean activations for all performance notifications with those of all novel foils. We reasoned that if fatigue caused the decrease in activation in response to performance notifications, then it would also decrease activation in response to the novel foils, given that the foils were randomly interspersed among the performance notifications. This would lead to no significant difference in activations between the performance notifications and the novel foils. However, activation for all novel foils was significantly greater than that for all performance notifications [$t(23) = 4.50$; $p < 0.001$; 95% CI, 0.094–0.253], consistent with the habituation theory and not fatigue (Figure 12, “performance” vs. “novel foil”).

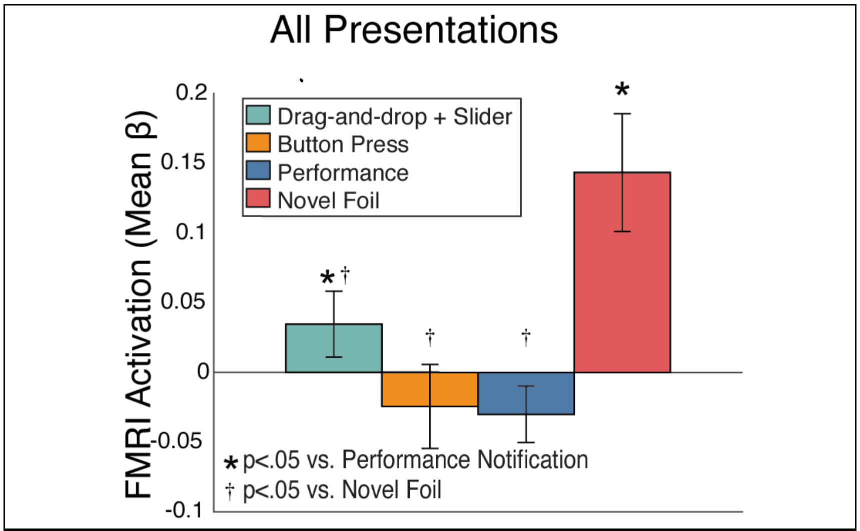


Figure 12. Mean fMRI activation for all security warnings, performance notifications, and novel foils (0 = baseline of image classification task). Error bars = ± standard error of the mean.

Testing H3: Whether the mode of interaction mitigates generalization

Next, we examined whether habituation to the performance notification generalized less to security warnings with a distinctive mode of interaction (H3). To test this, we compared the mean fMRI activation within the ventral visual stream cluster for all presentations of each warning type and for the performance notifications, collapsed across all trials, using the Batman image classification activations as the baseline (Figure 11). As the slider and drag-and-drop notifications had similar modes of interaction (i.e., press and hold a button on a controller in the MRI scanner), we collapsed these notification types for this analysis. We found that activation for the button-press warning type (Figure 9a) was not different from that of the performance notification, indicating generalization of habituation between these two messages, further strengthening H1. By contrast, activation was significantly higher for the slider and drag-and-drop warning types than for the performance notification [$t(23)=3.83$; $p <.001$ two-sided; 95%

CI, .03, .1; Cohen’s d=.083], which suggests a resistance to generalization of habituation for these types. In summary, the warnings with a distinctive mode of interaction (slider and drag-and-drop) showed substantially less generalization of habituation than the warning with the same mode of interaction (button press) as the performance notification. Therefore, experiment 2 supports H3.

Experiment 2: Discussion

The tests in experiment 2 effectively ruled out the rival explanations of fatigue and cognitive engagement for H1. First, the general trend of decreasing neural activity across the experimental task does not support the rival explanation of increasing cognitive engagement. Second, because participants showed a diminished neural response to the performance notification and the modally similar button-press warning, the neural response to the novel foil was significantly higher than the rest of the messages, this is strong evidence of generalization and not fatigue (Rankin et al., 2009). Experiment 2 also supported H3 by showing that habituation to the performance notification generalized less to the slider + drag-and-drop warnings with distinctive modes of interaction.

EXPERIMENT 3: ONLINE FIELD EXPERIMENT

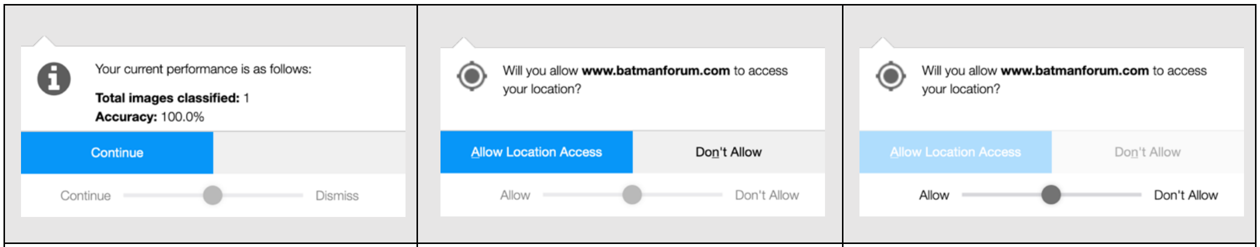
Experiment 2 provided evidence that security warnings with distinctive modes of interaction are more resistant to generalization of habituation, supporting H3. However, it is difficult to change the mode of interaction of a warning without also simultaneously changing its visual appearance. For example, in experiment 2, the warnings with the press-and-hold and drag-and-drop modes of interaction were more visually dissimilar to the performance notification than the extension warning owing to the differences in the UI widgets particular to those modes of interaction. Therefore, it is not possible from the results of experiment 2 alone to distinguish the effects of changing (a) the visual appearance and (b) the mode of interaction. To tease these effects apart, we conducted another field experiment with experiment 3 to test H3 while holding visual appearance constant.

Experiment 3: Participants

We recruited 241 participants from Amazon Mechanical Turk (mean age 39 years, 54% male) to participate in experiment 3. Seven outlier participants were removed on the basis of extreme reaction times in the task, leaving 234 participants. As in experiment 1, we followed Steelman et al. (2014) and required all participants to be from the United States. As in experiment 1, they were paid $1.50 ($1 plus a $.50 performance bonus) for a 5-minute task.

Experiment 3: Experimental Task and Treatments

The experimental task for experiment 3 was essentially the same as for experiment 1. However, in contrast to experiment 1, the experiment only involved a performance notification and permission warning. To hold visual appearance constant as much as possible, we incorporated both a traditional button and a disabled slider widget into the performance notification and permission warning (see Figure 13a and 13b). In this way, we could keep visual appearance constant, except for slightly fading a button or slider widget to indicate its disabled status). To increase the realism of this new notification design, we started with the standard Firefox HTML5-style notification and incorporated an actual slider from the Firefox setting dialogs, following the Firefox UI guidelines (e.g., for positioning of labels and length of the slider line). Our pilot testing indicated that participants understood how to interact with the different configurations.



Your current performance is as follows:

Total images classified: 1
Accuracy: 100.0%

Will you allow **www.batmanforum.com** to access your location?

Will you allow **www.batmanforum.com** to access your location?

(a) Performance notification **(b) button-click enabled** **(c) slider enabled**

Figure 13. Experiment 3 notification (a) and warnings (b) and (c)

Participants were randomly assigned to one of four treatments (Table 5). While all participants saw the performance notification with only the buttons enabled (see Figure 13a), half of them saw the permission request warning with the buttons enabled (see Figure 13b), while the other half received the permission request with the slider enabled (see Figure 13c). Within these groups, half of the participants were shown the permission request warning after just one exposure to the performance notification. The other half of the participants were shown the warning after 29 exposures to the performance notification. This is in contrast to experiment 1, which displayed the warning after 14 exposures. This change was necessary because our pilot testing revealed that more exposures were required for participants to become habituated to the novel combination of a button and slider present on the same notification.

Table 5. Experimental design for experiment 3

	Exposure position	
Permission request	**1**	**30**
Buttons enabled	*n* = 60	*n* = 59
Slider enabled	*n* = 60	*n* = 55

Experiment 3: Analysis

Testing for habituation and manipulation check

As with experiment 1, we verified that participants in the performance notification treatment experienced habituation. Figure 14 shows a clear pattern of habituation to the performance notification in terms of reaction time, with participants responding 18% faster at exposure 29 than at exposure 1 ($p < 0.05$). This demonstrates that the manipulation of habituation was successful.

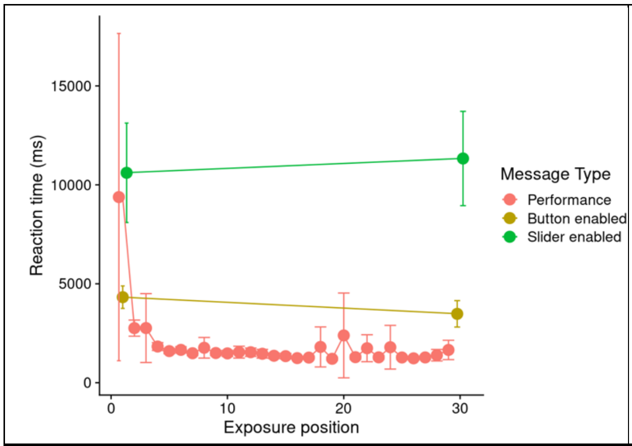


Figure 14. Comparison of the reaction times to the security warnings and performance notifications across exposure positions.

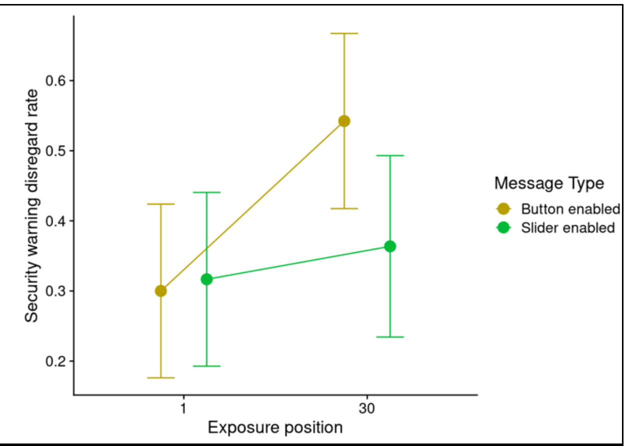


Figure 15. Comparison of the rates of security warning disregard at exposure positions 1 and 30.

Note: Only the slope for "Button enabled" was significant, indicating generalization to this warning with a similar mode of interaction to the performance notification. Standard error bars are shown at each point.

Testing H3: Whether mode of interaction mitigates generalization

Similar to the H2 test in experiment 1, we examined two dependent variables: (1) the natural log of reaction time to the security warning and (2) whether participants disregarded the warning. We fitted models for both dependent variables, with predictors representing the different warning types. Both models specified the main effect of the warning type and the interaction of each warning type with whether it appeared at exposure 30. To facilitate model interpretation, neither the main effect of the exposure position nor the model intercept was specified. In Table 6, the estimates for each interaction term indicate whether evidence of generalization existed for that warning type.

Table 6. Experiment 3 regression models predicting security warning disregard and reaction time

	DV: Warning disregard (Logistical regression)		**DV: ln(Reaction time) (Linear regression)**	
Predictor	*Odds Ratio*	*Standard Error*	*Estimate*	*Standard Error*
Permission warning (button enabled)	0.43**	0.12	8.26***	0.08
Permission warning (slider enabled)	0.46**	0.13	9.03***	0.08
Permission warning (button enabled) × displayed at exposure 30	2.77**	1.06	-0.38***	0.11
Permission warning (slider enabled) × displayed at exposure 30	1.23ns	0.49	0.10ns	0.12
Observations	234		234	
Tjur's (pseudo) R^2	0.040		0.995	
***p < 0.001; **p < 0.01; *ns* = not significant. All tests were one-tailed.				

The results show that the warning with the button enabled was 2.8 times more likely to be disregarded when shown at exposure 30 than at exposure 1. Participants also responded 32% faster to this warning when seen at exposure 30 than at exposure 1 (approximately 1.2 seconds faster).

By contrast, the warning with the slider enabled was no more likely to be disregarded at exposure 30 than at exposure 1 (the odds ratio was not significantly different from 1; see Figure 15). Likewise, participants responded no faster to this warning at exposure 30 than at exposure 1 (see Figure 14). These results strongly indicate that habituation generalized less to warnings with a distinctive mode of interaction, supporting H3.

Experiment 3: Discussion

In summary, experiment 3 revealed strong support for H3, which states that habituation to non-security-related notifications will generalize less to novel security warnings with a distinctive mode of interaction. For the button-enabled warning with the same mode of interaction as the performance notification, the likelihood of warning disregard increased, and the reaction time decreased between exposures 1 and 30. By contrast, the slider-enabled warning with a distinctive mode of interaction showed no significant difference between exposures 1 and 30, either for warning disregard or reaction time. Most importantly, these results were obtained while holding visual appearance essentially constant.

GENERAL DISCUSSION

The contributions of this study are summarized in Table 7. Although the effect of habituation on security warning disregard has been investigated previously (Anderson, Jenkins, et al., 2016; Anderson, Vance, et al., 2016; Vance et al., 2018), previous research has been narrow in scope in that it has considered only habituation to warnings. However, security warnings are relatively rare compared with ubiquitous notifications, and security scholars have long speculated that people’s automated reactions to notifications carry over to security warnings (e.g., Böhme & Köpsell, 2010; Vance et al., 2018; West, 2008). The results of this study demonstrate that the problem of habituation to security warnings is actually much greater than previous research has suggested because exposure to frequent notifications causes people to become habituated to similar security warnings they have never seen before.

Table 7 Contributions

Research question/hypothesis	Key contribution
RQ1: Does habituation to frequent non-security-related notifications generalize to infrequent security warnings? H1: Habituation to non-security-related notifications will generalize to security warnings with a similar visual appearance.	1. Demonstrated that generalization, in fact, occurs, consistent with habituation theory. The combination of two field experiments (experiments 1 and 3) and one fMRI laboratory experiment (experiment 2) increases overall precision and generalizability. 2. Measured generalization behaviorally (with reaction time and warning disregard) and neurally (with neural activation in response to warnings and notifications) to robustly test H1. 3. Ruled out rival explanations of fatigue and cognitive engagement by showing the neural basis for the generalization effect in experiment 2. 4. Demonstrated that previous guidance to mitigate habituation by “reducing the appearance of warnings” is incomplete because it does not prevent generalized habituation from non-security-related notifications. 5. Showed that the widely accepted UI principle of consistency poses a threat to users by desensitizing them to rare security warnings.
RQ2: How can generalization to security warnings be reduced? H2: Habituation to non-security-related notifications will generalize less to novel security warnings with a distinctive visual appearance. H3: Habituation to non-security-related notifications will generalize less to novel security warnings with a distinctive mode of interaction.	1. In experiment 1, we found that differentiating the visual appearance of a warning substantially reduced generalization, with the most distinctive warnings exhibiting the least generalization, supporting H2. 2. In experiments 2 and 3, we showed that differentiating the mode of interaction of a warning significantly reduced generalization, supporting H3. 3. In experiment 3, we controlled for visual appearance to isolate the effect of differentiating the mode of interaction. This test also strongly supported H3.

This experiment contributes by empirically demonstrating that habituation to non-security-related notifications carries over to novel security warnings. We quantified this effect behaviorally in experiment 1 in terms of reaction time and warning disregard behavior, showing that participants responded 30–39% faster to a novel warning and were 1.95–2.6 times more likely to disregard it after first habituating to a notification than when they received the warning without prior exposure to the notification. We also demonstrated this effect neurally using fMRI imaging, providing a comprehensive view of this phenomenon than is possible by measuring behavior alone (MacDougall-Shackleton, 2011; Marr, 1982). Our use of two field experiments and one laboratory experiment enhanced the overall precision and external validity of this finding.

More importantly, this study contributes to the literature by bringing theoretical depth to this issue by identifying the mechanism of generalization from habituation theory of neurobiology, which explains that the brain automatically diminishes responses to repeated stimuli and that this diminished response generalizes to novel stimuli that share similar characteristics. We also ruled out rival explanations of fatigue (experiments 1 and 2) and cognitive engagement (experiment 2), further supporting the explanation of generalization. Knowing the precise mechanism at work is important to theory and practice. For example, previous studies have conflated habituation and fatigue (e.g., Acer et al., 2017; Akhawe & Felt, 2013). Accordingly, the conventional wisdom to reduce security warning disregard is to reduce the frequency of warnings (Acer et al., 2017; Akhawe & Felt, 2013; Krol et al., 2012; Zeng et al., 2019). However, the accurate identification of the mechanism of generalization revealed that this approach will not work in an environment of frequent and similar notifications.

Furthermore, understanding the mechanism of generalization allowed us to contribute by developing and testing two theoretically driven interventions that directly address the problem. First, following the dual-process theory of habituation (Groves & Thompson, 1970), we hypothesized that differentiating the visual appearance of warnings so that it breaks the mental model of frequent notifications will prevent habituation from generalizing from notifications to warnings (H2). We tested this hypothesis in experiment 1 by comparing various actual warnings with an actual notification from the Firebox web browser to enhance ecological validity. Our findings strongly supported H2: participants who received warnings that were visually similar to non-security-related notifications responded 30–39% faster and were 1.95–2.6 times more likely to disregard the warning, whereas those who received distinctive warnings were no more likely to disregard the warning, even after repeated exposure to notifications. This approach also has the advantage of imposing no cost to the user in terms of time or effort.

Second, following schema theory (Rumelhart, 1980), we hypothesized that making the mode of interaction of a warning distinct from the conventional "click to dismiss" paradigm will break the automatic, muscle-memory schema of how to physically respond to notifications. This forces the user to consciously attend to the warning, so that warnings with a distinctive mode of interaction will exhibit less generalization (H3). We tested this hypothesis in experiment 2 using an fMRI experiment and found that warnings with a slider or drag-and-drop mode of interaction exhibited substantially less generalized habituation as measured by neural activation than warnings with the standard "click to dismiss" mode. In addition, because it is difficult to change the mode of interaction without also changing visual appearance, in experiment 3, we conducted another field experiment in which we tested H3 while holding visual appearance constant as much as possible. After exposing participants to a series of notifications, we found that those who received a warning with the "click to dismiss" mode of interaction responded 38% faster and were 2–2.8 times more likely to disregard the warning than those who received the warning with the slider mode. Moreover, those who received the slider-mode warnings showed no indication of generalization of habituation from repeated notifications. This second intervention is valuable because it provides developers an alternative to reduce generalization of habituation if changing visual appearance is not an option (e.g., due to a policy for UI consistency).

Overall, this research advances understanding of security warning disregard by revealing the problem of generalization of habituation and how this increases security warning disregard. However, our findings contradict the widely held UI principle in industry of making UI elements visually consistent (Apple, 2022; Google, 2022; Krug, 2014; Microsoft, 2022). Our results indicate that in the case of warnings, applying the UI principle of visual consistency poses a security threat because it promotes generalization of habituation. Furthermore, because adherence to this UI principle is widespread, so is the threat of generalization of habituation. Therefore, this research opens new research avenues, pointing the way for researchers and practitioners to develop and test security warning designs that are resistant to generalization of habituation and reduce security warning disregard.

Limitations

Our research is subject to several limitations. First, the fMRI experiment in experiment 2 necessarily involved low ecological validity because participants were required to perform the experimental task on their backs in an MRI scanner (Riedl et al., 2014). In addition, the experiment task was not on their own laptops, so participants might not have perceived a typical level of risk as they would on their own laptops in normal settings. However, these limitations were offset by experiments 1 and 3, which were field experiments conducted in participants' own environments and performed on their own computers. In this way, we followed the guidelines established by Kirwan et al. (2022) to complement NeuroIS experiments with behavioral experiments to enhance the overall internal, external, and ecological validity of a study.

Second, experiments 1–3 were designed to expose participants to notifications at a higher rate than normally encountered in the same amount of time during typical computer use. However, Vance et al. (2018) showed a similar rate of habituation of security warnings shown in a concentrated laboratory experiment compared with a more naturalistic 3-week window. In addition, participants' exposure to up to 15 notifications in experiment 1 or 30 in experiment 2 during the experimental sessions is close to the number of notifications reported in observational studies (Shirazi et al., 2014). Regardless, in future research, it would be interesting to explore if generalization of habituation occurs with the same number of exposures distributed across a longer time window.

CONCLUSION

Generalization of habituation is a serious problem because it causes users to tune out important security warnings, even if the warning has never been seen before, simply because they have become inured to visually or modally similar non-security-related notifications. This paper provides empirical evidence for this phenomenon. We also demonstrated in two online field experiments and an fMRI laboratory experiment that the effect of generalization can be reduced by changing either the visual appearance or the mode of interaction of a security warning to distinguish it from common and benign notifications. These solutions can enable software developers to create visually and interactively distinct security warnings that circumvent the neurobiological effects of generalization and create space for more conscious deliberation and facilitate greater adherence to security warnings.

REFERENCES

Acer, M. E., Stark, E., Felt, A. P., Fahl, S., Bhargava, R., Dev, B., Braithwaite, M., Sleevi, R., & Tabriz, P. (2017). Where the wild warnings are: Root causes of Chrome HTTPS certificate errors. *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security (CCS'17)*, New York, NY, USA, 1407–1420. https://doi.org/10.1145/3133956.3134007

Adkins, D. L., & Boychuck, J. (2006). Motor training induces experience specific patterns of plasticity across motor cortex and spinal cord. *Journal of Applied Physiology*, *101*(6), 1776–1782.

Akhawe, D., & Felt, A. P. (2013). *Alice in Warningland: A large-scale field study of browser security warning effectiveness*. The 22nd USENIX conference on security, Washington, D.C., 272–272.

Amran, A., Fitri, Z. Z., & Singh, M. K. M. (2018). Habituation effects in computer security warning. *Information Security Journal: A Global Perspective*, *27*(4), 192–204.

Anderson, B. B., Jenkins, J., Vance, A., Kirwan, C. B., & Eargle, D. (2016). Your memory is working against you: How eye tracking and memory explain susceptibility to phishing. *Decision Support Systems*, (92), 3–13.

Anderson, B. B., Vance, A., Kirwan, C. B., Jenkins, J., & Eargle, D. 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.

Anderson, E. M., Chang, Y. J., Hespos, S., & Gentner, D. (2022). No evidence for language benefits in infant relational learning. *Infant Behavior and Development*, 66, 101666, 1–17.

Apple. (2022). Human interface guidelines. https://developer.apple.com/design

Berry, A. S., Sarter, M., & Lustig, C. (2017). Distinct frontoparietal networks underlying attentional effort and cognitive control. *Journal of Cognitive Neuroscience*, *29*(7), 1212–1225.

Böhme, R., & Köpsell, S. (2010). Trained to accept? A field experiment on consent dialogs. *Proceedings of the SIGCHI conference on human factors in computing systems*, Atlanta, GA, 2403–2406.

Bravo-Lillo, C., Komanduri, S., Cranor, L. F., Reeder, R. W., Sleeper, M., Downs, J., & Schechter, S. (2013). *Your attention please: Designing security-decision UIs to make genuine risks harder to ignore. SOUPS '13: Proceedings of the ninth symposium on usable privacy and security*, Newcastle, UK.

Brustoloni, J. C., & Villamarín-Salomón, R. (2007). Improving security decisions with polymorphic and audited dialogs. *SOUPS '07: Proceedings of the third symposium on usable privacy and security*, Pittsburgh, PA, 76–85.

Bushman, B. J., & Anderson, C. A. (2009). Comfortably numb: Desensitizing effects of violent media on helping others. *Psychological Science*, *20*(3), 273–277.

CISA. (2022). Weak security controls and practices routinely exploited for initial access. https://www.cisa.gov/uscert/ncas/alerts/aa22-137a

Cooper, A., Reinmann, R., & Cronin, D. (2007). *About Face 3: The essentials of interaction design*. Wiley.

Crossler, R. E., Johnston, A. C., Lowry, P. B., Hu, Q., Warkentin, M., & Baskerville, R. (2013). Future directions for behavioral information security research. *Computers & Security*, (32), 90–101.

Curtis, B. (1989). Engineering computer "look and feel": User interface technology and human factors engineering. *Jurimetrics*, *30*(1), 51–78.

Dimoka, A. (2012). How to conduct a functional magnetic resonance (fMRI) study in social science research. *MIS Quarterly*, 36(3), 811–840.

Djamasbi, S., Siegel, M., & Tullis, T. (2011). Visual hierarchy and viewing behavior: An eye tracking study. In *International conference on human-computer interaction,* Jacko, J. Ed., Springer, Berlin, Heidelberg. 331–340.

Eargle, D., Gureckis, T., Rich, A. S., McDonnell, J., & Martin, J. B. (2022). *PsiTurk: An open platform for science on Amazon Mechanical Turk (V3.2.1)*. https://zenodo.org/record/7269012.

Eargle, D. W. (2017). Security messages or: How i learned to stop disregarding and heed the warning [Doctoral dissertation]. University of Pittsburgh.

Eickmeyer, K. (2022) Helping users stay safe: Blocking internet macros by default in Office, https://techcommunity.microsoft.com/t5/microsoft-365-blog/helping-users-stay-safe-blocking-internet-macros-by-default-in/ba-p/3071805

Epstein, L. H., & Carr, K. A. (2021). Food reinforcement and habituation to food are processes related to initiation and cessation of eating. *Physiology & Behavior*, 239, 113512, 1–12.

Google. (2022). Material design for Android. https://developer.android.com/design

Grizzard, M., Tamborini, R., Sherry, J. L., & Weber, R. 2017. Repeated play reduces video games' ability to elicit guilt: Evidence from a longitudinal experiment. *Media Psychology*, *20*(2), 267–290.

Grizzard, M., Tamborini, R., Sherry, J. L., Weber, R., Prabhu, S., Hahn, L., & Idzik, P. (2015). The thrill is gone, but you might not know: Habituation and generalization of biophysiological and self-reported arousal responses to video games. *Communication Monographs*, *82*(1), 64–87.

Gureckis, T.M., Martin, J., McDonnell, J., Rich, A.S., Markant, D., Coenen, A., Halpern, D., Hamrick, J.B., Chan, P. (2016) psiTurk: An open-source framework for conducting replicable behavioral experiments online. Behavioral Research Methods, 48 (3), 829-842. DOI: http://doi.org/10.3758/s13428-015-0642-8

Groves, P. M., & Thompson, R. F. (1970). Habituation: A dual-process theory. *Psychological Review*, *77*(5), 419–450.

James, I., Goodman, M., & Reichelt, F. K. (2013). What clinicians can learn from schema change in sport. *The Cognitive Behaviour Therapist*, *6*, e14, 1–9.

Jenkins, J. L., Anderson, B. B., Vance, A., Kirwan, C. B., & Eargle, D. (2016). More harm than good? How security messages that interrupt make us vulnerable. *Information Systems Research*, *27*(4), 880–896.

Kirwan, C. B., Bjornn, D. K., Anderson, B. B., Vance, A., Eargle, D., & Jenkins, J. L. (2020). Repetition of computer security warnings results in differential repetition suppression effects as revealed with functional MRI. *Frontiers in Psychology*, 11, 528079.

Kirwan, C. B., & Stark, C. E. (2007). Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learning & Memory*, 14(9), 625–633.

Krakauer, J. W., & Shadmehr, R. (2006). Consolidation of motor memory. *Trends in Neurosciences*, *29*(1), 58–64.

Krol, K., Moroz, M., & Sasse, M. A. (2012). *Don't work. Can't work? Why it's time to rethink security warnings*. 7th international conference on risk and security of internet and systems (CRiSIS), 1–8.

Krug, S. (2015). *Don't make me think, revisited: A common sense approach to web and mobile usability*. Pearson Education, New York, NY.

Lampinen, J. M., & Moore, K. N. (2016). Missing person alerts: Does repeated exposure decrease their effectiveness? *Journal of Experimental Criminology*, *12*(4), 587–598.

Li, X., Morgan, P. S., Ashburner, J., Smith, J., & Rorden, C. (2016). The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *Journal of Neuroscience Methods*, *264*, 47–56.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*(6843), 150–157.

MacDougall-Shackleton, S. A. (2011). The levels of analysis revisited. *Philosophical Transactions of the Royal Society B*, *366*(1574), 2076–2085.

Mackworth, J. F. (1968). Vigilance, arousal, and habituation. *Psychological Review*, *75*(4), 308.

Mandiant. (2021). *M-Trends 2021: FireEye Mandiant services special report*. https://www.mandiant.com/resources/m-trends-2021

Marr, D. (1982). *Vision: A computational approach*. MIT Press, Cambridge, MA.

Microsoft. (2022). *Microsoft design*. https://www.microsoft.com/Design

Mrug, S., Loosier, P. S., & Windle, M. (2008). Violence exposure across multiple contexts: individual and joint effects on adjustment. *American Journal of Orthopsychiatry*, *78*(1), 70–84.

Mrug, S., Madan, A., & Windle, M. (2016). Emotional desensitization to violence contributes to adolescents' violent behavior. *Journal of Abnormal Child Psychology*, *44*(1), 75–86.

Mrug, S., & Windle, M. (2010). Prospective effects of violence exposure across multiple contexts on early adolescents' internalizing and externalizing problems. *Journal of Child Psychology and Psychiatry*, *51*(8), 953–961.

MTurk. (2017). Tutorial: How to label thousands of images using the crowd. https://blog.mturk.com/tutorial-how-to-label-thousands-of-images-using-the-crowd-bea164ccbefc

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving bold activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage*, *59*(3), 2636–2643.

Pielot, M., Vradi, A., & Park, S. (2018). Dismissed! A detailed exploration of how mobile phone users handle push notifications. *MobileHCI '18: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, New York, NY, Article 3, 1–11. https://doi.org/10.1145/3229434.3229445

Rankin, C. H., Abrams, T., Barry, R. J., Bhatnagar, S., Clayton, D. F., Colombo, J., Coppola, G., Geyer, M. A., Glanzman, D. L., Marsland, S., McSweeney, F. K., Wilson, D. A., Wu, C.-F., & Thompson, R. F. (2009). Habituation revisited: An updated and revised description of the behavioral characteristics of habituation. *Neurobiology of Learning and Memory*, *92*(2), 135–138.

Reeves, L. M., Lai, J., Larson, J. A., Oviatt, S., Balaji, T. S., Buisine, S., ... & Wang, Q. Y. (2004). Guidelines for multimodal user interface design. *Communications of the ACM*, 47(1), 57-59.

Rosenfield, D., Jouriles, E. N., McDonald, R., & Mueller, V. (2014). Interparental conflict, community violence, and child problems: Making sense of counterintuitive findings. *American Journal of Orthopsychiatry*, *84*(3), 275–283.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro (ed.), *Theoretical issues in reading comprehension*. Lawrence Erlbaum.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Shirazi, A. S., Henze, N., Dingler, T., Pielot, M., Weber, D., & Schmidt, A. (2014). *Large-scale assessment of mobile notifications*. ACM.

Steelman, Z. R., Hammer, B. I., & Limayem, M. (2014). Data collection in the digital age: Innovative alternatives to student samples. *MIS Quarterly*, *38*(2), 355–378.

Thompson, R. F. (2009). Habituation: A history. *Neurobiology of Learning and Memory*, *92*(2), 127–134.

Thompson, R. F., & Spencer, W. A. (1966). Habituation: A model phenomenon for the study of neuronal substrates of behavior. *Psychological Review*, *73*(1), 16–43.

Turner, L. D., Allen, S. M., & Whitaker, R. M. (2019). The influence of concurrent mobile notifications on individual responses. *International Journal of Human-Computer Studies*, *132,* 70–80.

Vance, A., Brinton Anderson, B., Brock Kirwan, C., & Eargle, D. (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.

Vance, A., Jenkins, J. L., Anderson, B. B., Bjornn, D. K., & Kirwan, C. B. (2018). Tuning out security warnings: A longitudinal examination of habituation through fMRI, eye tracking, and field experiments. *MIS Quarterly*, *42*(2), 355–380.

Verizon. (2022). *2022 data breach investigations report*. https://www.verizon.com/business/resources/reports/2022/dbir/2022-data-breach-investigations-report-dbir.pdf

Weinberger, J., & Felt, A. P. (2016). A week to remember: The impact of browser warning storage policies. SOUPS '16: *Proceedings of the ninth symposium on usable privacy and security*, Denver, CO.

Weinert, C., Maier, C., Laumer, S., & Weitzel, T. (2022). Repeated IT Interruption: Habituation and Sensitization of User Responses. *Journal of Management Information Systems*, 39(1), 187–217.

West, R. (2008). The psychology of security. *Communications of the ACM*, *51*(4), 34–40.

Zeng, E., Li, F., Stark, E., Felt, A. P., & Tabriz, P. (2019). Fixing HTTPS misconfigurations at scale: An experiment with security notifications. *Workshop on the Economics of Information Security* (WEIS), Boston, MA.

APPENDIX A. MRI ACQUISITION AND ANALYSIS

MRI data were collected with a Siemens 3T Tim-Trio scanner with a 12-channel head coil. Structural MRI scans were acquired with an MP-RAGE sequence with the following parameters: 176 1-mm slices; TR, 1900 ms; TE, 2.3 ms; flip angle, 9°; 250-mm field of view; 256 × 256 acquisition matrix; voxel volume, $0.977 \times 0.977 \times 1$ mm^3. Functional data were acquired using an echo-planar imaging (EPI) sequence with the following parameters: 40 3-mm slices; TR, 2500 ms; TE, 28 ms; flip angle, 90°; 220-mm field of view; 64 × 64 acquisition matrix; and voxel volume, $3.4 \times 3.4 \times 3$ mm.^{3} The number of TRs collected varied according to participant completion time, with a minimum of 488 and maximum of 699.

MRI data were analyzed using the Analysis of Functional NeuroImages (AFNI) software package (version AFNI_20.1.10 "Otho") and *dcm2niix* (version v1.0.20180622; Li et al. 2016). All data and analysis scripts are available at the following links: [omitted for review].

Preprocessing was performed in accordance with the following steps: First, functional and structural data were converted from DICOM to NIfTI format using *dcm2niix* and were defaced for anonymity. The functional volume within each run with the least number of outlier voxel values was chosen as the base to which motion correction was performed for that run. Structural scans were also aligned with this functional volume and then skull stripped and warped into the MNI space using a nonlinear diffeomorphic transformation. The spatial transformations for motion correction and spatial normalization were concatenated and applied to the functional data in one step to minimize the number of data interpolations. Functional data resolution was maintained at the largest acquisition dimension (i.e., $3.4 \times 3.4 \times 3.4$ mm^3). Functional data were scaled using the mean of the signal within each voxel for each run. Functional volumes (TRs) with large motion events were excluded from the first-level regression analysis. Finally, coverage masks that excluded voxels with very low EPI signals were created. These coverage masks were combined with a gray-matter mask in the group analysis described below.

We performed two first-level single-subject regression analyses using the general linear model approach. In the first regression analysis, we examined trial-specific activation to obtain activation estimates for each separate presentation of the performance notification to test for habituation over multiple repetitions. This was performed by calculating a beta coefficient for each notification event (or "trial") separately using the iterative regression approach described by Mumford et al. (2012) and implemented in the AFNI program *3dLSS*. In this approach, a response model is estimated for each trial separately such that there is a single regressor for that trial and all other trials are combined into a second regressor. This is repeated iteratively for all trials, resulting in separate beta coefficients for each trial of interest. These trial-specific beta coefficients were blurred using an 8-mm FWHM Gaussian kernel and used to identify clusters that demonstrated habituation over multiple repetitions of the performance notifications in the group analysis (see below).

We also obtained activation estimates for the security warnings across all repetitions in a separate regression analysis, which included six regressors for motion (three translations and three rotations) in addition to polynomial regressors to code for low-frequency scanner drift. Behavioral regressors were created for the three styles of security warnings, for the performance notifications, and for the novel foils. Event onset times were convolved with the canonical hemodynamic response function. The image classification task served as the implicit baseline for the model, with variations in response time in the image classification task serving to provide jitter in the timing between the events of interest to avoid multicollinearity in the fMRI analysis (jitter median: 3.73 s, range: 2.75–78.98 s). Resulting beta coefficients were also blurred using an 8-mm FWHM Gaussian kernel.

Residuals from the single-subject regression analysis were also blurred with the same parameters, and the spatial smoothness of the noise (or the autocorrelation function) was estimated from these blurred residuals. The estimated smoothness of the data was then used to perform Monte Carlo simulations to determine corrections for multiple comparisons and control the family-wise error (FWE) rate. Given the observed smoothness in the functional data, obtaining a family-wise error (FWE < 0.01 with a voxel-wise threshold of $p < 0.0001$ would require a spatial extent threshold of just 6 voxels (with adjacent edges). However, given the strong linear effects observed in our group analysis, we chose more conservative thresholds to obtain separate clusters of activation. Our final thresholds were voxel-wise $p < 0.00001$, spatial extent $k > 40$ voxels with nearest-neighbor level 2 (adjacent edges), and two-sided thresholding, which represents a strict control for multiple comparisons.

The group-level analysis used AFNI program *3dRegAna* to perform a multiple linear regression analysis, with regressors coding for linear and quadratic trends across the first 15 repetitions of the performance notification in addition to regressors coding for each subject (i.e., subjects were treated as a random factor). Resulting activations were clustered according to the thresholds described above.

APPENDIX B. DETAILED MRI RESULTS

Table B1. Detailed MRI results by brain region

					Performance vs.							
		Peak voxel MNI			Novel foil		Extension		Slider		Drag and drop	
Label	**# of voxels**	***X***	***Y***	***Z***	***t*(23)**	***p***	***t*(23)**	***p***	***t*(23)**	***p***	***t*(23)**	***p***
Bilateral ventral visual stream	747	18.9	−93.6	−18.1	4.501	<0.001	0.287	0.777	2.047	0.052	4.432	<0.001
Left dorsal visual stream	238	−22.3	−83.3	50.7	5.812	<0.001	4.954	<0.001	6.447	<0.001	10.728	<0.001
Bilateral dorsomedial prefrontal cortex	181	−1.7	−0.8	71.3	5.004	<0.001	4.067	<0.001	2.649	0.014	4.939	<0.001
Right inferior frontal gyrus	156	49.8	23.2	−7.8	8.127	<0.001	6.21	<0.001	6.048	<0.001	9.198	<0.001