MIS Quarterly

TUNING OUT SECURITY WARNINGS: A LONGITUDINAL EXAMINATION OF HABITUATION THROUGH fMRI, EYE TRACKING, AND FIELD EXPERIMENTS[1]

Anthony Vance, Jeffrey L. Jenkins, Bonnie Brinton Anderson
Information Systems Department, Marriott School of Business, Brigham Young University,
Provo, UT 84602 U.S.A. {anthony@vance.name} {jeffrey_jenkins@byu.edu} {bonnie_anderson@byu.edu}

Daniel K. Bjornn
Department of Psychology, Brigham Young University,
Provo, UT 84602 U.S.A. {dbjornn@byu.edu} {kirwan@byu.edu}

C. Brock Kirwan
Department of Psychology and Neuroscience Center, Brigham Young University,
Provo, UT 84602 U.S.A. {kirwan@byu.edu}

Research in the fields of information systems and human-computer interaction has shown that habituation—decreased response to repeated stimulation—is a serious threat to the effectiveness of security warnings. Although habituation is a neurobiological phenomenon that develops over time, past studies have only examined this problem cross-sectionally. Further, past studies have not examined how habituation influences actual security warning adherence in the field. For these reasons, the full extent of the problem of habituation is unknown.

We address these gaps by conducting two complementary longitudinal experiments. First, we performed an experiment collecting fMRI and eye-tracking data simultaneously to directly measure habituation to security warnings as it develops in the brain over a five-day workweek. Our results show not only a general decline of participants' attention to warnings over time but also that attention recovers at least partially between workdays without exposure to the warnings. Further, we found that updating the appearance of a warning— that is, a polymorphic design—substantially reduced habituation of attention.

Second, we performed a three-week field experiment in which users were naturally exposed to privacy permission warnings as they installed apps on their mobile devices. Consistent with our fMRI results, users' warning adherence substantially decreased over the three weeks. However, for users who received polymorphic permission warnings, adherence dropped at a substantially lower rate and remained high after three weeks, compared to users who received standard warnings. Together, these findings provide the most complete view yet of the problem of habituation to security warnings and demonstrate that polymorphic warnings can substantially improve adherence.

Keywords: Security warnings, habituation, information security behavior, functional magnetic resonance imaging (fMRI), eye tracking, longitudinal experimental design, field experiment, mobile computing, NeuroIS

Introduction ▮▮▮

Research in the fields of information systems and human–computer interaction has shown that habituation—"decreased response to repeated stimulation" (Groves and Thompson 1970, p. 419)—is a serious threat to the effectiveness of security warnings. However, past studies share three critical limitations. First, they only examined habituation cross-sectionally (see Table 1). This technique is a substantial limitation because habituation is a neurobiological phenomenon that develops over time (Rankin et al. 2009). Further, a key characteristic of habituation is recovery—the increase of a response after a rest period in which the stimulus is absent (Rankin et al. 2009). Without a longitudinal design, it is not possible to examine whether recovery can sufficiently counteract the effect of habituation to warnings.

Second, past studies did not examine how habituation influences actual warning adherence in the field but instead used laboratory experiments that presented unrealistically high numbers of warnings to participants in a short laboratory session. Because users typically receive security warnings infrequently, presenting an artificially high number of warnings in a short laboratory session is too removed from reality to be ecologically valid (Straub et al. 2004). Consequently, for these reasons, the full extent of the problem of habituation is unknown.

Third, previous research (Anderson, Jenkins et al. 2016; Anderson, Vance et al. 2016b) proposed that repeatedly updating the appearance of a warning—that is, a polymorphic warning design—can be effective in reducing habituation. However, their findings were subject to the same limitations as above. Therefore, it is not clear (1) whether polymorphic warnings are effective over time or if users will quickly learn to ignore them and (2) whether the polymorphic design can actually lead to better security warning adherence.

In this article, our objective is to address these gaps by conducting two complementary longitudinal experiments. First, we conducted an experiment collecting fMRI and eye-tracking data simultaneously to directly measure habituation to security warnings as it develops in the brain over a five-day workweek. Our results not only showed a general decline of participants' attention to warnings over time but also that attention recovers at least partially between workdays without exposure to the warnings. Unfortunately, this recovery is not sufficient to compensate for the overall effect of habituation over time. However, we found that updating the appearance of a warning—that is, a polymorphic design—substantially reduced habituation of attention.

Second, we performed a three-week field experiment in which users were naturally exposed to privacy permission warnings as they installed apps on their mobile devices. Consistent with our fMRI results, users' warning adherence substantially decreased over the three weeks. However, for users who received polymorphic permission warnings, adherence dropped at a substantially lower rate and remained at a high rate after three weeks compared to users who received standard warnings. Together, these findings provide the most complete view yet of the problem of habituation to security warnings and demonstrate that polymorphic warnings can substantially improve warning adherence.

This paper proceeds as follows. After a brief review of habituation theory and recovery, we present the method, analysis, and discussion of Experiments 1 and 2. Finally, we integrate the results from both experiments in a general discussion and conclusion.

Habituation Theory and Hypotheses ▮▮

We developed our hypotheses around the two most prevalent characteristics of habituation: (1) *response decrement*, an attenuation of a response with multiple exposures, and (2) *response recovery*, the increase of a response after a rest period in which the stimulus is absent (Rankin et al. 2009). Hypotheses 1 and 2 explore how users' responses to security warnings weaken over repeated viewings and how polymorphic warnings (described below) can mitigate this effect. Hypotheses 3 and 4 explore how users' responses to warnings recover after the warning is withheld and how polymorphic warnings enhance this recovery. For further background on habituation, see Appendix A.

Response Decrement

Users experience a response decrement when exposed to repeated warnings. This response decrement may include paying less attention and responding less thoughtfully to the warning. Dual-process theory (Groves and Thompson 1970) explains that upon initial exposure to a stimulus, a mental model is created, and that when people see the same stimulus again, it is automatically and unconsciously compared to this model. If the models are similar, the behavioral responses to the stimulus are inhibited in favor of reliance on the mental model instead (Thompson 2009). In the context of security warnings, users will unconsciously compare subsequent warnings to their mental model of warnings they have seen previously. If users unconsciously determine that a warning is similar, they will give less attention to it after repeated ex-

Table 1. Review of Previous Research on Habituation to Security Warnings				
Citation	**Time Scale**	**Experiment Type**	**Measure of Habituation**	**Findings**
Anderson, Vance et al. (2016b)	Cross-sectional	Laboratory	Habituation as a mental state as measured by fMRI	Used fMRI to measure habituation in terms of decreased neural activity. Compared habituation to traditional warnings to polymorphic warnings that change their appearance and showed that decreases in neural activity in the visual processing center of the brain were less for polymorphic warnings compared to traditional warnings in a single laboratory session.
Anderson, Jenkins et al. (2016)	Cross-sectional	Laboratory	Habituation as a mental state as measured by eye tracking	Used eye tracking to measure habituation in terms of decreases in eye-gaze fixations. Showed that fixations decreased less for polymorphic warnings compared to static warnings in a single laboratory session.
Bravo-Lillo et al. (2013)	Cross-sectional	Amazon Mechanical Turk	Warning adherence	Measured habituation in terms of the percentage of users who immediately recognized that the contents of a dialog message changed after a rapid habituation period. Only 14% of users immediately recognized the change in the dialog message.
Bravo-Lillo et al. (2014)	Cross-sectional	Amazon Mechanical Turk	Warning adherence	Examined four different levels of frequency of exposure to a dialog message. Found that increasing the frequency with which a dialog was displayed caused a threefold decrease in the proportion of users who immediately recognized a change in the dialog message.
Brustoloni and Villamarín-Salomón (2007)	Cross-sectional	Laboratory	Inferred	Found that compared to conventional warnings, a security warning that randomized the position of its option buttons resulted in users ignoring the message less frequently in risky situations. Attributed this difference to a reduced habituation effect.
Egelman et al. (2008)	Cross-sectional	Laboratory	Inferred	Found a correlation between users disregarding warnings and recognizing warnings as previously viewed. Attributed this correlation to habituation.
Egelman and Schechter (2013)	Cross-sectional	Laboratory	Inferred	Found a correlation between users disregarding warnings and recognizing warnings as previously viewed. Attributed this correlation to habituation.
Krol et al. (2012)	Cross-sectional	Laboratory	Inferred	Observed that 81% of participants ignored a download warning. Of these, 56% later claimed they ignored the warning because of daily exposure to warnings in general.
Schechter et al. (2007)	Cross-sectional	Laboratory	Inferred	Presented participants with increasingly alarming security warnings in an online-banking experimental task. Despite this, the majority of users continued to ignore the warnings. Subsequent studies (e.g., Bravo-Lillo et al. 2013) pointed to this finding as evidence of habituation.
Sotirakopoulos et al. (2011)	Cross-sectional	Laboratory	Inferred	Experimentally compared adherence to a variety of web-browser SSL warnings. Observed no significant differences across treatments (i.e., adherence was poor in every treatment). Attributed this finding to habituation.
Sunshine et al. (2009)	Cross-sectional	Laboratory	Inferred	Observed that participants remembered their responses to previous interactive security warnings and applied them to new warnings, even if the level of risk or context changed. Attributed this result to habituation.

posures and rather rely on the mental model within a computing session and over time (Rankin et al. 2009). In summary, we hypothesize the following:

> H1a.   *Users habituate to warnings within a computing session.*

> H2a.   *Users habituate to warnings in computing sessions across days.*

We hypothesize that users will habituate more slowly to polymorphic warnings—warnings that change their appearance with repeated exposures (Anderson, Vance et al. 2016b)—than to static warnings. Wogalter (2006, p. 55) states that "habituation can occur even with well-designed warnings.... Where feasible, changing the warning's appearance may be useful in reinvigorating attention switch previously lost because of habituation." In terms of dual-process theory, users compare the current stimulus with a mental model of the stimulus as it was previously experienced. If a new or changed stimulus is experienced that does not match the mental model, users will give attention to the novel stimulus and the response strength will recover (Sokolov 1963). This process is known as *sensitization* (Groves and Thompson 1970) and counteracts habituation (Rankin et al. 2009). Consequently, changing the appearance of a warning will cause sensitization, and users will habituate less to polymorphic warnings at both neural and behavioral levels (Barry 2009).

We predict that polymorphic warnings will engender sensitization, reducing habituation within a single computing session as well as between computing sessions over multiple days. Within a computing session, polymorphic warnings slow habituation because the warning does not match the brain's mental model of the warning, resulting in a weaker and less stable mental model of the warning in the future (Cooke et al. 2015; Thompson 2009). Consequently, when users encounter a polymorphic warning in a future computing session, it may contradict this weaker mental model and be perceived as novel. In summary, we hypothesize the following:

> H1b.   *Users habituate less to polymorphic warnings than to static warnings within a computing session.*

> H2b.   *Users habituate less to polymorphic warnings than to static warnings across days.*

Response Recovery

Although users will habituate to warnings, we predict that they will partially recover from the habituation after a day's

rest period without seeing warnings. As habituation is a form of learning, it is subject to normal forgetting processes over time (Staddon 2005; Wixted 2004). When a warning is withheld for a day, the mental model of the warning will become weaker. When users see this warning in the future, it will be less likely to match the mental model and will appear novel. In response to this novelty, the response strength will recover and the sensitization process will increase a person's attention to the warning, countering habituation (Terry 2015).

Although the mental model diminishes with time, it is unlikely to fade completely within a single day. The brain will still inhibit the behavioral response to the stimulus and habituation will occur. However, this response inhibition or habituation is likely to be weaker when users see a warning after it has been withheld for a day compared to when they see it repeatedly within a single computing session (Rankin et al. 2009). In summary, we hypothesize the following:

> H3a.   *If warnings are withheld after habituation occurs, the response recovers at least partially the next day.*

We predict that the amount of recovery from day to day will be greater for polymorphic warnings than for static warnings. As previously discussed, the mental models of polymorphic warnings are weaker and less stable than the models of static warnings. Less stable mental models fade more quickly than stable models (Rankin et al. 2009; Wagner 1976). Thus, after users have not seen a warning for a day, they are more likely to perceive the polymorphic warning as more novel than static warnings and recovery will increase.

Further, as the polymorphic warning continues to change its appearance from the previous day, it is even more likely to differ from the existing mental model, weakening the behavioral inhibition, increasing sensitization, and enhancing response recovery (Rankin et al. 2009). Conversely, with static warnings, response recovery will be weaker because the mental model is more robust, reinforced by repeatedly seeing the same warning on previous days (Grill-Spector et al. 2006; Groves and Thompson 1970). The behavioral response will be inhibited to a greater degree, and habituation will be more influential (Rankin et al. 2009). In summary, we hypothesize the following:

> H3b.   *If warnings are withheld after habituation occurs, response recovery is stronger for polymorphic warnings than for static warnings the next day.*

We predict that the amount of recovery between days will decrease as users continue to see security warnings day after day. For example, if users see a warning three days in a row,

their response recovery will be weaker between days 2 and 3 compared to between days 1 and 2, and so on. This weakening recovery occurs because users' mental models of the warning become stronger and more stable as they view warnings across additional computing sessions (Cooke et al. 2015; Groves and Thompson 1970). As a result, the behavioral response will be inhibited to a greater degree each successive day and the degree of habituation to the warning will be greater. In summary, we hypothesize the following:

H4a. *The amount of recovery will decrease across days.*

Finally, we predict that response recovery across days will decrease less for polymorphic warnings than for static warnings. The mental models of polymorphic warnings will be less accurate and stable than the mental models of static warnings. Because polymorphic warnings change their appearance, the neural response will not be inhibited (Sokolov 1963; Wagner 1979) and sensitization will counter the habituation (Groves and Thompson 1970). Therefore, the amount of recovery each day will be greater for polymorphic warnings than for static warnings, and response recovery will decrease less for polymorphic warnings than for static warnings (Sanderson and Bannerman 2011). In summary, we hypothesize the following:

H4b. *The amount of recovery will decrease less for polymorphic warnings than for static warnings across days.*

Polymorphic Warning Design

Anderson, Vance et al. (2016b) developed 12 polymorphic warning artifacts based on an extensive review of the warning-science literature. Through testing of the different polymorphic variations using fMRI data, they found four of the variations maintained attention better than the rest: (1) including a pictorial symbol, (2) changing the warning's background color to red, (3) using a "jiggle" animation when the warning appears, and (4) using a zoom animation to make the warning increase in size. Figure 1 shows each variation for one sample warning with its supporting sources. Given this support, we used these four variations for our polymorphic warning to test our hypotheses.

We evaluate the effectiveness of polymorphic warnings over time using NeuroIS. NeuroIS can be an effective approach for the evaluation of IT artifacts (vom Brocke and Liang 2014), and it has previously been shown to be effective in examining security warnings specifically (Jenkins et al. 2016; Vance et al. 2014). Riedl, Davis, and Hevner (2014, p. ii)

explain that NeuroIS measures are beneficial "to the design of ICT artifacts." Riedl, Banker et al. (2010, p. 250) explain that "IS researchers could use the theory of controlled and automatic brain processes to . . . allow for a better design of IT artifacts and other interventions." Further, Dimoka et al. (2011) argued that NeuroIS measures should be used as dependent variables in evaluating IT-artifact designs:

> Rather than relying on perceptual evaluations of IT artifacts, the brain areas associated with the desired effects can be used as an objective dependent variable in which the IT artifacts will be designed to affect (p. 700).
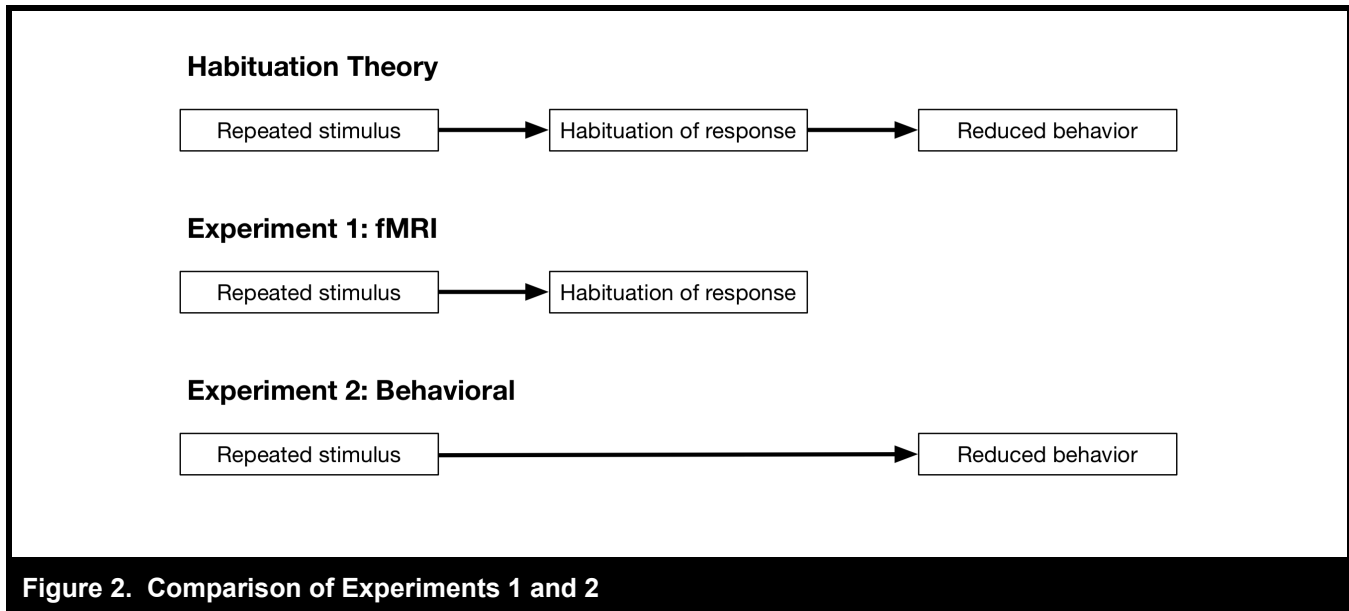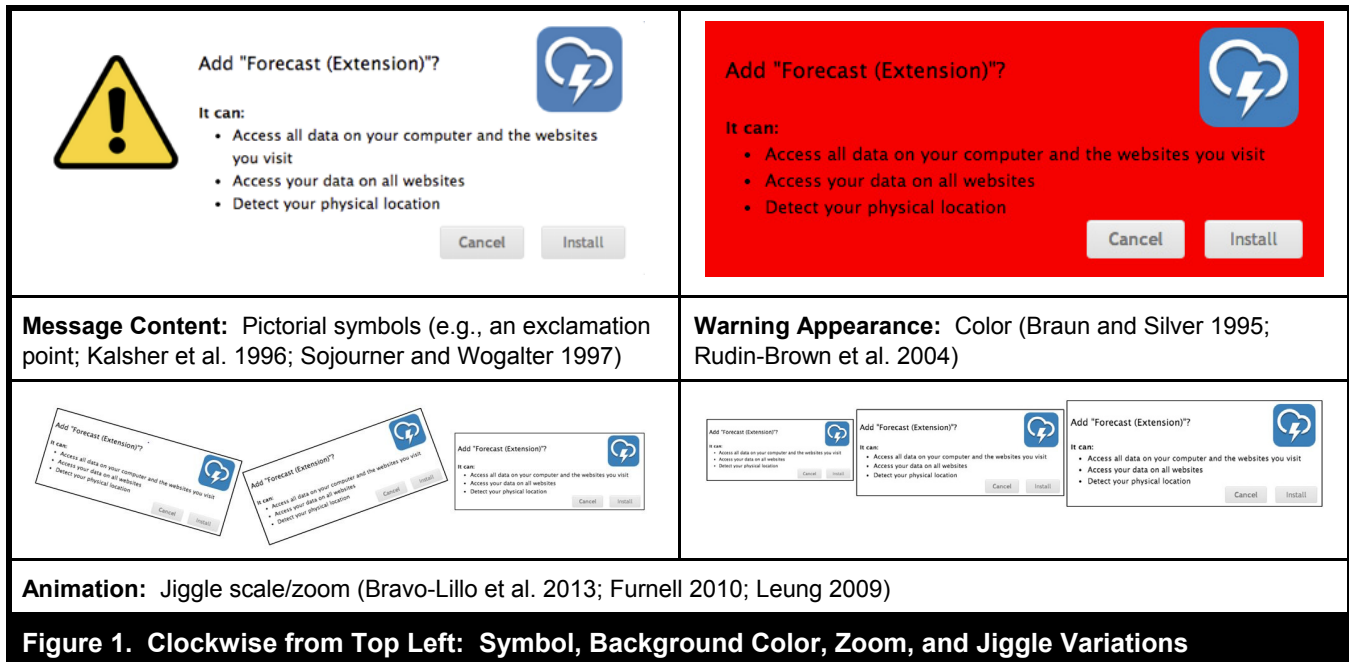
We used this approach to evaluate our polymorphic warning design.

Methodology

To test our hypotheses, we conducted two complementary experiments. Using fMRI and eye tracking, Experiment 1 directly measured habituation in the brain over time with repeated exposure to warnings. However, it did not measure actual behavior and had various ecological validity limitations (described in the "Discussion" section). In contrast, Experiment 2 measured how warning adherence decreases over time with repeated exposure to warnings in a more ecologically valid field experiment. However, Experiment 2 did not directly measure habituation in the brain. The results of both experiments, therefore, evaluate habituation theory from different perspectives (see Figure 2). This is consistent with the observation that "no single neurophysiological measure is usually sufficient on its own, and it is advisable to use many data sources to triangulate across measures" (Dimoka et al. 2012, p. 694). Together, Experiments 1 and 2 provided a powerful evaluation of habituation and its effect on behavior, and whether polymorphic warnings, as an IT artifact, can substantially reduce habituation and improve warning adherence over time.

Experiment 1: fMRI

In Experiment 1, we built on prior fMRI habituation research by examining habituation in the brain in response to static and polymorphic warnings through repeated exposure to the stimuli over the course of a workweek. In addition, we sought to understand potential response recovery associated with the rest periods within the time frame of the experiment.

Message Content: Pictorial symbols (e.g., an exclamation point; Kalsher et al. 1996; Sojourner and Wogalter 1997)	**Warning Appearance:** Color (Braun and Silver 1995; Rudin-Brown et al. 2004)

| **Animation:** Jiggle scale/zoom (Bravo-Lillo et al. 2013; Furnell 2010; Leung 2009) | |

Figure 1. Clockwise from Top Left: Symbol, Background Color, Zoom, and Jiggle Variations

Figure 2. Comparison of Experiments 1 and 2

Experiment 1: Context

To test our hypotheses, we conducted a multimethod NeuroIS experiment, simultaneously collecting both fMRI and eye-tracking data. This allowed us to capitalize on the strengths of each method while mitigating their limitations (Venkatraman et al. 2015). We used these methods to directly measure the effect of the IT artifact (security warnings) on the underlying neurocognitive process of habituation.

A neural manifestation of habituation to visual stimuli in the brain is called repetition suppression (RS)—the reduction of neural responses to stimuli that are repeatedly viewed, which is a robust indicator of habituation (Grill-Spector et al. 2006). We used the differential RS effect in various brain regions to map sensitivity to repetitive security warning stimuli. In our case, the high spatial resolution afforded by fMRI (Dimoka 2012) was important because it allowed us to disentangle RS effects from sensory adaptation or fatigue effects (Rankin et al. 2009).

Concurrent with the fMRI scan, we used an eye tracker to measure the eye-movement memory (EMM) effect—another robust indicator of habituation (Ryan et al. 2000). The EMM effect manifests in fewer eye-gaze fixations and less visual sampling of the regions of interest within the visual stimulus. Memory researchers have discovered that the EMM effect is a pervasive phenomenon in which people unconsciously pay less attention to images they have viewed previously (Smith and Squire 2017). With repeated exposure, the memories become increasingly available, thus requiring less visual sampling of an image (Heisz and Shore 2008).

One strength of eye tracking is its temporal resolution, which allows researchers to measure with millisecond precision the attentional process of participants' responses to repeated stimuli. Thus, fMRI (with high spatial resolution) and eye tracking (with high temporal resolution) complement each other, measuring both a behavioral manifestation of attention (i.e., eye movements) and the neural activity that drives attention.

Experiment 1: Method

Participants

We recruited 16 participants from a large U.S. university (8 male, 8 female); this number of participants is consistent with other fMRI studies (Dimoka et al. 2012). Participants were between 19 and 29 years of age (mean 23.3 years), were right-handed, were native English speakers, had normal or corrected-normal visual acuity, and were primarily Microsoft Windows users. One subject was excluded from the study due to a scanner malfunction, resulting in 15 total participants (8 male, 7 female).[2] Each participant engaged in five fMRI scans: one scan at the same time each day for five consecutive days. Upon arrival, participants were screened to ensure MRI compatibility. They were then given instructions about the task and placed in the scanner. Each scan lasted 30 minutes, beginning with a structural scan and followed by two functional scans that displayed the warnings and images.

---

[2]We conducted a pilot study (Anderson et al. 2014) that revealed a large estimated effect size for the repetition effect (partial eta^2 = .7). Using this estimated effect size, an *a priori* power analysis indicated that we would need four subjects to achieve power greater than .8. The low sample size required for adequate power is due to the large effect size and is not typical of NeuroIS research, where samples of 15 to 25 participants are common. For example, Anderson, Vance et al. (2016b), *n* = 25; Jenkins et al. (2016), *n* = 24; Dimoka (2010), *n* = 15; Riedl, Hubert, and Kenning (2010), *n* = 20; Riedl, Mohr et al. (2014), *n* = 18; Warkentin et al. (2016), *n* = 17.

Experiment Design

For each participant, warning stimuli were randomly assigned to conditions that remained the same across the five-day experiment. In addition, the order of the presentation of the warning stimuli was randomized per participant, per day, as described below and pictorially in Figure B1 of Appendix B. First, 40 images of various computer-security warnings (e.g., browser malware and SSL warnings, antivirus software warnings, and software signing errors) were randomly split into two pools: one for the static condition and the other for the polymorphic condition. The 20 warnings in the static condition were repeated four times at random times each day over the course of the study (20 static warnings × 4 repetitions = 80 static warning images to display).

The four polymorphic variations were then applied to the 20 warnings in the polymorphic pool so that each of the polymorphic variations randomly appeared once a day during the experiment. For each polymorphic warning, all four polymorphic variations were shown each day (20 polymorphic warnings × 4 variations = 80 polymorphic images to display). The order of these variations was randomized so that each day had a different presentation order for the variations of each warning.

Next, 120 images of general software (e.g., Windows Explorer, Control Panel, Microsoft Outlook) were also randomly split into two sets. The first set consisted of 20 images that were randomly repeated four times each day (20 general software images × 4 repetitions = 80 general software images to display). The remaining 100 images were divided evenly across days so that a new set of 20 images were randomly displayed each day. These images were used to create a baseline of unique presentations throughout the task. By comparing the responses for each repeated image to the unique baseline images, we could distinguish the habituation effect from attention decay attributable to participants' fatigue over time.

Upon completion of the randomization, there were 260 images to be presented each day (80 static images + 80 polymorphic images + 80 general software images + 20 new general software images for the day = 260). These 260 images were randomly displayed across two blocks of 7.7 minutes each, with a two-minute break in between blocks. Images were displayed for three seconds each, with a 0.5-second interstimulus interval. When participants saw the warnings, they were required to rate the severity of the content of the warning on a four-point scale. They did this using an MRI-compatible button-input device. The purpose of this task was to help keep the participant engaged in a context relevant to the warnings.

Protocol

Scan sessions occurred at the same time each day over a period of five days for each participant, resulting in five scans per participant. Upon arrival at the facility, participants completed a screening form to ensure MRI compatibility. Participants were verbally briefed about the MRI procedures and the task and were then situated lying on their backs in the scanner. Visual stimuli were viewed using a mirror attached to the head coil; this reflected a large monitor outside the scanner that was configured to display images in reverse so they appeared normal when viewed through the mirror.

We first performed a 10-second localizer scan, followed by a 7-minute structural scan. Following these scans, we started the experimental task. We used SR Research Experiment Builder software to display the stimuli and synchronize the display events and scanner software. The total scan time was 26.6 minutes for each day. Upon completion of the scan on days 1–4, participants were thanked and reminded of the next day's scan. After the completion of the scan on day 5, participants were again thanked, debriefed, and given $60 compensation. All *ex post* tests revealed that no subjects needed to be excluded (e.g., due to abnormalities or excessive movement).

Experiment 1: Analysis ■■■■

We analyzed each hypothesis separately for the fMRI and eye-tracking data. We describe each analysis below, followed by the testing of our hypotheses.

fMRI Analysis

MRI data were analyzed with the Analysis of Functional NeuroImages (AFNI) suite of programs (Cox 1996). Details of our scans and procedures can be found in Appendix B. Whole-brain, multivariate model analyses were conducted on the fMRI data to identify significant clusters of activation, or regions of interest (ROIs), that demonstrated activation consistent with the hypothesized pattern. All of our hypothesis tests utilized the same ROIs. Figures 3 and 4 graph the activation in two of the brain regions for polymorphic and static warnings analyzed across repetitions and across days.

Eye-Tracking Analysis

Eye-tracking data were collected using an MRI-compatible SR Research EyeLink 1000 Plus (see Figure B2 of Appendix B). Eye-fixation data were processed with DataViewer software (SR Research Ltd., version 1.11.900) to identify fixations and saccades. *Saccades* were defined as eye movements that met three different criteria: eye movement of at least .1°, velocity of at least 30°/second, and acceleration of at least 8,000°/second. *Fixations* were defined as periods of time between the saccades that were not also part of blinks. Fixation count was used as the dependent variable in each analysis.[3]

The number of fixations for polymorphic and static warnings per warning repetition per day is shown in Figure 5. The mean and standard deviations of fixation count and fixation duration per day are shown in Table 2. Some of the polymorphic warnings were animated, which prevented participants from fixating upon the warning during the animation. To control for this, we normalized all intercepts to zero and controlled for warning type in the analysis, allowing for individual warning intercepts. This control allowed us to focus on and accurately analyze how fixations change over time as an indicator of habituation.

H1a Analysis: Users Habituate to Warnings Within an Experimental Session

fMRI Analysis: We conducted a whole-brain, multivariate model analysis (Chen et al. 2014) on the fMRI data with sex, day, repetition number, and stimulus type (static warning and polymorphic warning) as fixed factors. We then conducted a linear trend analysis on repetition number, collapsing across days and stimulus types. We identified six significant ROIs where neural responses demonstrated a linear decrease in activation across repetitions (i.e., regions that demonstrated habituation). These regions are listed in Table 3a. Of particular note are the left and right ventral visual processing streams, both of which exhibited robust habituation effects [left: $F(1,597) = 17.71, p < .001$; right: $F(1, 597) = 20.49, p < .001$]. Accordingly, the fMRI data analysis in both right and left ventral visual processing streams supported H1a (see Figures 6 and 7).

Eye-Tracking Analysis: In a linear mixed-effects model, we included fixation count as the dependent variable and the subject ID, day number, and warning ID as random factors. The presentation number was treated as a fixed factor, and

---

[3]We chose fixation count as a more appropriate measure of habituation than fixation duration because the warning stimuli were displayed to subjects for the same duration. However, we also replicated all analyses using fixation duration as the dependent variable. This analysis is reported in Appendix C. The hypothesis testing results are the same as those using fixation count as the dependent variable.
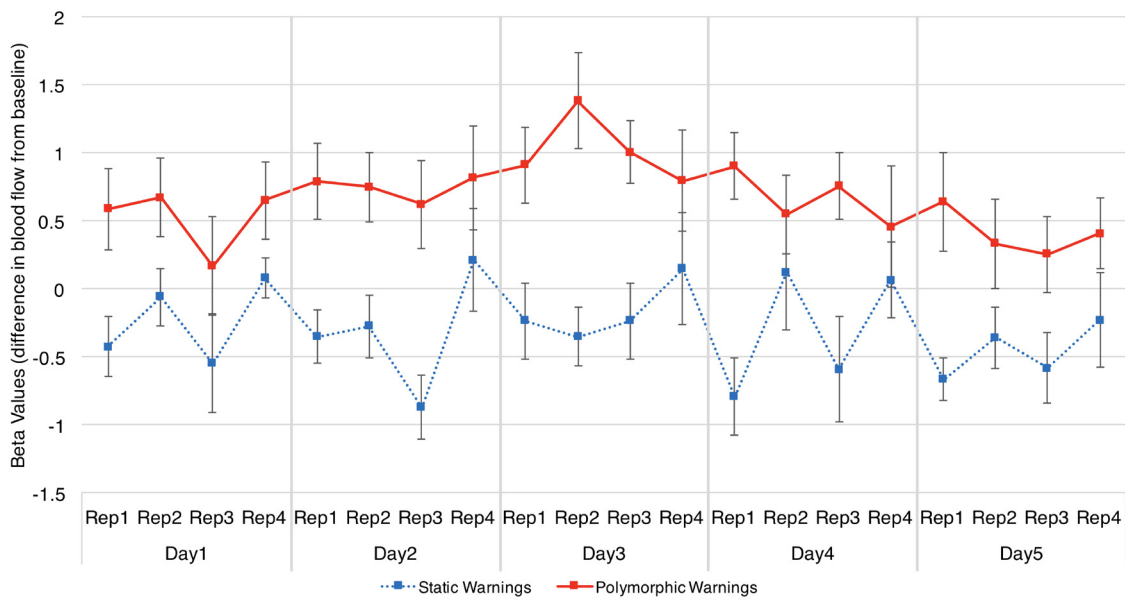
Beta values were extracted from a whole-brain analysis for each subject and then averaged across subjects according to stimulus condition.

Figure 3. Activity in the Right Inferior Temporal Gyrus in Response to Each Presentation of Static and Polymorphic Warnings
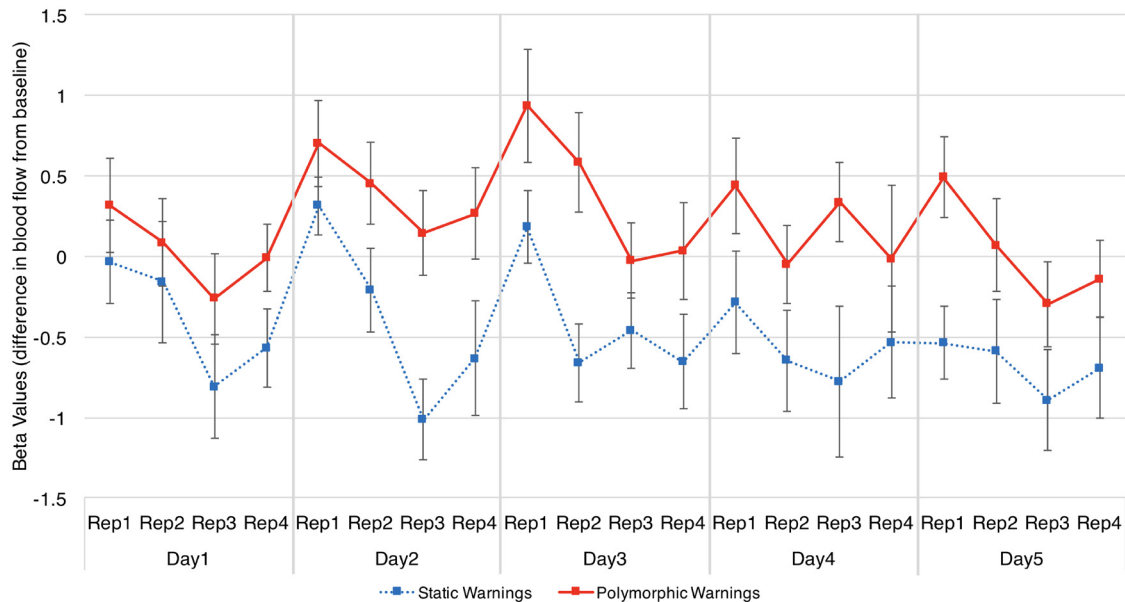
Beta values were extracted from a whole-brain analysis for each subject and then averaged across subjects according to stimulus condition.

Figure 4. Activity in the Right Ventral Visual Pathway in Response to Each Presentation of Static and Polymorphic Warnings
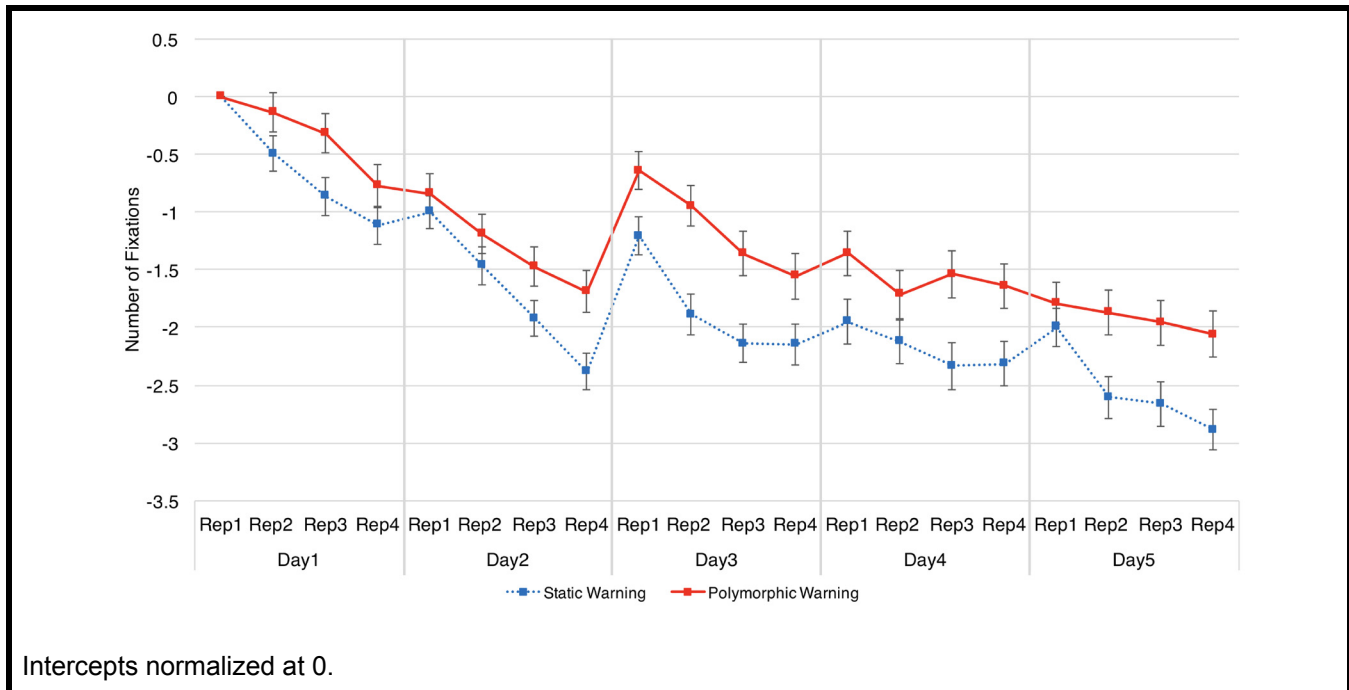
Figure 5. Change in Eye-Gaze Fixations Across Viewings

Table 2. Absolute Fixation Count and Fixation Duration by Day

	Day 1	Day 2	Day 3	Day 4	Day 5
Fixation count mean milliseconds (ms)	9.1	8.08	8.09	7.71	7.35
Fixation count *SD* (ms)	2.65	2.18	2.27	2.48	2.32
Fixation duration mean (ms)	2349.55	2204.45	2135.69	2113.39	2081.29
Fixation duration SD (ms)	450.94	325.25	384.14	441.1	444.35

visual complexity[4] was included as a covariate. The analysis supported H1a: The presentation number beta was significantly negative, indicating that habituation had occurred: $\chi^2(1, N = 11,976) = 212.58$, $p < .001$, $\beta = -0.254$. In addition, visual complexity was significant: $\chi^2(1, N = 11,976) = 35.38$, $p < .001$, $\beta = 0.4533$. The R^2 of the model was .26.

H1b Analysis: Users Habituate Less to Polymorphic Warnings than to Static Warnings Within an Experimental Session

fMRI Analysis: We examined the previous whole-brain, multivariate model analysis to identify ROIs where there were

significant interactions between stimulus type and the linear trend over repetition number. We identified eight ROIs (see Table 3b). Beta values were extracted for these regions and tested using a within-subjects, repeated-measures ANOVA. The stimulus type × repetition number interaction supported H1b.

Eye-Tracking Analysis: We specified the same mixed-effects model as in H1a, except that we included an interaction term between presentation number and whether or not the warning was polymorphic. The eye-tracking analysis supported H1b. Both main effects for presentation number [$\chi^2(1, N = 11,976)$ = 159.95, $p < .001$, $\beta = -0.309$] and polymorphism [$\chi^2(1, N = 11,976) = 78.89$, $p < .001$, $\beta = -0.849$] were significant, as were the interaction [$\chi^2(1, N = 11,976) = 10.15$, $p < .001$, $\beta = 0.1096$] and visual complexity [$\chi^2(1, N = 11,976) = 82.79$, $p < .001$, $\beta = -0.4734$]. The R^2 of the model was .266.

---

[4]Visual complexity was calculated through a script in MATLAB (Rosenholtz et al. 2007).

Table 3. Regions of Interest (ROIs) for Habituation Within Experimental Sessions

(a) ROIs for Main Effect of Repetition

Region	# Voxels	Peak x	Peak y	Peak z	*F* Value	*p* Value
R. ventral visual stream	569	-28	58	-15	20.49	< .001
L. inferior frontal gyrus	479	49	-10	33	29.87	< .001
L. ventral visual stream	418	28	88	18	17.71	< .001
B. dorsomedial prefrontal cortex	111	-1	-7	48	17.75	< .001
R. inferior frontal gyrus	101	-49	-10	33	15.11	< .001
L. posterior middle temporal gyrus	90	55	46	6	16.23	< .001

(b) ROIs for Repetition by Stimulus-Type Interaction

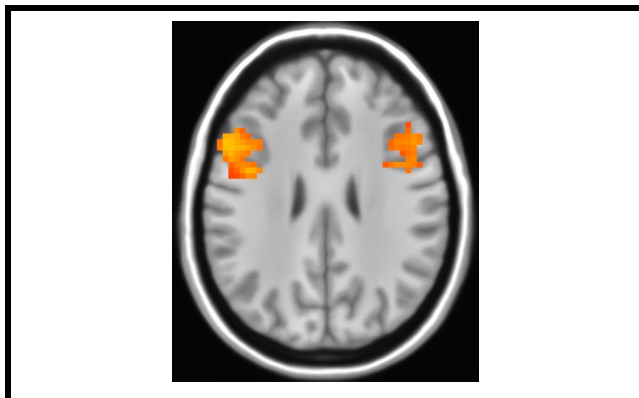| Region	# Voxels	Peak x	Peak y	Peak z	*F* Value	*p* Value
R. superior temporal gyrus	307	-61.5	43.5	9	7.23	< .001
R. anterior insula	255	-46.5	-40.5	6	5.66	.02
R. medial frontal gyrus	76	-1.5	-37.5	33	4.64	.03
L. anterior insula	70	31.5	-13.5	-12	5.73	.02
R. inferior temporal gyrus	70	-46.5	67.5	0	4.91	.03
L. superior temporal gyrus	54	52.5	40.5	21	4.61	.03
L. middle frontal gyrus	53	43.5	-34.5	27	3.87	.05
R. middle frontal gyrus	51	-31.5	-34.5	42	3.08	.08

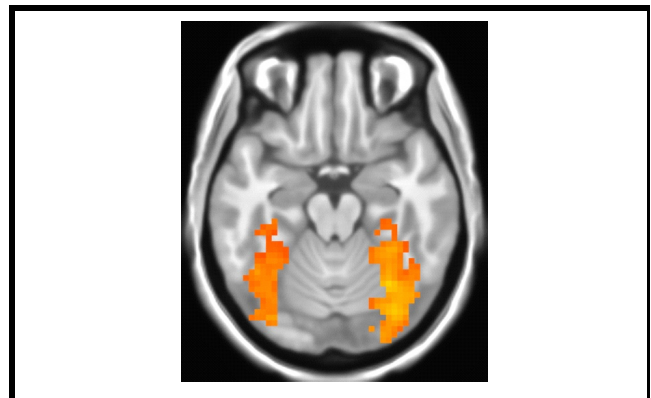**Figure 6. Left and Right Inferior Frontal Gyri**

Figure 7. Left and Right Ventral Visual Pathways

Table 4. Regions of Interest for Habituation Across Days

ROIs for Main Effect of Day

Region	# Voxels	Peak x	Peak y	Peak z	*F* Value	*p* Value
R. insula	160	-43	-16	3	67.87	< .001
L. insula	158	40	-16	0	86.19	< .001

ROIs for Day by Stimulus-Type Interaction

Region	# Voxels	Peak x	Peak y	Peak z	*F* Value	*p* Value
L. middle frontal gyrus	190	49	-31	18	5.19	.02
L. middle occipital gyrus	118	25	76	39	4.70	.03

H2a Analysis: Users Habituate to Warnings Across Days

fMRI Analysis: We conducted a whole-brain, multivariate model analysis to find areas that responded to a linear trend on day number, collapsing across repetitions and stimulus types. This analysis identified two main ROIs: the right and left insula. To quantify the extent of the decrease in these ROIs, beta values were extracted for these regions and tested using a within-subjects, repeated-measures ANOVA. Both the right [F (1,597) = 67.87, $p < .001$] and left insula [F (1,597) = 86.19, $p < .001$] displayed a significant habituation effect across days (see Table 4). Thus, the fMRI analysis supported H2a.

Eye-Tracking Analysis: In a linear mixed-effects model, we included fixation count as the dependent variable and the subject ID and warning ID as random factors. The presentation number (across days) was treated as a fixed factor, and visual complexity was included as a covariate. The eye-tracking analysis supported H2a; the beta of presentation number across days was significantly negative [$\chi^2(1, N = 11,976) = 212.89, p < .001, \beta = -.1031$], indicating habituation. Visual complexity was also significant [$\chi^2(1, N = 11,976) = 34.85, p < .001, \beta = 0.3815$]. The R^2 of the model was .13.

H2b Analysis: Users Habituate Less to Polymorphic Warnings than to Static Warnings Across Days

fMRI Analysis: We conducted a whole-brain analysis for a day-by-stimulus-type interaction. Two ROIs, the left middle frontal gyrus [$F(1,595) = 5.188, p < .05$] and left middle occipital gyrus [$F(1,595) = 4.697, p < .05$], displayed a significant habituation interaction across days and between stimulus types (see Table 4).

Eye-Tracking Analysis: We specified the same mixed-effects model as in H2a, except that we included an interaction term between presentation number (across days) and whether or not the warning was polymorphic (coded as 1 for polymorphic and 0 for static). The eye-tracking analysis supported H2b; the interaction between presentation number and polymorphic warning type was significantly positive [$\chi^2 (1, N = 11,976) = 10.70, p < .001, \beta = 0.024$], indicating that participants habituate less to polymorphic warnings across days than static warnings. Both main effects for presentation number [$\chi^2 (1, N = 11,976) = 493.42, p < .001, \beta = -0.115$] and polymorphism [$\chi^2 (1, N = 11,976) = 64.71, p < .001, \beta = -0.725$] were also significant. Visual complexity, however, was not significant: $\chi^2 (1, N = 11,976) = 0.17, p > .05, \beta = 0.026$. The R^2 of the model was .137.

H3a Analysis: If Warnings Are Withheld After Habituation Occurs, the Response Recovers at Least Partially the Next Day

fMRI Analysis: We first calculated recovery scores by subtracting the mean beta value of the last display of each stimulus type from the first display of that stimulus type on the following day (i.e., Day 2 Display 1 – Day 1 Display 4, etc.). A whole-brain, multivariate model analysis was then conducted to test for regions that displayed changes from baseline activation, which, collapsing across days, revealed four ROIs in which significant recovery occurred (see Table 5). Note that the ROIs identified in this analysis overlapped considerably with those identified in the analysis of H1a. *Post hoc* analysis comparing specific days showed significant recovery for days 2–4 in nearly every area, with no significant recovery on day 5 (see Table 5). Thus, H3a was supported by the fMRI data.

Eye-Tracking Analysis: We subtracted the fixation count for the first viewing of a warning on one day from the fixation count for the last viewing of the warning on the previous day. We then conducted a *t*-test to test this hypothesis. The analysis supported H3a: participants experienced significantly positive recovery ($m = 0.369, SD = 3.171$) from day to day: $t(2377) = 5.672, p < .001, d = 0.233$.

H3b Analysis: If Warnings Are Withheld After Habituation Occurs, Response Recovery Is Stronger for Polymorphic Warnings than for Static Warnings the Next Day

fMRI Analysis: We analyzed the ROIs found for H3a but added stimulus type (polymorphic or static) to the model as a factor. None of the regions displayed a significant recovery by stimulus-type interaction (see Table 5); thus, H3b was not supported.

Eye-Tracking Analysis: We subtracted the fixation count for the first viewing of a warning on one day from the fixation count for the last viewing of the warning on the previous day. We specified a linear mixed-effects model that tested whether warning type (polymorphic or static) predicted this difference. The subject ID, day interval (e.g., the difference between day 1 and day 2 was coded as 1), and warning ID were included as random factors. Polymorphism was included as a fixed factor, and visual complexity was included as a covariate. The eye-tracking analysis did not support H3b. Neither the warning type [$\chi^2 (1, N = 2,400) = 1.92, p > .05, \beta = -0.166$] nor visual complexity [$\chi^2 (1, N = 2,400) = 1.16, p > .05, \beta = 0.072$] significantly predicted recovery between days.

Table 5. Regions of Interest (ROIs) for Recovery											
				Recovery > 0		**Recovery by Stimulus-Type Interaction**		**Recovery Across Days**		**Day by Stimulus-Type Interaction**	
Region	# Voxels	Peak	Peak	Peak	*t* Value	*p* Value	*F* Value	*p* Value	*F* Value	*p* Value	*F* Value	*p* Value
R. ventral visual stream	334	-31	73	-9	4.52	< .001	0.015	.90	2.79	.10	1.02	.31
L. ventral visual stream	206	40	52	-12	4.00	< .001	0.02	.89	1.33	.25	.004	.95
L. inferior frontal gyrus	188	40	-1	27	5.31	< .001	0.083	.77	.01	.91	.06	.81
R. inferior frontal gyrus	54	-37	-1	30	4.35	< .001	1.115	.29	2.06	.15	.66	.42

H4a Analysis: The Amount of Recovery Will Decrease Across Days

fMRI Analysis: Using the recovery scores from H3a, we tested whether the day interval predicted the difference between scores in a within-subjects, repeated-measures ANOVA. The day interval did not significantly predict recovery scores in any of the ROIs (see Table 5). Analysis of H3a revealed that recovery was significant in the early periods but not significant in the later periods. Thus, H4a was not supported by the fMRI data.

Eye-Tracking Analysis: We again used the difference in the fixation count from the last viewing of the warning on the previous day as our dependent variable. We used a mixed-effects model that tested whether the day interval predicted this difference. Subject ID and warning ID were included as random factors. The day interval was treated as a fixed factor, and visual complexity was included as a covariate. The eye-tracking analysis supported H4a. The day interval was significantly negative [$\chi^2(1, N = 2,400) = 2.64$, $p < .05$, $\beta = -0.093$], indicating that the recovery decreased across days. Visual complexity was insignificant: $\chi^2(1, N = 2,400) = 1.66$, $p > .05$, $\beta = 0.092$. The R^2 of the model was .041.

H4b Analysis: The Amount of Recovery Will Decrease Less for Polymorphic Warnings than for Static Warnings Across Days

fMRI Analysis: We conducted the same analysis as in H4a but with stimulus type as a factor in the model. None of the four ROIs displayed a significant day-by-stimulus-type inter-

action (see Table 5). The fMRI data, therefore, did not support H4b.

Eye-Tracking Analysis: We specified the same model as in H4a, with the addition of an interaction term between day intervals and whether the warning was polymorphic. The eye-tracking analysis did not support H4b. Both main effects for day number [$\chi^2(1, N = 2,400) = 0.58$, $p > .05$, $\beta = -0.062$] and polymorphism [$\chi^2(1, N = 2,400) = 0.00$, $p > .05$, $\beta = -0.015$] were nonsignificant; the interaction was also nonsignificant: $\chi^2(1, N = 2,400) = 0.30$, $p > .05$, $\beta = -0.063$. Likewise, visual complexity was nonsignificant: $\chi^2(1, N = 2,400) = 0.94$, $p > .05$, $\beta = 0.071$.

Experiment 2: Behavioral ▮▮▮▮

Although Experiment 1 provided valuable neural insights into the process of habituation, it did not measure actual warning adherence. It also made necessary sacrifices in ecological validity that may limit the generalizability of the findings to real life. The objectives of Experiment 2 were to measure actual warning adherence in an ecologically valid field setting. Experiment 2 improved on Experiment 1 in the following ways.

First, in Experiment 1 users were exposed to 260 warnings per day. In contrast, in Experiment 2 participants only saw a small number—an average of 4.74 (*SD* = 1.29) warnings per day. Moreover, we used mobile devices as our context, on which users typically saw multiple notifications and warnings per day. An analysis of 40,191 Android users showed that users on average encounter nearly 26 notifications per day on

their mobile phones (these notifications include app permission warnings, app notifications, email notifications, system notifications, etc.; Shirazi et al. 2014).

Second, in Experiment 1 participants viewed warnings (1) presented on a laboratory computer, (2) while lying in an MRI machine, (3) in a laboratory, and (4) at a set time each day. By comparison, in Experiment 2 participants viewed warnings (1) presented on their personal mobile devices, (2) in their natural environment, (3) at a place of their choosing, and (4) at a time of their choosing, which varied each day.

Third, Experiment 1 examined habituation of attention in terms of neural correlates and eye-tracking fixations but did not measure warning adherence itself. It was, therefore, unknown whether habituation to security warnings did in fact lead to diminished warning adherence. Likewise, it was unclear whether polymorphic warnings were actually effective in improving warning adherence over time. In contrast, Experiment 2 used warning adherence as the dependent variable. This allowed us to observe how repeated exposure to warnings influences warning adherence over time, as well as whether polymorphic warnings are effective in improving warning adherence.

Fourth, Experiment 1 examined habituation in the brain for a workweek of five days. While this allowed us to observe recovery effects over time, it may have been too short to observe whether people accepted polymorphic warnings from a usability standpoint or instead learned to ignore them—or, worse yet, disliked the forced novelty of the polymorphic warning design. By comparison, Experiment 2 observed participants' behavior over 3 workweeks or 15 days (three times as long as Experiment 1). We were, therefore, able to assess the effectiveness of polymorphic warnings over a longer period of time.

In Experiment 2, the participants often saw only one risky warning per day; thus, we could not test H1a and H1b, both of which focus on habituation within computing sessions. However, we were able to test the remaining hypotheses that address the influence of habituation, recovery, and polymorphic warnings across days.

Experiment 2: Context

In Experiment 2, we monitored participants' behavior as they accepted or rejected mobile app permission warnings—that is, warnings shown before an app is granted access to information or resources. These warnings can be shown when the app is downloaded or attempts to access a resource (i.e., just-in-time warnings). App stores may also display permission warnings before an app is downloaded (e.g., the Google Play store displays the permission warnings before downloading if just-in-time warnings are not used).

Permission warnings are frequent. By May 2016, 65 billion Android apps had been downloaded by smartphone users (Statista 2017), most of which display a permission warning during installation or use. Additionally, the average Android user has 95 apps installed on his or her mobile device (Sawers 2014). Furthermore, users often experience multiple permission warnings in a short period of time. For example, when evaluating apps, people may download several apps (and, therefore, see several associated warnings) in a short period of time. When using apps with just-in-time warnings, users will typically see a series of separate permission requests when first using them. Mobile apps, therefore, represent a realistic scenario where people frequently see warnings and are thus an appropriate context for studying longitudinal habituation to security messages.

Experiment 2 involved a third-party Android app store. Third-party app stores are common on the Android platform (e.g., Amazon Underground, GetJar, Mobogenie, SlideME, AppBrain, Aptoide Cloud Store, BAM, Top Apps, AppGratis, MyApp, MIUI app store, Baidu app store, F-Droid, etc.). Some of these app stores compete with the Google Play app store by offering app specials (e.g., free or reduced-price apps), serving markets that have restricted access to Google Play (e.g., GetJar in China), or both. Others complement the Google Play store by providing customized experiences (e.g., app of the day, in-depth app reviews, categorized apps, recommended apps) with apps that link directly to the Google Play store (e.g., AppGratis). Some of these app stores are standalone apps that can be downloaded (e.g., Amazon Underground), while others must be accessed via a web browser on the Android phone (e.g., Mobogenie). For Experiment 2, we created a browser-based, third-party app store that allowed us to monitor participants' responses to permission warnings over time.

Experiment 2: Method

Participants

Participants were students recruited from a variety of majors at a university in the United States. They received course credit for their participation in the experiment. To encourage continued participation, participants were also given $10 for completing the first week, $10 for completing the second, and $20 for completing the third, for a possible total of $40. Of an initial group of 134 subjects, 32 failed to participate beyond the first 7 days. Thus, we had 102 valid responses. These subjects were 61 percent male and had an average age of 22.1 years ($SD = 2$ years).
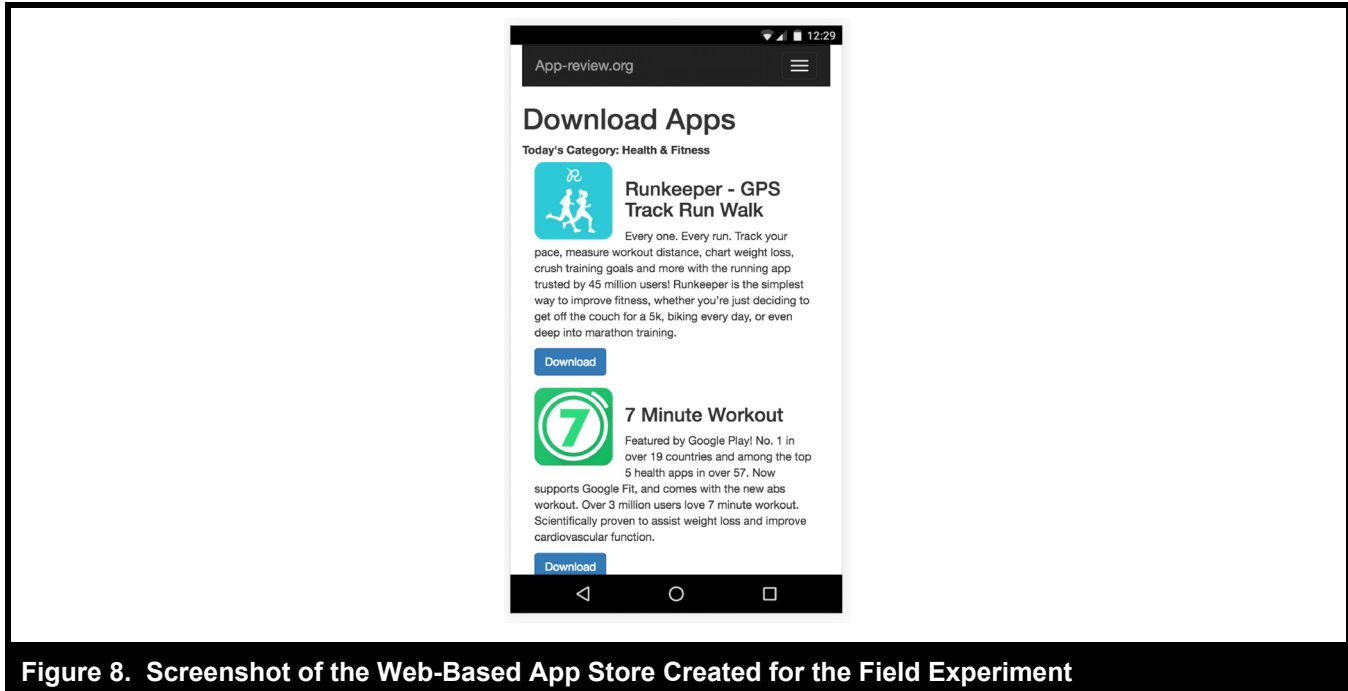
Figure 8. Screenshot of the Web-Based App Store Created for the Field Experiment

Experiment Design

Participants were asked to rank apps on an app store created specifically for this study (see Figure 8). This app store operated as a legitimate app store, and following an IRB-approved deception protocol to improve realism, participants were instructed that the app store was unaffiliated with the research team. Participants were told that the purpose of the experiment was to study how people rank apps in various categories.

The app store presented 10 apps from a different category each day (e.g., utilities, education, entertainment, travel, finance, etc.). Participants were instructed to download, install, and evaluate three of the 10 apps within the daily category on their personal Android device and rank each app from 1 (best) to 3 (worst). Participants then completed an apparently unaffiliated survey sent via email from the research team each day that allowed participants to report and rank the apps they downloaded on that day. These steps were repeated each day for three weeks (excluding weekends).

When participants clicked to download an app, they were shown a permission warning that listed the app's requested permissions, as per the Google Play store (for apps that do not implement just-in-time permissions). The permissions displayed were randomly drawn from two categories: safe and risky (see Table 6). Safe permissions were selected from the Android Developer Guide (Android 2016) and were chosen

because we determined that they would be thought of by participants as low risk across app categories. We also created four risky permissions to (1) heighten respondents' perception of risk in ignoring the permission warning and installing the app and (2) ensure that the requested permission was inappropriate regardless of the category of app on a given day.

As a second deception to increase realism, although the research team did in fact control the apps and validated their security, a set of instructions (presented in Figure 9) was issued.

Before starting the experiment, participants were required to pass a short quiz verifying that they knew which permissions were considered risky by the researchers. This allowed us to have an objective measure of security behavior: whether people knowingly installed apps with these disallowed permissions. After completing the quiz, we provided participants with the mobile URL for the app store and the separate URL for the daily survey. In addition, we sent two daily reminders to participants, once in the morning and once in the evening (if they had not already completed the task that day) to remind them to download and review three apps from the app store.

We only allowed participants to submit one category evaluation each day (we told them that this was a feature of the app store). This ensured that participants viewed a realistic number of permission warnings each day. Further, participants had to evaluate a new category each day so that they did

Table 6. Safe and Risky Permissions Displayed in the App Store Permission Warnings	
Safe Permissions	
Send notifications	Set an alarm
Pair with Bluetooth devices	Alter the phone's set time zone
Change the size of the status bar	Change the phone's displayed wallpaper
Install shortcut icons	Uninstall shortcut icons
Connect to the Internet	Use vibration for notifications or interactions
Change phone volume and audio settings	Temporarily prevent phone from sleeping (for viewing videos)
Ask permission to download additional features	
Risky Permissions	
Charge purchases to your credit card	Record microphone audio any time
Delete your photos	Sell your web-browsing data

Be aware that the research team is not affiliated with App-Review.org in any way, so we cannot verify that the apps are all safe. Before you download an app, be sure to check the permissions that the app requires. This app store displays the permissions before directing you to the Google Play store.

Make sure that the permissions required by the app do not contain any of the following:
- Charge purchases to your credit card
- Delete your photos
- Record microphone audio any time
- Sell your web-browsing data

If the app has any of these permissions, DO NOT download it. These apps are potentially dangerous and can harm your privacy and/or phone. Not only is your own device at risk if you install these apps, but also if you positively review these apps, it will put future users at risk. Therefore, if you review too many apps with dangerous permissions, you may not receive the course credit for this experiment.

Figure 9. Participant Instructions for the Field Experiment
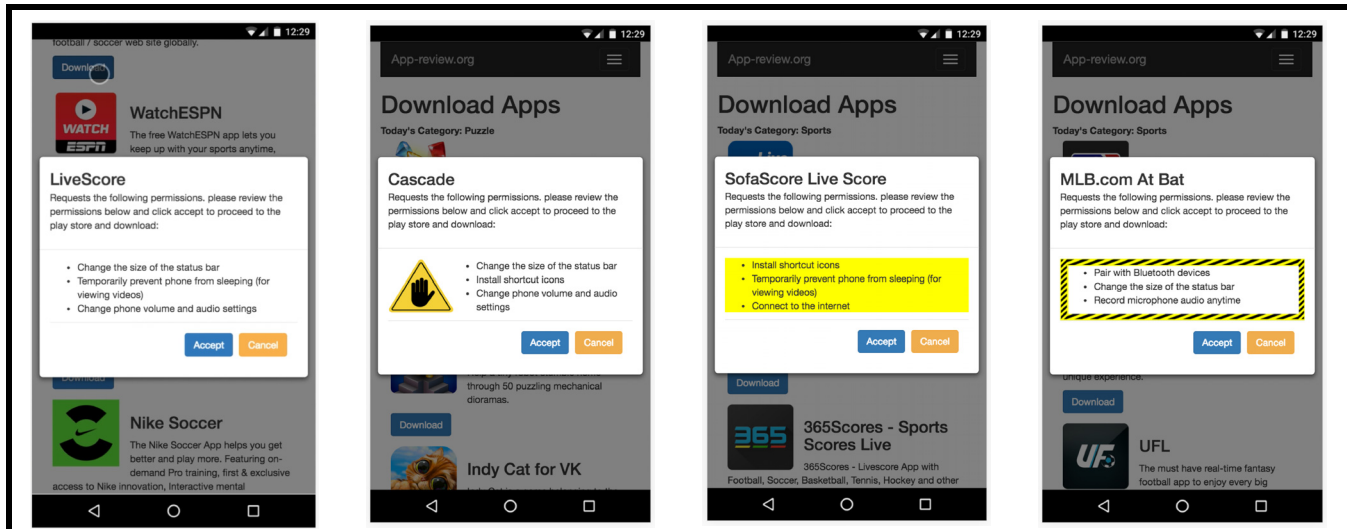
The leftmost warning shows the appearance of the warning in the static condition. The other three warnings are a sample of the 15 variations used in the polymorphic condition.

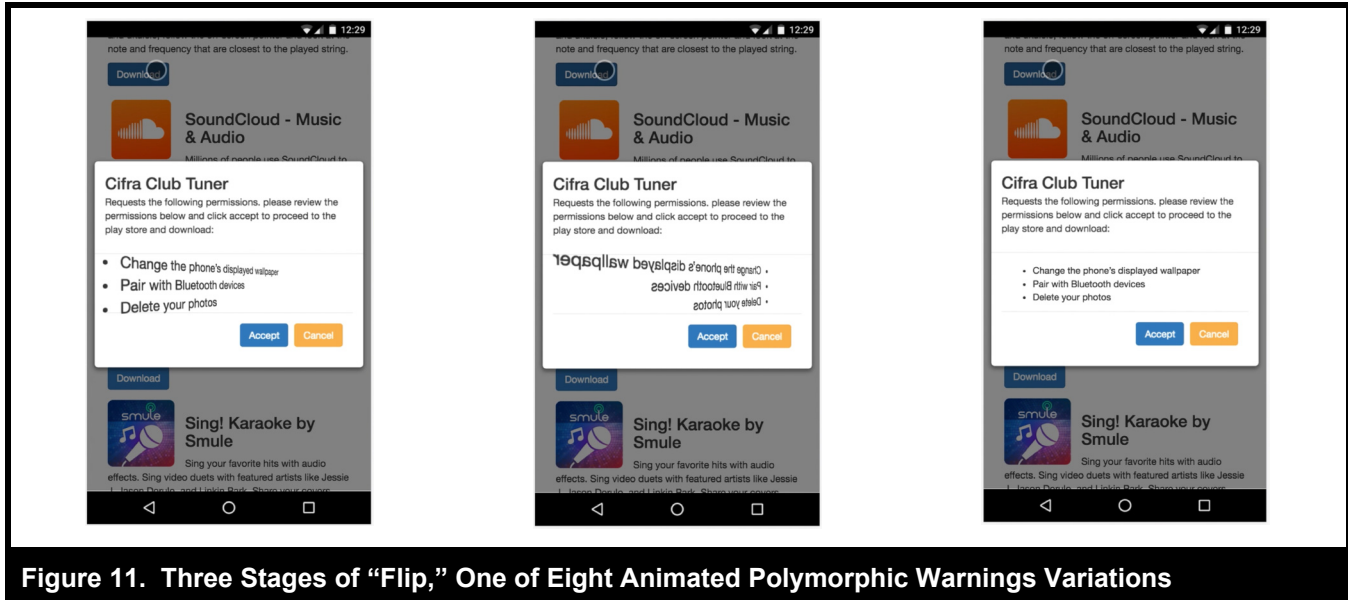**Figure 10. Sample Static and Polymorphic Permission Warnings**

Figure 11. Three Stages of "Flip," One of Eight Animated Polymorphic Warnings Variations

not evaluate the same app twice. Importantly, apps in each category were not well-known apps, reducing the likelihood that participants trusted a particular brand.

Each category had 10 apps to choose from. When clicking on the "Download" button on an app, the app store displayed the permission warning for the app (see Figure 10, leftmost screenshot). If the user accepted the permission warning, the app store completed the app installation through Google Play—a pattern shared by many of the "customized experience" app stores (e.g., Mobile App Store, Cloud Store, BAM, Top Apps, AppGratis). Participants did not see a permission warning again through Google Play when the app was downloaded.

Dependent Variable

Our dependent variable was warning adherence, which we operationalized as a binary variable: whether or not participants rejected apps with risky permissions. We randomized the permissions for each warning to ensure that participants would encounter at least one risky permission among the first three apps they selected. Each app selected beyond the first three had a 50 percent chance of displaying a risky permission. The experimental app store recorded whether or not participants ignored permission warnings containing a risky permission.

Experimental Conditions

We implemented a between-subject study design in which participants were randomly assigned to either the static or polymorphic warning condition when they first created an account on the app store. Users were required to log in to the app store, which ensured they had the same condition across all three weeks. The static warning condition always had the same look and feel for the duration of the experiment, although the requested permissions changed for each app. In contrast, the polymorphic warning condition randomly changed the appearance of the permission warning each time it was shown.

We created 16 variations of the polymorphic warning; one-half of these involved animations and one-half did not. Further, for each participant, we randomly iterated through four polymorphic warning variations every four days. This was done to maintain the novelty of the polymorphic treatment. We deliberately set the interval for changing the polymorphic versions at the fourth day of each set so we would be able to detect if the level of habituation changed due to the warning treatments rather than being due to the weekend and time away from the task. Example static and polymorphic warnings are shown in Figures 10 and 11. See Appendix D for all warning variations.

Daily Survey

After downloading and installing three apps, participants completed a survey from the research team. The survey asked participants to list and rank the three apps they had downloaded as 1 (best) to 3 (worst) for the daily category. We deliberately branded the survey as coming from the researchers, and not the app store, so it would appear that the two were not connected. This helped promote the story that

the app store was not affiliated with the researchers and, therefore, could contain risky apps. To ensure that participants actually downloaded the apps from the app store, we enabled a "sharing" feature in the store through which participants shared with us which apps they had downloaded. Although we actually captured all behavioral data in the app store regardless of this functionality, the feature again helped promote the story that the app store was not associated with the research team.

Debriefing Survey

At the end of the three-week experiment, participants received a debriefing survey. In the survey, participants were asked, "How concerned were you about each of the following permissions?" and were then presented with the permissions listed in Table 6. The response scale was 1 "Not at all concerned" to 7 "Extremely concerned." The average for risky permissions was 6.03 ($SD = 1.40$), while the average for safe permissions was 1.97 ($SD = 1.34$), a significant difference ($t = -38.9$, $df = 653.9$, $p < .001$). This indicated that participants did see a difference in concern for the two categories of permissions. We also asked a manipulation check question to ensure that participants in the experimental condition noticed the polymorphic treatment (Straub et al. 2004). All participants in the polymorphic condition responded affirmatively.

Experiment 2: Analysis ▐▬▬▬▬▬

We limited our data to participants who completed at least one-half of the days (seven days or more) of the experiment. This resulted in 102 participants—55 in the static condition and 47 in the polymorphic condition—who, all together, viewed 7,248 warnings over three weeks or 15 weekdays. This amounted to an average of 4.74 ($SD = 1.29$) permission warnings viewed per weekday, per participant. Of these, 2,695 (about one-third) were apps with risky permissions. Thus, the N for our analysis was 2,695.

To analyze our data, we specified a logistic linear mixed-effects model because it is robust to uneven observations (Cnaan et al. 1997). In other words, the analysis was robust even if participants saw a different number of warnings each day. Linear mixed-effects modeling also allows for the inclusion of fixed effects (observations that are treated as nonrandom or nonindependent) and random effects (observations that are treated as random or independent). Thus, we accounted for the within-subject nature of our experiment by including the participant identifier as a random effect. Finally, the logistic linear mixed-effects model was designed to handle binary dependent variables, such as ours (McCulloch and Neuhaus 2001).

To test H2a, we included the warning number (how many warnings the participant had seen up to that point) as a fixed effect to measure the stimulus repetition. We also included the treatment as a binary fixed effect (1 = polymorphic warnings, 0 = static warnings). We then included an interaction effect between warning number and treatment to test H2b. To test H3a, we included a binary variable of whether the participant saw the risky warning after a recovery period of a day or more (1 = recovery period, 0 = no recovery period). We then added an interaction between the recovery period and treatment to test H3b. To test H4a, we included an interaction term between the recovery period and what day of recovery it was for the participant (i.e., the recovery period after the first day, second day, etc.). Finally, to test H4b, we added the treatment to the prior interaction term (resulting in a three-way interaction) to see if the effect of the prior interaction differed by treatment. As stated previously, our dependent variable was whether the user adhered to a warning containing a risky permission and canceled their installation of the app (coded as 1), or disregarded the warning and installed the app anyway (coded as 0).

The results are shown in Table 7. The warning number negatively predicted whether the user rejected the app with the risky permission ($\beta = -0.027$, $p < .001$). Over the course of the three-week experiment, average adherence substantially dropped from 87 percent to 64 percent, a significant difference of 23 percent ($x^2 = 14.514$, $df = 1$, $p < .001$). Thus, H2a was supported. Likewise, the interaction between warning number and treatment was significant ($\beta = 0.015$, $p < .01$). Participants' accuracy in rejecting risky permission warnings decreased more slowly when viewing polymorphic warnings compared to static warnings, supporting H2b. Note that the prior two estimated coefficients (β) are relatively small (compared to the coefficient of the recovery period discussed below) because they represent the amount by which the log odds of adherence would change for each incremental warning. Thus, the overall effect is additive for each additional warning a user sees, which can add up quickly. Finally, the recovery period significantly influenced warning adherence, supporting H3a ($\beta = 0.787$, $p < .05$). H3b, H4a, and H4b were not statistically significant, which is consistent with the fMRI analysis. To examine the effect size of these predictors, we examined both R^2 and changes in the percentage of adherence. First, the conditional R^2 (R^2 associated with the random effects) of the model was .352, and the marginal R^2 was .098 (R^2 associated with the fixed effects).

Second, to explore the extent of the interaction between warning number and treatment, we graphed the trends. Figure 12 displays how each treatment group's accuracy rate (percent correct in rejecting risky apps) changed over the three-week (15-day) experiment as well as trend lines fitted to the data. Interestingly, after the three weeks, the accuracy rate of parti-

Table 7. Logistic Mixed-Effects Model Results Predicting Whether Participants Rejected Apps with Risky Permissions

	Estimate	Std. Error	*z*-value	*p*-value	*Hypothesis*
Intercept	2.082	0.216	9.642	< .001	—
Warning Number	-0.027	0.003	-8.714	< .001	H2a
Polymorphic Treatment	0.238	0.334	0.712	.238	—
Recovery Period	0.787	0.414	1.900	< .05	H3a
Warning Number × Polymorphic Treatment	0.015	0.005	2.947	< .01	H2b
Recovery Period × Polymorphic Treatment	0.055	0.630	0.088	.465	H3b
Recovery Period Number × Recovery Period	-0.008	0.008	0.998	.159	H4a
Recovery Period Number × Recovery Period × Polymorphic Treatment	-0.013	0.013	-1.058	.145	H4b
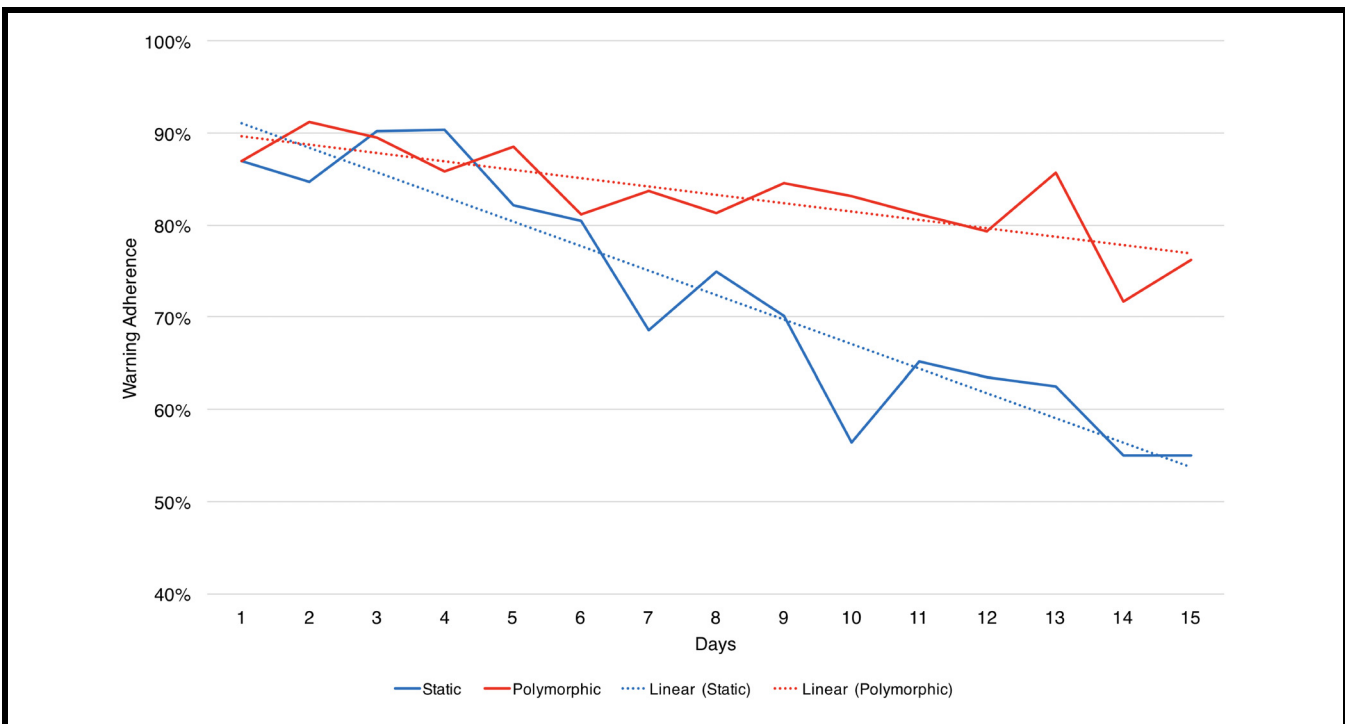
Figure 12. Percentage of Warning Adherence in Rejecting Risky Warnings Across 15 Weekdays for Each Treatment Group

cipants in the polymorphic condition was 76 percent, whereas the accuracy of participants in the static condition was 55 percent. This difference of 21 percent was significant ($x^2 = 7.172$, $df = 1$, $p < .01$). Overall, accuracy in the polymorphic condition dropped from 87 percent at the start of the three weeks to 76 percent at the end. In contrast, accuracy in the static condition dropped from 87 percent to 55 percent.

Discussion

Table 8 summarizes our results. We note that although the dependent variables, methods, and experimental designs of Experiments 1 and 2 were quite different, they were consistent overall in their tests of the hypotheses. We discuss our contributions below.

Table 8. Summary of Results			
Hypothesis	**Experiment 1:**		**Experiment 2:**
	Eye Tracking	**fMRI**	**Behavior**
H1a: Users habituate to warnings within an experimental session	**Supported**	**Supported**	Not tested
H1b: Users habituate less to polymorphic warnings than to static warnings in an experimental session	**Supported**	**Supported**	Not tested
H2a: Users habituate to warnings across days	**Supported**	**Supported**	**Supported**
H2b: Users habituate less to polymorphic warnings than to static warnings across days	**Supported**	**Supported**	**Supported**
H3a: If warnings are withheld after habituation occurs, the response recovers at least partially the next day	**Supported**	**Supported**	**Supported**
H3b: If warnings are withheld after habituation occurs, response recovery is stronger for polymorphic warnings than for static warnings the next day	Not supported	Not supported	Not supported
H4a: The amount of recovery will decrease across days	**Supported**	Not supported	Not supported
H4b: The amount of recovery will decrease less for polymorphic warnings than for static warnings across days	Not supported	Not supported	Not supported

This paper makes three principal contributions to the field of information systems: (1) measuring habituation of attention to security warnings longitudinally, including response decrement and recovery; (2) examining the effect of habituation on warning adherence in the field; and (3) demonstrating that polymorphic warnings are effective in reducing habituation over time in terms of both diminished attention and warning adherence. We discuss each of these points below.

First, measuring habituation directly in terms of neural activity and eye tracking is valuable because brain activity "can be used as a mediator between the IT artifact and IT behavior" (Riedl and Léger 2016, p. 20), which can allow researchers to understand not only how people behave but why they behave that way. Neurobiology plays a key role in security behavior because habituation is an unconscious mental process that is difficult to measure without neurophysiological methods (Dimoka et al. 2011). These contributions are discussed further in Appendix E.

We extend prior research by showing how habituation develops over time. The longitudinal nature of our experiments allowed us to capture these hidden mental processes over the course of several days. We were able to capture not only the response decrement in the habituation process but also the daily recovery or increase in response strength. Although our findings do not support a greater recovery (increased response to stimuli after a rest period) associated with the polymorphic warnings, they do demonstrate that withholding warnings for a time can help increase sensitivity and response to warnings.

Further, in our attempts to develop IT artifacts that are more resistant to habituation, neurophysiological tools can be used to instruct developers on the neural mechanisms influencing the response to the IT artifact (Riedl and Léger 2016). Specifically, these tools can capture the effects of the IT artifacts on users more objectively than traditional measures (vom Brocke et al. 2013). In addition, by using both fMRI and eye tracking, we demonstrated that the latter is a valid measure of habituation that can be used in more ecologically valid settings.

Second, whereas previous studies have examined habituation in laboratory settings (see Table 1), we demonstrated in Experiment 2 how habituation affects actual warning adherence in the field. This is an important contribution in terms of ecological validity because habituation is inseparably related to the frequency of stimuli received. Due to memory effects and other factors, people may exhibit different patterns of habituation to multiple warnings in a one-hour laboratory session than they would to the same number of warnings over the course of a day. This is necessarily the case even if the number of warnings presented in a one-hour laboratory session is proportionally accurate to the number of warnings received over an entire day (see Figure 13) because the time period over which the warnings are displayed is compressed. Thus, it is not possible to conduct a laboratory experiment of habituation to warnings in a perfectly ecologically valid way.[5]

---

[5]We thank the associate editor and an anonymous reviewer for clarifying these points, and the associate editor for Figure 13.
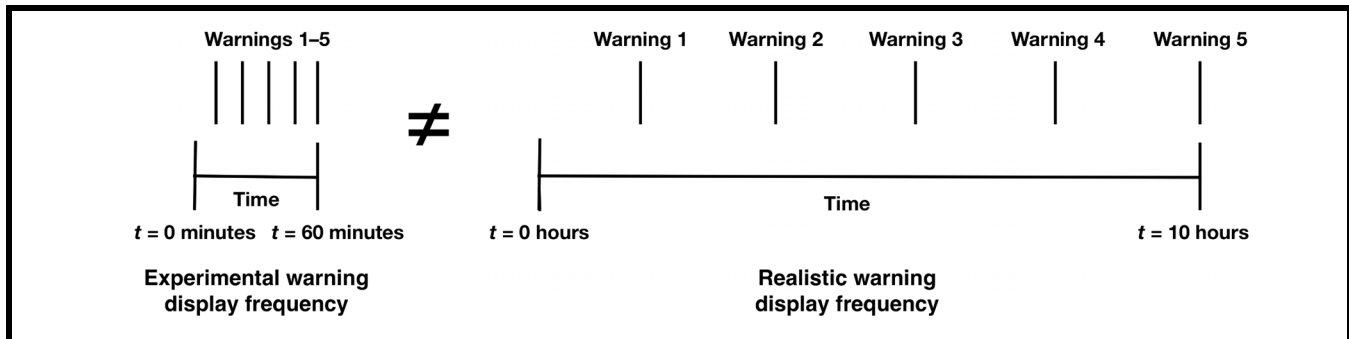
Figure 13. Habituation to Warnings in Compressed, Yet Proportionally Accurate, Time Frame May Not Necessarily Equal Habituation in a Natural, True-to-Life Time Frame

This problem is exacerbated when an unrealistically high number of warnings is presented in laboratory sessions (such as in Anderson, Jenkins et al. 2016; Anderson, Vance et al. 2016b; Bravo-Lillo et al. 2014; Bravo-Lillo et al. 2013; Brustoloni and Villamarín-Salomón 2007; and Experiment 1). As a consequence, it is unclear how previous findings of laboratory experiments correspond to real-life habituation and warnings.

Experiment 2 contributes by providing an ecologically valid field experiment design that for the first time showed how a realistic repetition of warnings in the field (an average of 4.74 warnings per day over three weeks) results in a decrease of warning adherence. The results of Experiment 2 are consistent with habituation theory and the neuroimaging and eye-tracking results of Experiment 1, providing strong evidence of the negative influence of habituation on warning adherence. Interestingly, the results of Experiment 2 show a pattern of habituation that is similar to those of laboratory experiments that measure habituation in a condensed and artificial way. Therefore, Experiment 2 also contributes by validating past results and suggesting that laboratory experiments can serve as useful proxies for real-world habituation to warnings.

An additional contribution of our field study design is that it allowed us to show how habituation of attention to security warnings maps to actual behavior. A weakness of Experiment 1 and past studies of habituation (Anderson, Jenkins et al. 2016; Anderson, Vance et al. 2016b) is that they assumed that habituating to security warnings would result in lower warning adherence (i.e., behavior). However, this may not necessarily be the case—it is possible that users pay diminished attention to a warning and still exhibit high warning adherence because they already recognize the warning and have consciously decided to respond in a secure way. Experiment 2 contributes by showing that adherence decreases significantly with each exposure to a warning. Over the course of the three-week experiment, average adherence

substantially dropped from 87 percent to 64 percent, a significant difference of 23 percent.

Third, although previous research has proposed polymorphic warnings (Anderson, Jenkins et al. 2016; Anderson, Vance et al. 2016b), the design was only evaluated cross-sectionally, despite the fact that habituation is a phenomenon that develops over time. As a result, it was unclear whether polymorphic warnings would maintain their advantage over time or lose their saliency to the point that users would react to static and polymorphic warnings in the same way. Moreover, their designs only examined the effect of polymorphic warnings on attention, rather than on behavior itself. Therefore, it was unknown whether higher levels of attention paid to polymorphic warnings would translate to improved behavior or if users would become indifferent—or worse, respond negatively—to them over time.

This study contributes by addressing these questions. In Experiment 1, we showed that polymorphic warnings maintained substantially higher levels of attention over the five-day experiment—both in terms of neural and eye-gaze activity—compared to static warnings. In addition, in Experiment 2 we demonstrated that polymorphic warnings sustain substantially higher levels of warning adherence compared to static warnings. Although users did habituate to the polymorphic warnings, the rate of habituation was significantly slower than that of static warnings, as evidenced by the significant interaction between the warning exposure number and the polymorphic treatment. Further, after three weeks, the warning adherence rate of participants in the polymorphic condition was 76 percent, compared to 55 percent for those in the static condition, a significant difference of 21 percent. Overall, accuracy in the polymorphic condition dropped from 87 percent at the start of the three weeks to 76 percent at the end. In contrast, accuracy in the static condition dropped from 87 percent to 55 percent.

Moreover, although participants were conscious of the polymorphic warnings (100 percent of participants in the experimental treatment correctly answered the manipulation check question in our post-survey), their warning adherence was significantly higher compared to those in the static group. This indicates that the polymorphic design had a sustained advantage both in higher attention and adherence over time, providing a clear contribution to the IS security literature.

Finally, Experiments 1 and 2 complement each other and together provide the most complete test yet of habituation theory as applied to security warnings. Experiment 1 directly measured habituation in the brain over time with repeated exposure to warnings. However, it did not measure actual behavior. In contrast, Experiment 2 measured how warning adherence decreases over time with repeated exposure to warnings, but it did not measure habituation directly. The results of both experiments evaluate habituation theory from different angles, showing its neural and behavioral effects (Figure 2). Combined, Experiments 1 and 2 provide strong evidence that the polymorphic warning, as an IT artifact, can substantially reduce habituation and improve warning adherence.

Limitations and Future Research

Our research is subject to several limitations. First, fMRI methods require stimuli to be repeated for the sake of reliability of measurement. However, as noted above, this results in poor ecological validity in the context of habituation to warnings. We discuss this limitation in further detail in Appendix F. However, Experiment 2 partially compensated for this limitation by showing how warning adherence diminishes with repeated exposure. The pattern of habituation we observed in Experiment 2 closely resembles the fMRI and eye-tracking results of Experiment 1, suggesting that the experimental design of Experiment 1 is a good proxy for real-world habituation to warnings, despite its artificiality.

Second, Experiments 1 and 2 used different experimental designs and are, therefore, not directly comparable. This is a consequence of the methodological limitations of fMRI, as well as the different objectives for each experiment. Additionally, both experiments used different sets of stimuli: Experiment 1 involved images of a large variety of warnings with four polymorphic variations, while Experiment 2 used a single warning with 16 polymorphic variations. For these reasons, we argue that Experiment 2 complements rather than replicates Experiment 1. Nevertheless, both experiments are consistent in their (1) application of habituation theory, (2) use of polymorphic design principles, and (3) results.

Third, although the goal of Experiment 2 was to provide a realistic field experiment, it, too, necessarily involved some artificiality. For example, participants completed an assigned task (i.e., download and evaluate three mobile apps a day from a designated app store), rather than behaving freely as in a true field study. Similarly, we presented participants with four artificial risky permissions that are not actually used by the Android OS (see Table 6). These design decisions are typical of the field experiment method, which imposes experimental conditions for the sake of precision and control but does so in a natural environment for enhanced realism (Boudreau et al. 2001).

Fourth, participants did not experience any negative consequences in Experiment 2 for installing an app that requested a risky permission. Therefore, an alternative explanation for the diminishing of security warning adherence over the three weeks is the "crying wolf" effect, in which people ignore a warning for which the associated danger never materializes (Sunshine et al. 2009), rather than due to habituation, as we theorize. However, this was true of both the static and polymorphic conditions; and yet, warning adherence in the polymorphic condition remained high throughout the experiment. In addition, the results are consistent with those of Experiment 1, which clearly showed the effects of habituation when warnings are repeated. These two findings provide strong evidence that habituation was a key factor in Experiment 2, although other factors may also have been at play.

Finally, security warnings may fail for a variety of reasons, such as lack of comprehension on the part of recipients (Felt et al. 2015), distraction or interfering noise in the environment (Wogalter 2006), or conscious decisions to ignore them (Herley 2009). Future research should investigate these and other factors (Anderson, Vance et al. 2016a). Similarly, besides response decrement and recovery, other facets of habituation may influence the users' responses to security warnings, such as dishabituation (the recovery of a response by encountering another strong stimulus) and generalization (the carryover of habituation from one stimulus to another novel stimulus; Rankin et al. 2009).

Conclusion

This paper contributes to the IS literature by providing the most complete examination to date of the problem of habituation to security warnings. We added to past research by examining habituation longitudinally, both via fMRI and eye tracking, as well as through a field experiment involving security warning adherence. Our results illustrate that habitu-

ation occurs over time at the neurobiological level. We also demonstrated that exposure to repeated warnings results in diminishing security warning adherence. Finally, we showed that polymorphic warnings are effective in reducing habituation over time, manifested as both attention at the neurobiological level and in actual security warning adherence.

Acknowledgments

References

Anderson, B. B., Jenkins, J., Vance, A., Kirwan, C. B., and Eargle, D. 2016. "Your Memory Is Working Against You: How Eye Tracking and Memory Explain Susceptibility to Phishing," *Decision Support Systems* (25:4), pp. 3-13.

Anderson, B., Vance, A., Kirwan, B., Eargle, D., and Howard, S. 2014. "Users Aren't (Necessarily) Lazy: Using NeuroIS to Explain Habituation to Security Warnings," in *Proceedings of the 35th International Conference on Information Systems Proceedings*, Auckland, New Zealand.

Anderson, B. B., Vance, A., Kirwan, C. B., Eargle, D., and Jenkins, J. L. 2016a. "How Users Perceive and Respond to Security Messages: A NeuroIS Research Agenda and Empirical Study," *European Journal of Information Systems* (25:4), pp. 364-390.

Anderson, B. B., Vance, A., Kirwan, C. B., Jenkins, J. L., and Eargle, D. 2016b. "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), pp. 713-743.

Android. 2016. "Android Developers Guide" (https://developer.android.com/guide/topics/permissions/requesting.html).

Barry, R. J. 2009. "Habituation of the Orienting Reflex and the Development of Preliminary Process Theory," *Neurobiology of Learning and Memory* (92:2), pp. 235-242.

Boudreau, M. C., Gefen, D., and Straub, D. 2001. "Validation in Information Systems Research: A State-of-the-Art Assessment," *MIS Quarterly* (25:1), pp. 1-16.

Braun, C. C., and Silver, N. C. 1995. "Interaction of Signal Word and Colour on Warning Labels: Differences in Perceived Hazard and Behavioural Compliance," *Ergonomics* (38:11), pp. 2207-2220.

Bravo-Lillo, C., Cranor, L., Komanduri, S., Schechter, S., and Sleeper, M. 2014. "Harder to Ignore? Revisiting Pop-Up

Fatigue and Approaches to Prevent It," in *Proceedings of the 10th Symposium on Usable Privacy and Security*, Menlo Park, CA: USENIX Association, pp. 105-111.

Bravo-Lillo, C., Komanduri, S., Cranor, L. F., Reeder, R. W., Sleeper, M., Downs, J., and Schechter, S. 2013. "Your Attention Please: Designing Security-Decision UIs to Make Genuine Risks Harder to Ignore," in *Proceedings of the 9th Symposium on Usable Privacy and Security*, New York: ACM Press, Article 6.

Brustoloni, J. C., and Villamarín-Salomón, R. 2007. "Improving Security Decisions with Polymorphic and Audited Dialogs," in *Proceedings of the 3rd Symposium on Usable Privacy and Security*, New York: ACM Press, pp. 76-85.

Chen, G., Adleman, N. E., Saad, Z. S., Leibenluft, E., and Cox, R. W. 2014. "Applications of Multivariate Modeling to Neuroimaging Group Analysis: A Comprehensive Alternative to Univariate General Linear Model," *NeuroImage* (99), pp. 571-588.

Cnaan, A., Laird, N. M., and Slasor, P. 1997. "Tutorial in Biostatistics: Using the General Linear Mixed Model to Analyse Unbalanced Repeated Measures and Longitudinal Data," *Statistics in Medicine* (16), pp. 2349-2380.

Cooke, S. F., Komorowski, R. W., Kaplan, E. S., Gavornik, J. P., and Bear, M. F. 2015. "Visual Recognition Memory, Manifested as Long-Term Habituation, Requires Synaptic Plasticity in V1," *Nature Neuroscience* (18:2), pp. 262-271.

Cox, R. W. 1996. "AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages," *Computers and Biomedical Research* (29:3), pp. 162-173.

Dimoka, A. 2010. "What Does the Brain Tell Us About Trust and Distrust? Evidence from a Functional Neuroimaging Study," *MIS Quarterly* (34:2), pp. 373-396.

Dimoka, A. 2012. "How to Conduct a Functional Magnetic Resonance (fMRI) Study in Social Science Research," *MIS Quarterly* (36:3), pp. 811-840.

Dimoka, A., Banker, R. D., Benbasat, I., Davis, F. D., Dennis, A. R., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P. H., Pavlou, P. A., Müller-Putz, G., Riedl, R., vom Brocke, J., and Weber, B. 2012. "On the Use of Neurophysiological Tools in IS Research: Developing a Research Agenda for NeuroIS," *MIS Quarterly* (36:3), pp. 679-702.

Dimoka, A., Pavlou, P. A., and Davis, F. D. 2011. "Research Commentary—NeuroIS: The Potential of Cognitive Neuroscience for Information Systems Research," *Information Systems Research* (22:4), pp. 687-702.

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

Egelman, S., and Schechter, S. 2013. "The Importance of Being Earnest [in Security Warnings]," in *Financial Cryptography and Data Security*, A. R. Sadeghi (ed.), Berlin: Springer, pp. 52-59.

Felt, A. P., Ainslie, A., Reeder, R. W., Consolvo, S., Thyagaraja, S., Bettes, A., Harris, H., and Grimes, J. 2015. "Improving SSL Warnings: Comprehension and Adherence," in *Proceedings of the 33rd Annual Conference on Human Factors in Computing Systems*, New York: ACM Press, pp. 2893-2902.

Furnell, S. 2010. "Usability Versus Complexity—Striking the Balance in End-User Security," *Network Security* (2010:12), pp. 13-17.

Grill-Spector, K., Henson, R., and Martin, A. 2006. "Repetition and the Brain: Neural Models of Stimulus-Specific Effects," *Trends in Cognitive Sciences* (10:1), pp. 14-23.

Groves, P. M., and Thompson, R. F. 1970. "Habituation: A Dual-Process Theory," *Psychological Review* (77), pp. 419-450.

Heisz, J. J., and Shore, D. I. 2008. "More Efficient Scanning for Familiar Faces," *Journal of Vision* (8:1), pp. 1-10.

Herley, C. 2009. "So Long, and No Thanks for the Externalities: The Rational Rejection of Security Advice by Users," in *Proceedings of the 2009 Workshop on New Security Paradigms*, Oxford, UK, pp. 133-144.

Jenkins, J. L., Anderson, B. B., Vance, A., Kirwan, C. B., and Eargle, D. 2016. "More Harm Than Good? How Messages That Interrupt Can Make Us Vulnerable," *Information Systems Research* (27:4), pp. 880-896.

Kalsher, M. J., Wogalter, M. S., and Racicot, B. M. 1996. "Pharmaceutical Container Labels: Enhancing Preference Perceptions with Alternative Designs and Pictorials," *International Journal of Industrial Ergonomics* (18:1), pp. 83-90.

Krol, K., Moroz, M., and Sasse, M. A. 2012. "Don't Work. Can't Work? Why It's Time to Rethink Security Warnings," in *Proceedings of the 7th International Conference on Risk and Security of Internet and Systems*, Cork, Ireland, pp. 1-8.

Leung, C. M. 2009. "Depress Phishing by CAPTCHA with OTP," in *Proceedings of the 3rd International Conference on Anti-Counterfeiting, Security, and Identification in Communication*, Hong Kong, pp. 187-192.

McCulloch, C. E., and Neuhaus, J. M. 2001. *Generalized Linear Mixed Models*, Hoboken, NJ: John Wiley & Sons, Ltd.

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., and Thompson, R. F. 2009. "Habituation Revisited: An Updated and Revised Description of the Behavioral Characteristics of Habituation," *Neurobiology of Learning and Memory* (92:2), pp. 135-138.

Riedl, R., Banker, R. D., Benbasat, I., Davis, F. D., Dennis, A. R., Dimoka, A., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P., Müller-Putz, G., Pavlou, P. A., Straub, D. W., vom Brocke, J., and Weber, B. 2010. "On the Foundations of NeuroIS: Reflections on the Gmunden Retreat 2009," *Communications of the Association for Information Systems* (27), Article 15.

Riedl, R., Davis, F. D., and Hevner, A. R. 2014. "Towards a NeuroIS Research Methodology: Intensifying the Discussion on Methods, Tools, and Measurement," *Journal of the Association for Information Systems* (15:10), pp. i-xxxv.

Riedl, R., Hubert, M., and Kenning, P. 2010. "Are There Neural Gender Differences in Online Trust? An fMRI Study on the Perceived Trustworthiness of eBay Offers," *MIS Quarterly* (34:2), pp. 397-428.

Riedl, R., and Léger, P. M. 2016. *Fundamentals of NeuroIS: Information Systems and the Brain*, Berlin: Springer.

Riedl, R., Mohr, P. N. C., Kenning, P. H., Davis, F. D., and Heekeren, H. R. 2014. "Trusting Humans and Avatars: A Brain Imaging Study Based on Evolution Theory," *Journal of Management Information Systems* (30:4), pp. 83-114.

Rosenholtz, R., Li, Y., and Nakano, L. 2007. "Measuring Visual Clutter," *Journal of Vision* (7:2), pp. 1-22.

Rudin-Brown, C. M., Greenley, M. P., Barone, A., Armstrong, J., Salway, A. F., and Norris, B. J. 2004. "The Design of Child Restraint System (CRS) Labels and Warnings Affects Overall CRS Usability," *Traffic Injury Prevention* (5:1), pp. 8-17.

Ryan, J. D., Althoff, R. R., Whitlow, S., and Cohen, N. J. 2000. "Amnesia Is a Deficit in Relational Memory," *Psychological Science* (11:6), pp. 454-461.

Sanderson, D. J., and Bannerman, D. M. 2011. "Competitive Short-Term and Long-Term Memory Processes in Spatial Habituation," *Journal of Experimental Psychology: Animal Behavior Processes* (37:2), pp. 189-199.

Sawers, P. 2014. "Android Users Have an Average of 95 Apps Installed on Their Phones, According to Yahoo Aviate Data," *The Next Web* (https://thenextweb.com/apps/2014/08/26/android-users-average-95-apps-installed-phones-according-yahoo-aviate-data/#.tnw_p8IqFXt0).

Schechter, S. E., Dhamija, R., Ozment, A., and Fischer, I. 2007. "The Emperor's New Security Indicators," in *Proceedings of the 2007 IEEE Symposium on Security and Privacy*, Berkeley, CA, pp. 51-65.

Shirazi, A. S., Henze, N., Dingler, T., Pielot, M., Weber, D., and Schmidt, A. 2014. "Large-Scale Assessment of Mobile Notifications," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, M. Jones and P. Palanque (eds.), New York: ACM Press, pp. 3055-3064.

Smith, C. N., and Squire, L. R. 2017. "When Eye Movements Express Memory for Old and New Scenes in the Absence of Awareness and Independent of Hippocampus," *Learning & Memory* (24:2), pp. 95-103.

Sojourner, R. J., and Wogalter, M. S. 1997. "The Influence of Pictorials on Evaluations of Prescription Medication Instructions," *Drug Information Journal* (31:3), pp. 963-972.

Sokolov, E. 1963. "Higher Nervous Functions: The Orienting Reflux," *Annual Review of Physiology* (25), pp. 545-580.

Sotirakopoulos, A., Hawkey, K., and Beznosov, K. 2011. "On the Challenges in Usable Security Lab Studies: Lessons Learned from Replicating a Study on SSL Warnings," in *Proceedings of the 7th Symposium on Usable Privacy and Security*, New York: ACM Press, pp. 3:1-3:18.

Staddon, J. E. R. 2005. "Interval Timing: Memory, Not a Clock," *Trends in Cognitive Sciences* (9:7), pp. 312-314.

Statista. 2017. "Cumulative Number of Apps Downloaded from the Google Play as of May 2016 (in Billions)" (https://www.statista.com/statistics/281106/number-of-android-app-downloads-from-google-play/).

Straub, D., Boudreau, M. C., and Gefen, D. 2004. "Validation Guidelines for IS Positivist Research," *Communications of the Association for Information Systems* (13:24), pp. 380-427.

Sunshine, J., Egelman, S., Almuhimedi, H., Atri, N., and Cranor, L. F. 2009. "Crying Wolf: An Empirical Study of SSL Warning Effectiveness," in *Proceedings of the 18th Conference on USENIX*

Security Symposium, Berkeley, CA: USENIX Association, pp. 399-416.

Terry, W. S. 2015. *Learning and Memory: Basic Principles, Processes, and Procedures* (4th ed.), New York: Routledge.

Thompson, R. F. 2009. "Habituation: A History," *Neurobiology of Learning and Memory* (92:2), pp. 127-134.

Vance, A., Anderson, B. B., Kirwan, C. B., and 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), pp. 679-722.

Venkatraman, V., Dimoka, A., Pavlou, P. A., Vo, K., Hampton, W., Bollinger, B., Hershfield, H. E., Ishihara, M., and Winer, R. S. 2015. "Predicting Advertising Success Beyond Traditional Measures: New Insights from Neurophysiological Methods and Market Response Modeling," *Journal of Marketing Research* (52:4), pp. 436-452.

vom Brocke, J., and Liang, T.-P. 2014. "Guidelines for Neuroscience Studies in Information Systems Research," *Journal of Management Information Systems* (30:4), pp. 211-234.

vom Brocke, J., Riedl, R., and Léger, P.-M. 2013. "Application Strategies for Neuroscience in Information Systems Design Science Research," *Journal of Computer Information Systems* (53:3), pp. 1-13.

Wagner, A. R. 1976. "Priming in STM: An Information-Processing Mechanism for Self-Generated or Retrieval-Generated Depression in Performance," in *Habituation: Perspectives from Child Development, Animal Behavior, and Neurophysiology*, T. J. Tighe and R. N. Leaton (eds.), Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 95-128.

Wagner, A. R. 1979. "Habituation and Memory," in *Mechanisms of Learning and Motivation: A Memorial Volume to Jerzy Konorski*, A. Dickinson and R. A. Boakes (eds.), Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 53-82.

Warkentin, M., Walden, E., Johnston, A. C., and Straub, D. W. 2016. "Neural Correlates of Protection Motivation for Secure IT Behaviors: An fMRI Examination," *Journal of the Association for Information Systems* (17:3), pp. 194-215.

Wixted, J. T. 2004. "The Psychology and Neuroscience of Forgetting," *Annual Review of Psychology* (55), pp. 235-269.

Wogalter, M. S. 2006. "Communication-Human Information Processing (C-HIP) Model," in *Handbook of Warnings*, M. S. Wogalter (ed.), Mahwah, NJ: Lawrence Erlbaum Associates, pp. 51-61.

About the Authors

Anthony Vance is an associate professor in the Information Systems Department in the Marriott School of Business at Brigham Young University and in the Information Technology Management Department of the Shidler College of Business at the University of Hawai'i at Mānoa. He has earned Ph.D. degrees in Information Systems from Georgia State University; the University of Paris— Dauphine, France; and the University of Oulu, Finland. His work is published in outlets such as *MIS Quarterly*, *Information Systems Research*, *Journal of Management Information Systems*, *Journal of the Association for Information Systems*, *European Journal of Information Systems*, *Journal of the American Society for Information Science and Technology*, and proceedings of the ACM Conference on Human Factors in Computing Systems. His research focuses on behavioral and neuroscience applications to information security. He currently is an associate editor at *MIS Quarterly* and serves on the editorial board of *Journal of the Association for Information Systems*. Anthony will be joining the Management Information Systems Department in the Fox School of Business at Temple University later this year.

Jeffrey L. Jenkins is an assistant professor of Information Systems at the Marriott School of Business, Brigham Young University. He graduated with a Ph.D. in Management Information Systems from the University of Arizona. His active research includes human–computer interaction and behavioral information security. In a human–computer interaction context, Jeff's research explores how to infer human states using computer input devices such as the computer mouse, keyboard, or touchscreen. His research has been published in various journals and conference proceedings, including *MIS Quarterly*, *Information Systems Research*, *Journal of Management Information Systems*, *European Journal of Information Systems*, *Computers in Human Behavior*, and proceedings of the ACM Conference on Human Factors in Computing Systems. Prior to earning his Ph.D., Jeff was a software engineer in both the public and private sectors.

Bonnie Brinton Anderson is a professor of Information Systems and is chair of the Information Systems Department in the Marriott School of Business at Brigham Young University. She received her Ph.D. from Carnegie Mellon University. Her work has been published in *MIS Quarterly*, *Information Systems Research*, *Journal of Management Information Systems*, *Journal of the Association for Information Systems*, *European Journal of Information Systems*, *Decision Support Systems*, proceedings of the ACM Conference on Human Factors in Computing Systems, and other outlets. She currently researches the intersection of decision neuroscience and behavioral information security.

Daniel K. Bjornn is a Ph.D. candidate in Psychology, emphasizing in cognitive and behavioral neuroscience. He received his B.S. in Psychology from Brigham Young University in 2013. Dan is interested in how cognitive neuroscience can be applied to the field of user experience. His research focus is in memory, looking specifically at what allows us to differentiate very similar memories as well as how memory can be used to improve user security behavior. He has published in *MIS Quarterly* and the proceedings of the ACM Conference on Human Factors in Computing Systems and has presented research at several conferences including the Annual Meeting of the Society for Neuroscience and the Interdisciplinary Symposium on Decision Neuroscience.

C. Brock Kirwan is an associate professor of Psychology and Neuroscience at Brigham Young University. He received his Ph.D. in Psychological and Brain Sciences from Johns Hopkins University in 2006. Brock has a decade of experience conducting fMRI scans with patient populations at Johns Hopkins University; the University of California, San Diego; the University of Utah; and now BYU. He has published numerous papers reporting fMRI and neuropsycho-logical results in journals such as *Science*, *Proceedings of the National Academy of Sciences*, *Neuron*, and *Journal of Neuro-science*, as well as journals in the field of information systems such as *MIS Quarterly*, *Information Systems Research*, *Journal of Man-agement Information Systems*, *Journal of the Association for Infor-mation Systems*, and *European Journal of Information Systems*.

MIS Quarterly

TUNING OUT SECURITY WARNINGS: A LONGITUDINAL EXAMINATION OF HABITUATION THROUGH fMRI, EYE TRACKING, AND FIELD EXPERIMENTS

Anthony Vance, Jeffrey L. Jenkins, Bonnie Brinton Anderson
Information Systems Department, Marriott School of Business, Brigham Young University,
Provo, UT 84602 U.S.A. {anthony@vance.name} {jeffrey_jenkins@byu.edu} {bonnie_anderson@byu.edu}

Daniel K. Bjornn
Department of Psychology, Brigham Young University,
Provo, UT 84602 U.S.A. {dbjornn@byu.edu} {kirwan@byu.edu}

C. Brock Kirwan
Department of Psychology and Neuroscience Center, Brigham Young University,
Provo, UT 84602 U.S.A. {kirwan@byu.edu}

Appendix A

Theoretical and Methodological Background of Habituation

Habituation and Habit Compared

Although the words have the same Latin root, the construct of habituation is very different from the construct of habit. *Habit* is defined as "learned sequences of acts that have become automatic responses to specific cues, and are functional in obtaining certain goals or end-states" (Verplanken and Aarts 1999, p. 104). Further, habits are "created by frequently and satisfactorily pairing the execution of an act in response to a specific cue" (Verplanken and Orbell 2003, p. 1314). These descriptions show that habit occurs at the behavioral level and is a form of associative learning, in which behaviors are associated with specific outcomes.

Habituation, in contrast, occurs at the neurobiological level (Ramaswami 2014); it is a form of nonassociative learning, in which an organism filters out stimuli that in the past have not led to relevant outcomes (Rankin et al. 2009). Further, habituation does not require subsequent behavior but occurs involuntarily, that is, without conscious awareness. Indeed, habituation and habit correspond with opposite ends of the C–HIP model: the "attention" and "behavior" stages, respectively. Although habitual behavior is also relevant to security warnings, we restrict our focus to habituation in this study.

What Is Habituation?

Habituation is widely recognized as "the simplest and most basic form of learning" (Rankin 2009, p. 125); it is believed to be ubiquitous in the animal kingdom, having been found "in every organism studied, from single-celled protozoa, to insects, fish, rats, and people" (Rankin 2009, p. 125; see also Christoffersen 1997). In contrast to associative learning, in which a response to stimulus is associated with another stimulus (e.g., Pavlovian conditioning), habituation is a form of nonassociative learning, because an organism undergoing habituation adjusts the way in which it responds to a stimulus without pairing it with another stimulus, such as a specific consequence (Çevik 2014).

Habituation is an important survival mechanism because it allows organisms to filter out stimuli in the environment that are not relevant, thus conserving energy to respond to stimuli that predict things that are good or bad for survival (Schmid et al. 2014). Not surprisingly, humans exhibit habituation to a wide variety of stimuli—visual, auditory, and others—and it is evident as early as infancy (Colombo and Mitchell 2009).

Repetition Suppression: A Neurobiological View of Habituation

A neural manifestation of habituation to visual stimuli in the brain is called *repetition suppression* (RS): the reduction of neural responses to stimuli that are repeatedly viewed (Grill-Spector et al. 2006). RS has been observed in a range of neural measurement techniques, including single-cell recording (Kaliukhovich and Vogels 2010), functional magnetic resonance imaging (fMRI) (Hawco and Lepage 2014; Summerfield et al. 2008; Vidyasagar et al. 2010), electroencephalography (EEG) (Summerfield et al. 2011), and magnetoencephalography (MEG) (Todorovic and de Lange 2012). Researchers have observed activation decreases for repeated stimuli at delays ranging from mere milliseconds to days (Grill-Spector et al. 2006; van Turennout et al. 2000).

Despite the robustness of the findings on RS, debate continues on the neural and cognitive reasons for reduced neural responses to repeated stimuli and how they relate to long-term behavioral habituation. The two most prevalent (and opposing) explanations for RS are the bottom-up and top-down models (Valentini 2011). According to the bottom-up (or fatigue) model, RS is due to the refractory period of local neural generators in response to physical stimulation (Grill-Spector et al. 2006). In contrast, the top-down (or predictive coding) model holds that RS is not local in nature but instead is due to higher levels of cognition wherein the brain determines the expected probability with which a stimulus will occur (Mayrhauser et al. 2014). Recent research suggests that RS is likely a result of a combination of the bottom-up and top-down mechanisms (Hsu et al. 2014; Mayrhauser et al. 2014; Valentini 2011).

Eye Movement-Based Memory: An Eye Tracking-Based View of Habituation

Another manifestation of habituation is the eye movement–based memory (EMM) effect (Ryan et al. 2000). The EMM effect is apparent in fewer eye-gaze fixations and less visual sampling of the regions of interest within the visual stimulus. Memory researchers have discovered that the EMM effect is a pervasive phenomenon in which people unconsciously pay less attention to images they have viewed before. With repeated exposure, the memories become increasingly available, thus requiring less visual sampling of an image (Heisz and Shore 2008). People's attention fundamentally decreases in a systematic fashion with repeated viewings, even when they do not consciously recognize that they have seen an image before (Hannula et al. 2010). For these reasons, the EMM effect is a robust means of directly observing habituation to security warnings and of evaluating warning designs intended to reduce its occurrence.

Appendix B

fMRI and Eye-Tracking Experimental Details ▐▬▬▬▬▬

Equipment

MRI scanning took place at a university MRI research facility with the use of a Siemens 3T TIM Trio scanner. For each scanned participant, we collected a high-resolution structural MRI scan for functional localization, in addition to a series of functional scans to track brain activity during the performance of various tasks. Structural images for spatial normalization and overlay of functional data were acquired with a T1-weighted magnetization-prepared rapid gradient-echo (MP-RAGE) sequence with the following parameters: matrix size = 224×256; TR = 1900 ms; TE = 2.26 ms; field of view = 219×250 mm; NEX = 1; slice thickness = 1.0 mm; voxel size = $1 \times .977 \times .977$ mm^3; flip angle = 9°; number of slices = 176. Functional scans were acquired with a T2*-weighted gradient-echo echoplanar pulse sequence with the following parameters: matrix size = 64×64; field of view = 192 mm; slice thickness = 3 mm; TR = 2000 ms; 229 TRs; TE = 28 ms; number of slices = 39; voxel size = $3 \times 3 \times 3$ mm; flip angle = 90°. Slices were aligned parallel with the rostrum and the splenium of the corpus callosum. The first three volumes acquired were discarded to allow for T1 stabilization.

Eye-tracking data were collected on each scan using an MRI-compatible SR Research EyeLink 1000 Plus long-range eye tracker (see Figure B2) with a spatial resolution of 0.01° and sampling at 1,000 Hz. Eye movements were recorded for the right eye. A nine-point calibration routine was used to map eye position in order to screen coordinates prior to each scanning block. Eye-fixation data were processed with DataViewer software (SR Research Ltd., version 1.11.900) to identify fixations and saccades. *Saccades* were defined as eye movements that

met three parameters: eye movement of at least .1°, velocity of at least 30°/second, and acceleration of at least 8,000°/second. *Fixations* were defined as periods of time that were between the saccades and that were not part of blinks. Image size was normalized to subtend approximately 8.5° of the visual angle on the images' longest axis.

Engagement Check

To ensure that participants were attentive to the task in the scanner, they were instructed to rate the severity of the content of each item as it was presented to them; the answer choices available to them were "extremely severe," "somewhat severe," "somewhat not severe," and "extremely not severe." Answers were given any time during each trial by pressing a button on an MRI-compatible button box. Following Dimoka (2012), we performed two checks to ensure that participants were engaged in the task. First, we explored whether participants ranked each stimulus on the severity scale or ignored the ranking. We found that participants ranked stimuli 99.8% of the time, which is a strong indicator of engagement. Second, we explored whether participants ranked the security warnings as more severe than the software prompts, which would suggest that participants were giving thoughtful responses. A *t*-test indicated that participants did indeed report that the security warnings ($m = 2.998$, SD = 1.478) were more severe than the software prompts [($m = 2.366$, SD = 1.826), $t(13463) = 25.236$, $p < .001$, $d = 0.435$].

fMRI Data Analysis Details

Functional data were slice-time corrected to account for differences in acquisition time for different slices of each volume; then, each volume was registered with the middle volume of each run to account for low-frequency motion. A three-dimensional automated image registration routine—program 3dVolreg (Cox and Jesmanowicz 1999), which uses Fourier interpolation—was applied to the volumes to realign them with the first volume of the first series used as a spatial reference. Data from each run were aligned with the run nearest in time to the acquisition of the structural scan. The structural scan was then coregistered to the functional scans. Spatial normalization was accomplished by calculating a transformation from each subject's structural scan to a template brain with advanced neuroimaging tools (ANTs) and then applying the transformation to the structural and functional data for each subject. The experimental design is represented pictorially in Figure B1.

Behavioral vectors were created that coded for stimulus type (e.g., security warnings, general software images) and repetition number. These were then entered separately into single-participant regression analyses for each day. Stimulus events were modeled using a stick function convolved with the canonical hemodynamic response. Regressors that coded for motion and scanner drift were also entered into the model as nuisance variables. Spatial smoothing was conducted by blurring the resulting beta values with a 5-mm FWHM Gaussian kernel to increase the signal-to-noise ratio. Beta values for the conditions of interest were then entered into group-level analyses as we tested each hypothesis (below). Group comparisons were corrected for multiple comparisons using a voxel-wise threshold of $p < .02$ and a spatial-extent threshold of 40 contiguous voxels (1080 mm^3) for an overall corrected *p*-value < .05, as determined through Monte Carlo simulations (Xiong et al. 1995).
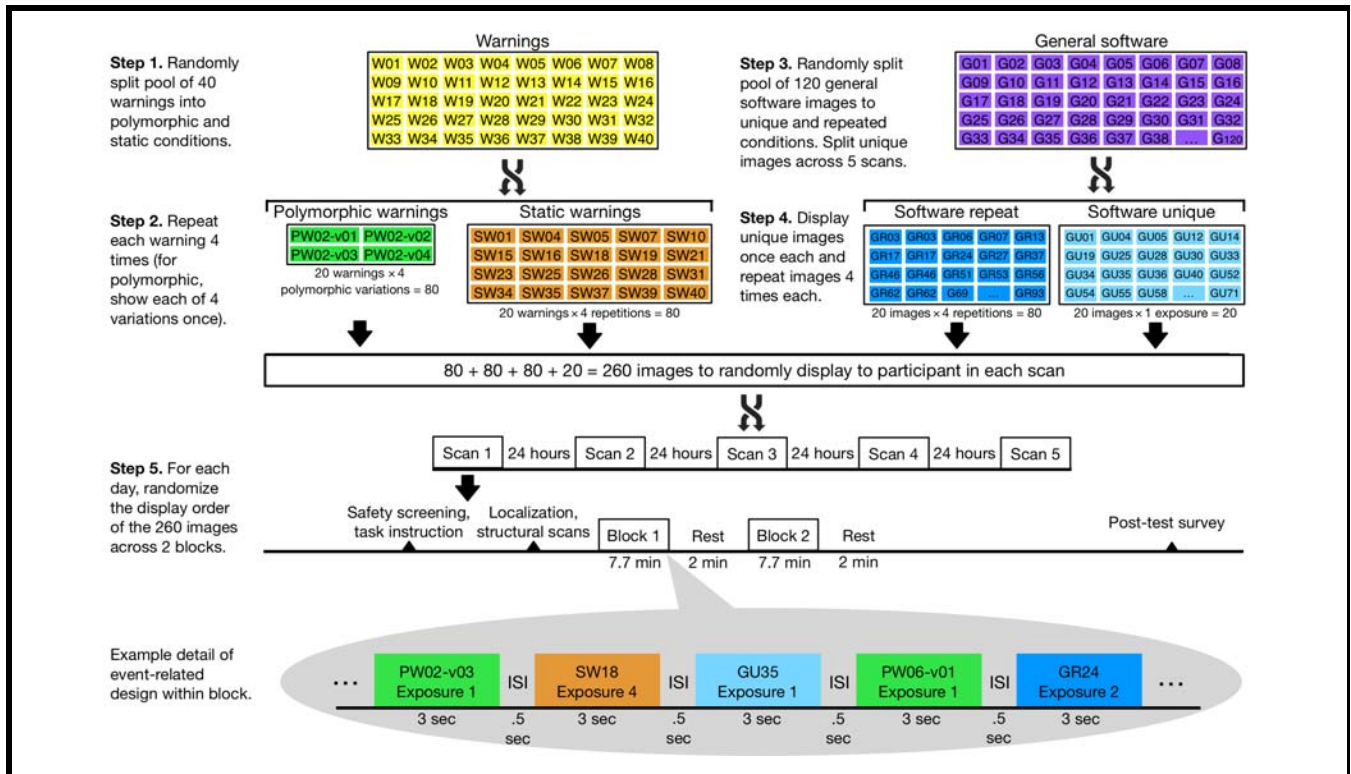
Figure B1. fMRI Repetition-Suppression-Effect (RSE) Longitudinal Protocol
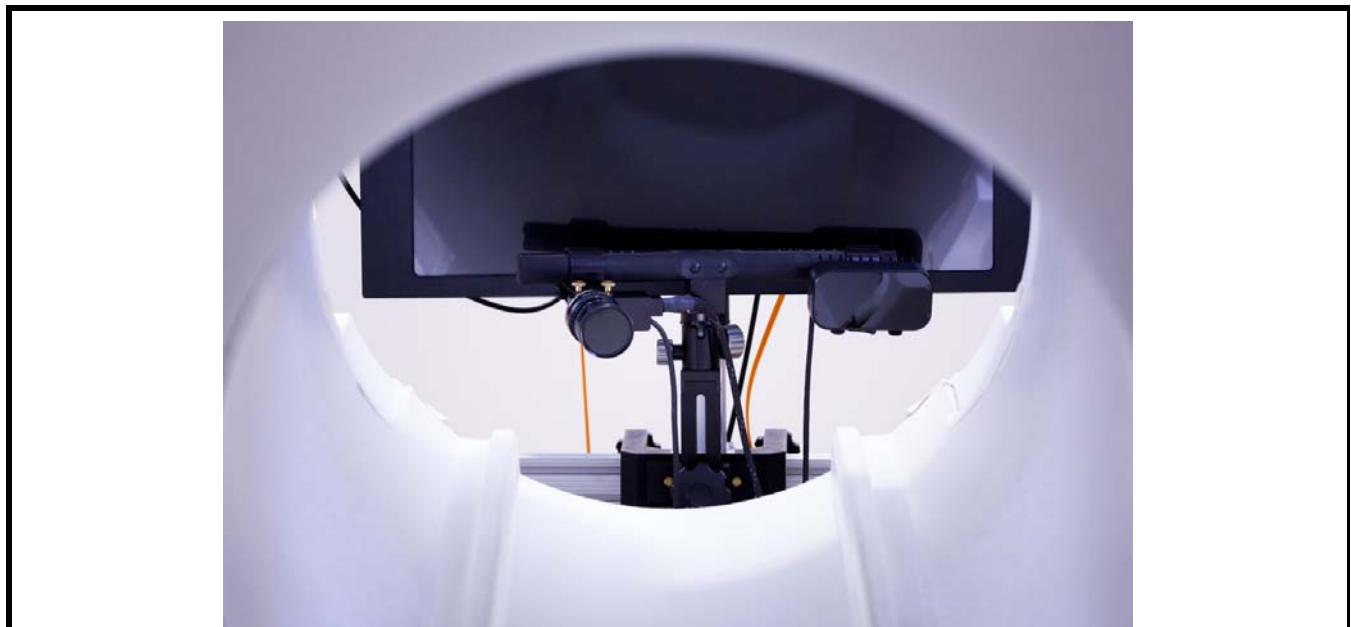

Figure B2. EyeLink 1000 Plus Long-Range Eye Tracker (mounted under the MRI viewing monitor)

Appendix C

Supplemental Analysis of Fixation Duration

In our main analysis, we tested the hypotheses with the eye-tracking data, using fixation count as the dependent variable. As a supplemental analysis, we tested our hypotheses using fixation duration, or the total time a person is fixated on a security message, as the dependent variable. The results are summarized below.

H1a Analysis: Users Habituate to Warnings Within an Experimental Session

We included fixation duration as the dependent variable; the subject ID, day number, and warning ID were included as random factors in a linear mixed-effects model. The presentation number was treated as a fixed factor, and visual complexity[1] was included as a covariate. The analysis supported H1a; the presentation number beta was significantly negative, indicating that habituation had occurred: $\chi^2(1, N = 11,976)$ = 174.37, $p < .001$, $\beta = -34.30$. However, visual complexity was not significant: $\chi^2(1, N = 11,976) = 2.84$, $p > .05$, $\beta = 12.25$. The R^2 of the model was .386.

H1b Analysis: Users Habituate Less to Polymorphic Warnings than to Static Warnings Within an Experimental Session

We specified the same mixed-effects model as in H1a, except that we included an interaction term between presentation number and determination of whether the warning was polymorphic. If the interaction was significantly positive, this would indicate that the participants habituated less to polymorphic warnings. The eye-tracking analysis supported H1b. Both main effects for presentation number [$\chi^2(1, N = 11,976) = 132.89$, $p < .001$, $\beta = -42.26$] and polymorphism [$\chi^2(1, N = 11,976) = 5.62$, $p < .01$, $\beta = -33.60$] were significant; in addition, the interaction was significant: $\chi^2(1, N = 11,976) = 2.95$, $p > .05$, $\beta = 15.94$. Visual complexity was not significant: $\chi^2(1, N = 11,976) = 82.79$, $p < .001$, $\beta = 12.48$. The R^2 of the model was .387.

H2a Analysis: Users Habituate to Warnings Across Days

In the linear mixed-effects model, we included fixation duration as the dependent variable, with the subject ID and warning ID as random factors. The presentation number (across days) was treated as a fixed factor, and visual complexity was included as a covariate. The eye-tracking analysis supported H2a; the beta of presentation number across days was significantly negative [$\chi^2(1, N = 11,976) = 769.24$, $p < .001$, $\beta = -15.35$], indicating habituation. Visual complexity was not significant: $\chi^2(1, N = 11,976) = 2.94$, $p > .05$, $\beta = 12.40$. The R^2 of the model was 0.269.

H2b Analysis: Users Habituate Less to Polymorphic Warnings than to Static Warnings Across Days

We specified the same mixed-effects model as in H2a, except that we included an interaction term between presentation number (across days) and determination of whether the warning was polymorphic (coded as 1 for polymorphic and 0 for static). The eye-tracking analysis supported H1b; the interaction between presentation number and polymorphic-warning type was significantly positive [$\chi^2(1, N = 11,976) = 3.25$, $p < .05$, $\beta = 1.962$], indicating that participants habituated less to polymorphic warnings across days than to static warnings. The main effect for presentation number [$\chi^2(1, N = 11,976) = 441.44$, $p < .001$, $\beta = -16.328$] was significant, but the main effect for polymorphism [$\chi^2(1, N = 11,976) = 1.03$, $p > .05$, $\beta = -13.405$] was not. Visual complexity was not significant: $\chi^2(1, N = 11,976) = 3.09$, $p > .05$, $\beta = 12.686$. The R^2 of the model was 0.153.

---

[1]Visual complexity was calculated using a script in MATLAB (Rosenholtz et al. 2007).

H3a Analysis: If Warnings Are Withheld After Habituation Occurs, the Response Recovers at Least Partially the Next Day

We subtracted the fixation duration for the first viewing of a warning on a day from the fixation duration for the last viewing of the warning on the previous day. We then conducted a *t*-test to test this hypothesis. The analysis supported H3a; participants experienced significant positive recovery ($m = 35.887$, SD = 483.903) from day to day: $t(2288) = 3.5482$, $p < .001$, $d = 0.148$.

H3b Analysis: If Warnings Are Withheld After Habituation Occurs, Response Recovery Is Stronger for Polymorphic Warnings than for Static Warnings the Next Day

We again subtracted the fixation duration for the first viewing of a warning on a day from the fixation duration for the last viewing of the warning on the previous day. We then specified a linear mixed-effects model that tested whether warning type (polymorphic versus static) predicted this difference. The subject ID, day interval (e.g., the difference between day 1 and day 2 was coded as 1), and warning ID were included as random factors. Polymorphism (coded as 1 for polymorphic and 0 for static) was included as a fixed factor, and visual complexity was included as a covariate. The eye-tracking analysis did not support H3b. Neither the warning type [$\chi^2(1, N = 2,400) = 2.05$, $p > .05$, $\beta = -24.14$] nor visual complexity [$\chi^2(1, N = 2,400) = 3.28$, $p > .05$, $\beta = 18.53$] significantly predicted recovery between days.

H4a Analysis: The Amount of Recovery Will Decrease Across Days

We again used the difference in the fixation duration for the last viewing of the warning on the previous day as our dependent variable. We specified a mixed-effects model that tested whether the day interval (e.g., 1 for the difference between days 1 and 2) predicted this difference. The subject ID and warning ID were included as random factors. The day interval was treated as a fixed factor, and visual complexity was included as a covariate. The eye-tracking analysis supported H4a. The day interval was significantly negative [$\chi^2(1, N = 2,400) = 9.77$, $p < .001$, $\beta = -27.73$], indicating that recovery decreased across days. Visual complexity was not significant: $\chi^2(1, N = 2,400) = 3.39$, $p > .05$, $\beta = 18.75$. The R^2 of the model was .079.

H4b Analysis: The Amount of Recovery Will Decrease Less for Polymorphic Warnings than for Static Warnings Across Days

We specified the same linear mixed-effects model as in H4a, except that we included an interaction term between day intervals (the difference between days) and another term for whether the warning was polymorphic (coded as 1 for polymorphic and 0 for static). The eye-tracking analysis did not support H4b. The main effect for day number [$\chi^2(1, N = 2,400) = 7.05$, $p > .01$, $\beta = 33.25$] was significant. However, the main effect for polymorphism [$\chi^2(1, N = 2,400) = 0.01$, $p > .05$, $\beta = -4.86$] and for the interaction were nonsignificant: $\chi^2(1, N = 2,400) = 0.40$, $p > .05$, $\beta = -11.12$. Likewise, visual complexity was nonsignificant: $\chi^2(1, N = 2,400) = 2.93$, $p > .05$, $\beta = 17.48$. Table C1 compares the results for fixation count, fixation duration, and fMRI data.

Table C1. Summary of Results			
Hypothesis	**Fixation Count**	**Fixation Duration**	**fMRI**
H1a: Users habituate to warnings within an experimental session.	Supported	Supported	Supported
H1b: Users habituate less to polymorphic warnings than to static warnings in an experimental session.	Supported	Supported	Supported
H2a: Users habituate to warnings across days.	Supported	Supported	Supported
H2b: Users habituate less to polymorphic warnings than to static warnings across days.	Supported	Supported	Supported
H3a: If warnings are withheld after habituation occurs, the response recovers at least partially the next day.	Supported	Supported	Supported
H3b: If warnings are withheld after habituation occurs, response recovery is stronger for polymorphic warnings than for static warnings the next day.	Not supported	Not supported	Not supported
H4a: The amount of recovery will decrease across days.	Supported	Supported	Not supported
H4b: The amount of recovery will decrease less for polymorphic warnings than for static warnings across days.	Not supported	Not supported	Not supported

Appendix D

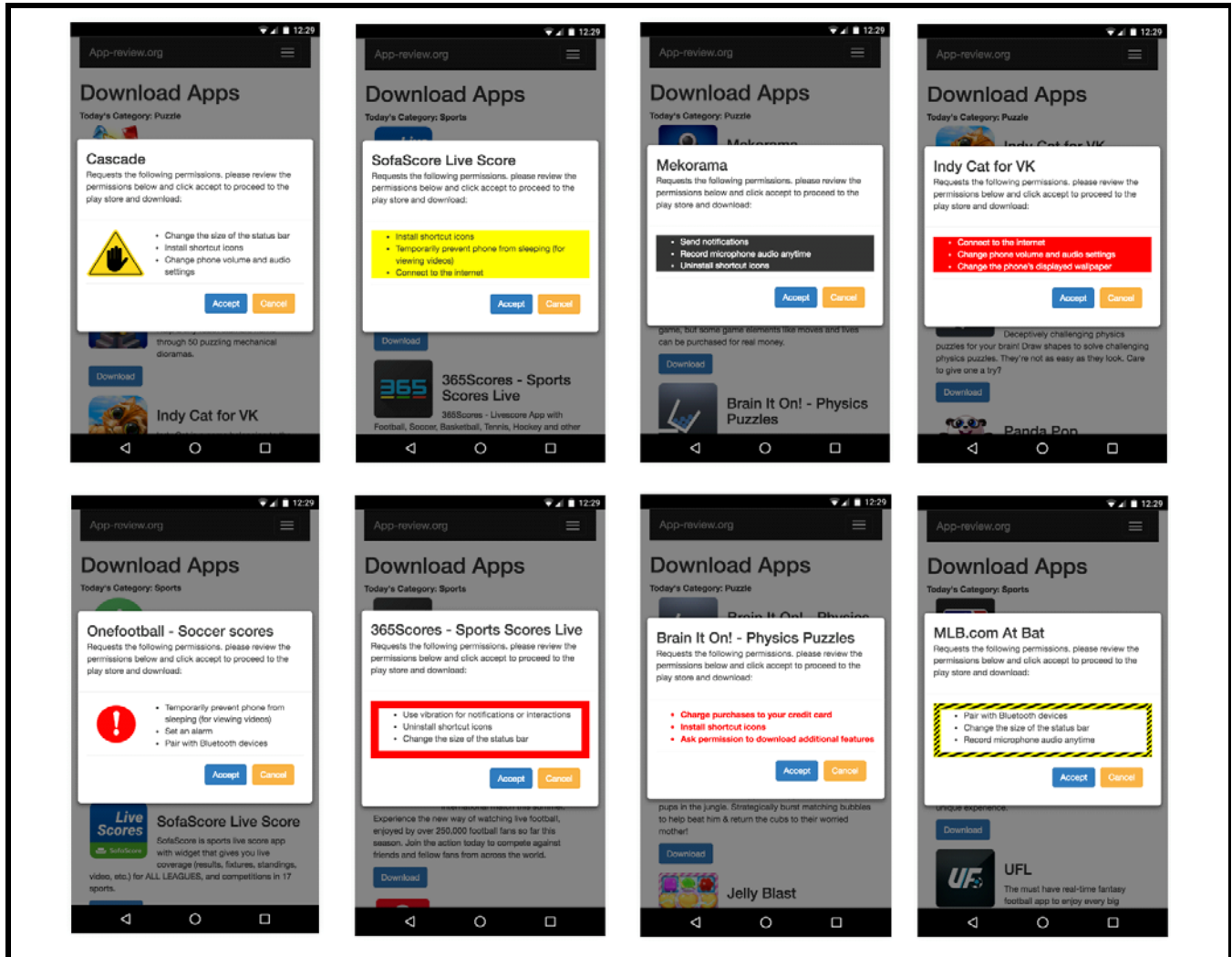## Polymorphic Warning Variations Used in Experiment 2 ▮▮▮▮▮▮

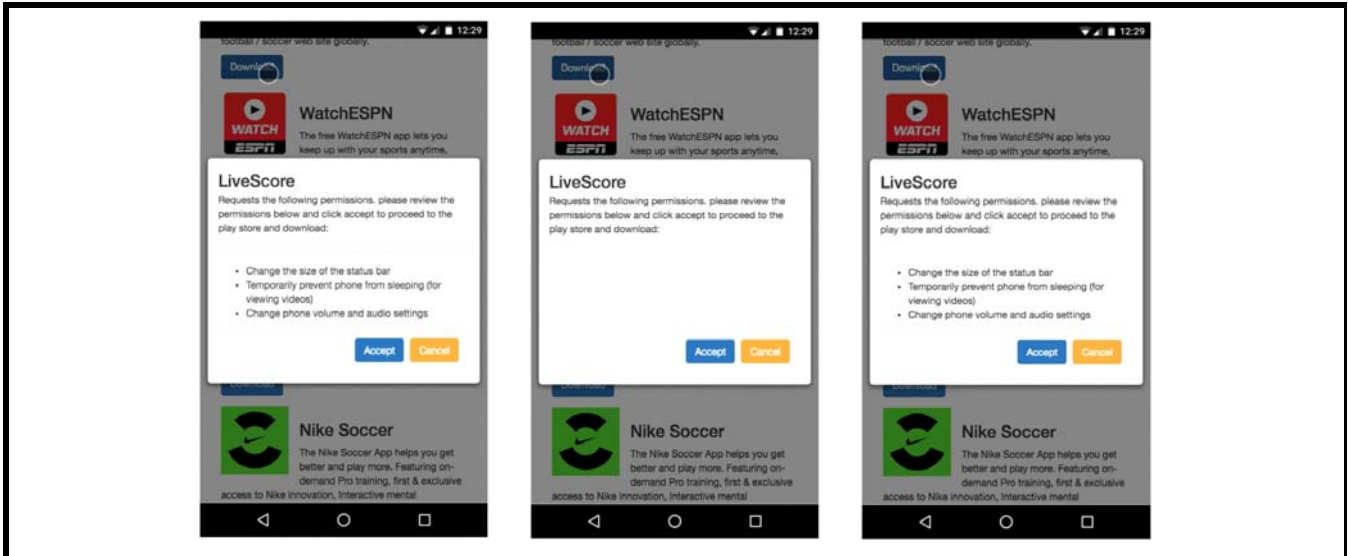**Figure D1. Nonanimated Polymorphic Variations**

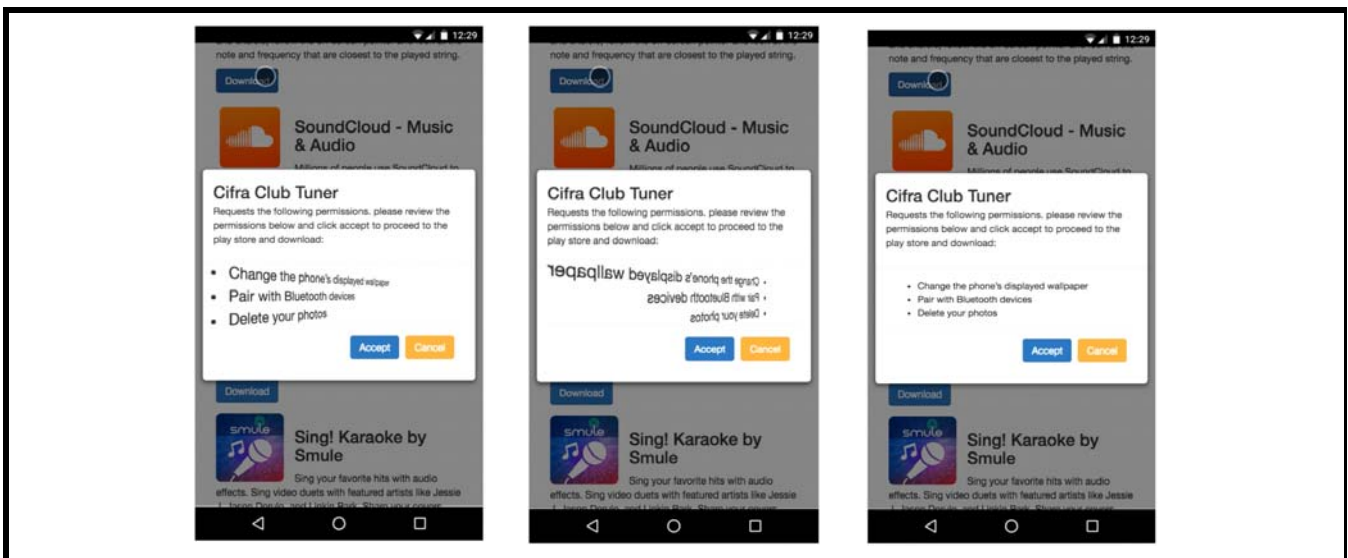**Figure D2. Animated Polymorphic Variation: Flash**
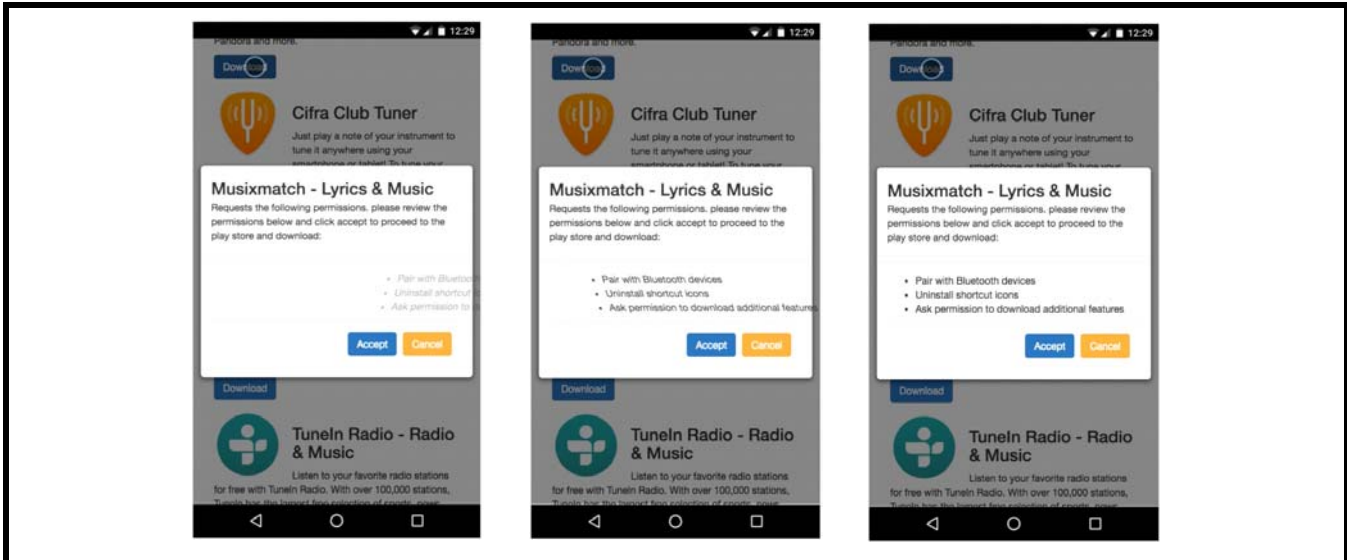

Figure D3. Animated Polymorphic Variation: Flip

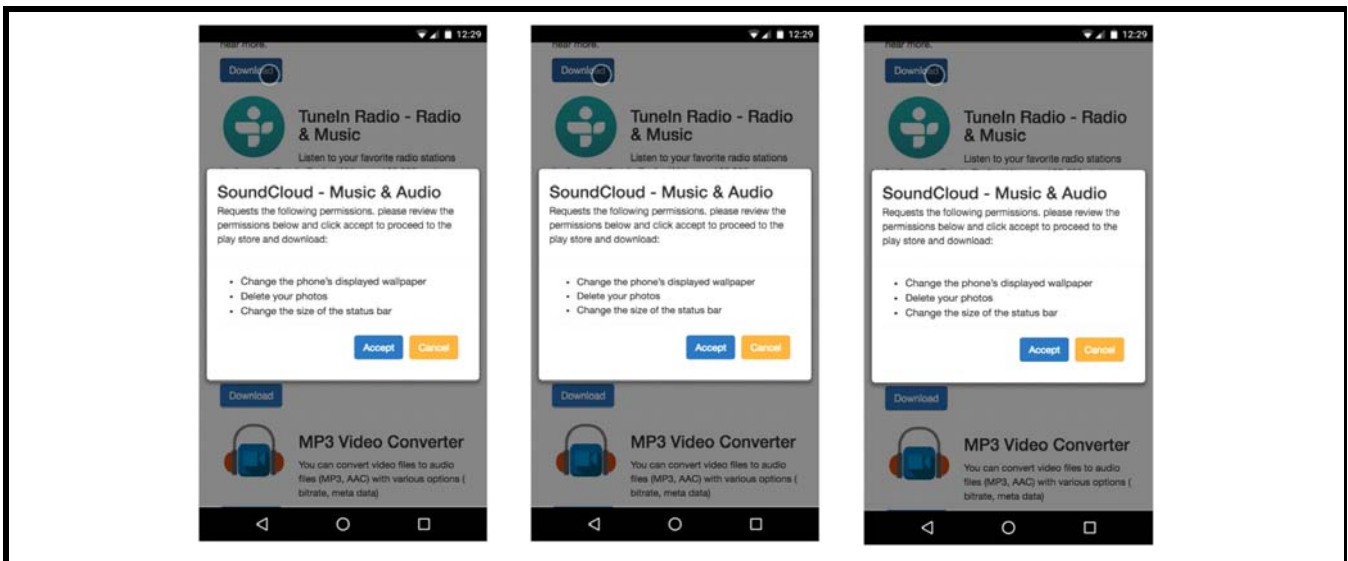**Figure D4. Animated Polymorphic Variation: "Light Speed In"**
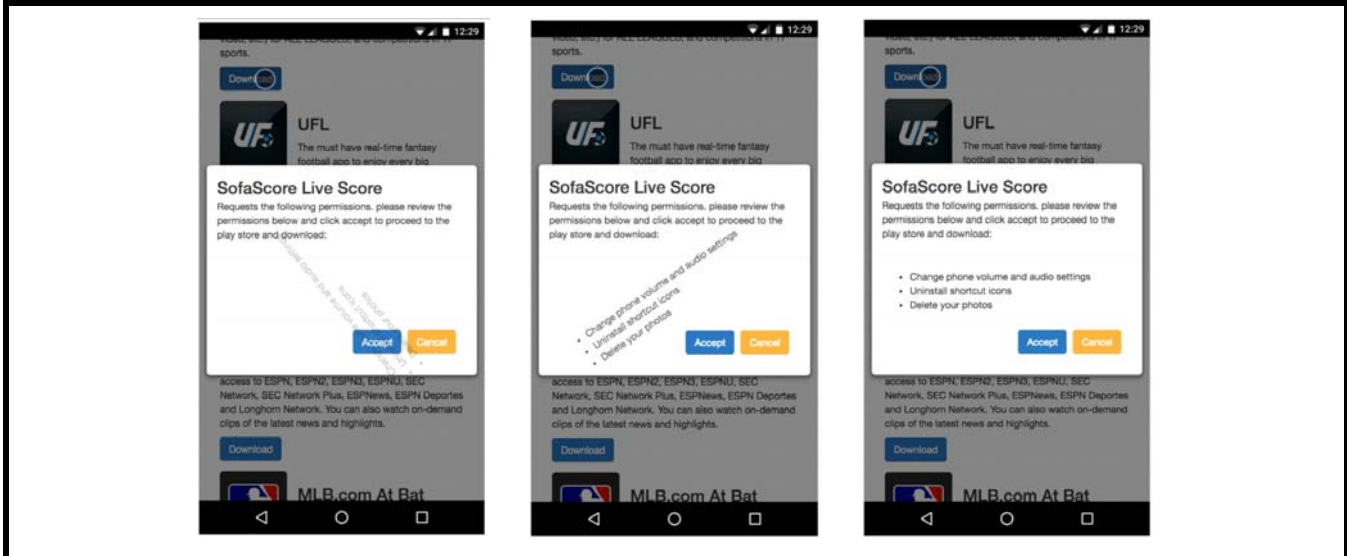

Figure D5. Animated Polymorphic Variation: Pulse

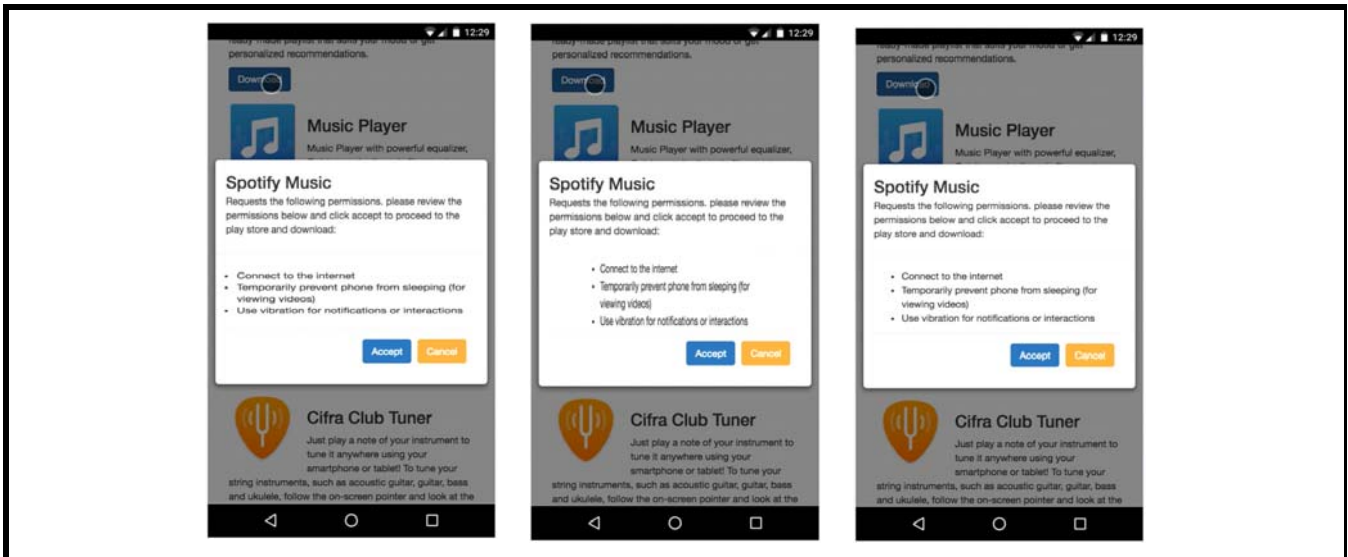**Figure D6. Animated Polymorphic Variation: "Rotate In"**

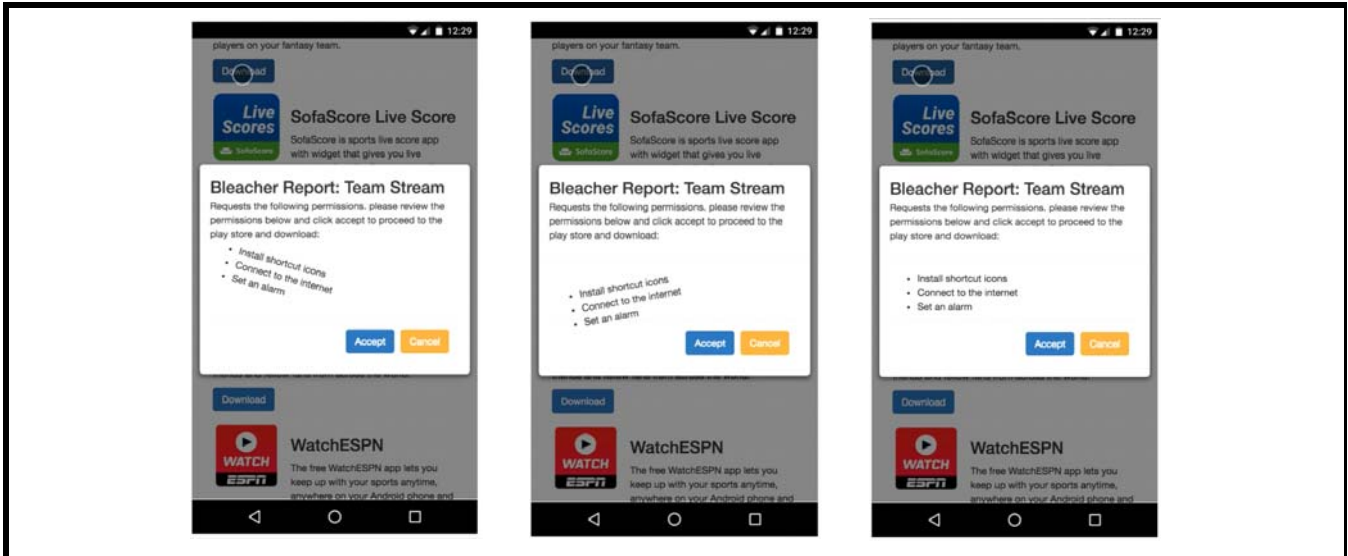**Figure D7. Animated Polymorphic Variation: "Rubber Band"**
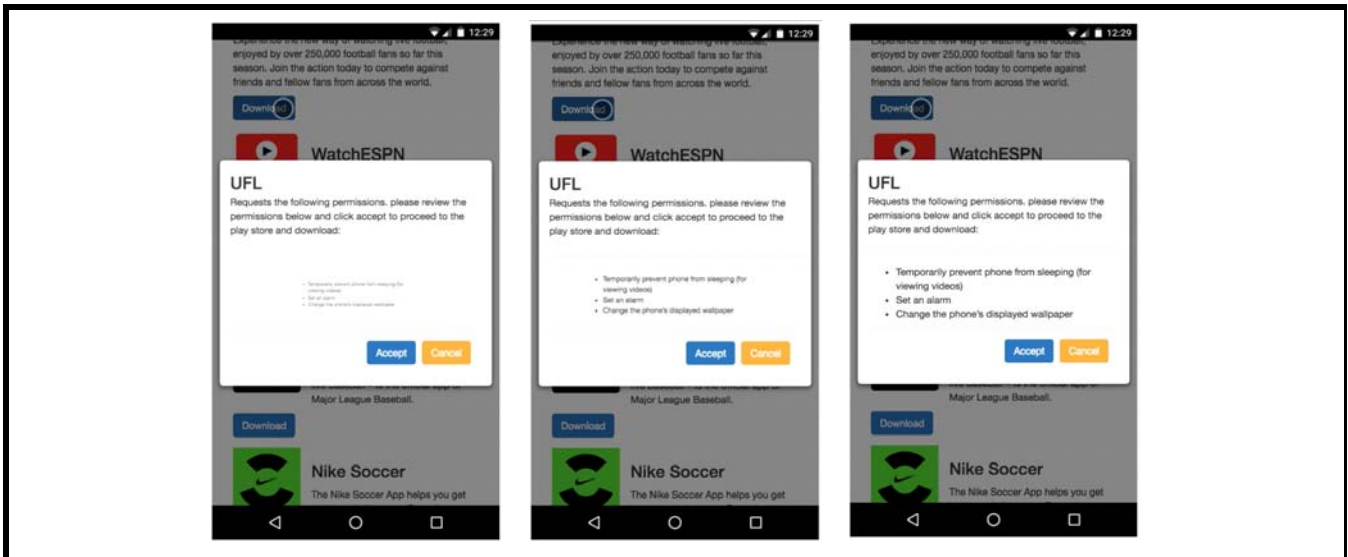
Figure D8. Animated Polymorphic Variation: Swing

Figure D9. Animated Polymorphic Variation: "Zoom In"

Appendix E

NeuroIS Contributions of Experiment 1 ▮▮▮▮▮▮▮▮▮▮▮▮▮▮

Experiment 1 used fMRI to measure underlying brain activity when individuals were presented with warning messages. As with any research method, the use of fMRI has both strengths and weaknesses. It is important to consider both, along with the purpose of the study, when evaluating the methodology of the experiment.

In information systems research, neurobiological methods can contribute to a deeper understanding of hidden (automatic or unconscious) mental processes; these mental processes are one of seven research opportunities for NeuroIS suggested by Dimoka et al. (2011). Habituation is one such mental process that does not require the conscious awareness of the organism (Groves and Thompson 1970; Sokolov 1963). As such, self-report measures are not useful because users may not even recognize that the process of habituation is occurring (Dimoka et al. 2011; Riedl and Léger 2016; vom Brocke et al. 2013).

Further, although behavioral methods can provide evidence that the artifact reduces warning adherence behavior, these approaches are not able to reveal the extent to which habituation in the brain plays a part in this behavior. Therefore, the use of fMRI yields deeper insights by moving past conscious experience toward looking at the neurobiological processes that are involved (Riedl and Léger 2016). fMRI allows the examination of neuronal processes through the blood oxygen level-dependent (BOLD) response, which is correlated with neuronal activation (Goense and Logothetis 2008), giving an accurate picture of which structures are involved and allowing for inferences about what to manipulate in the IT artifact. In Experiment 1, for example, we confirmed that even with long-term habituation, visual processing areas of the brain show strong habituation effects. Such a finding shows that users are not merely disregarding a warning when they are presented with it; they may not even notice that a warning is novel. This finding lends support to the need for a visual change in the IT artifact (Riedl and Léger 2016).

Finally, existing sources of data—in this case, eye tracking—were complemented in Experiment 1 with brain imaging data, another of the seven research opportunities enjoined by Dimoka et al. (2011). Although eye tracking has been used before to examine habituation to warnings in a cross-sectional experiment (Anderson et al. 2016), Experiment 1 contributes by having simultaneously collected eye-tracking and neuroimaging data over a five-day period. We found that the eye-tracking results closely mimicked the fMRI results, suggesting that eye tracking is a valid index of the mental process of habituation. This suggests that eye tracking is a cost-effective alternative to fMRI for studying habituation to warnings as a mental process, enabling future researchers to conduct less intrusive habituation studies that use eye tracking in a normal computing environment.

Appendix F

Ecological Validity Limitations of Experiment 1 ▮▮▮▮▮▮▮▮▮▮▮▮▮

The objective of Experiment 1 was to evaluate habituation of attention in the brain in response to static and polymorphic warnings. Accomplishing this purpose required a controlled laboratory setting to enable a precise test of habituation theory, as well as the use of an MRI scanner. The results of Experiment 1 provide unique insights that would not be possible using traditional behavioral methods, as explained in Appendix E. However, these neural insights come at the expense of ecological validity, which concerns whether an "effect is representative of what happens in everyday life" (Crano et al. 2015, p. 136).

There are two primary reasons for this tradeoff. The first is inherent in all fMRI studies. As Riedl et al. (2010, p. 255) observed:

> During an fMRI experiment, for example, participants are required to lie still on their back within the scanner while their head is restrained with pads to prevent head motion. Within the scanner, participants can use simple devices to react to the stimuli presented by pressing a button (e.g., to state a chosen alternative in a decision task). . . . An fMRI scanner is also relatively noisy, posing a potential distraction and making auditory stimulus presentation difficult. Therefore, experimental situations in fMRI studies are artificial, because in real life, computer users usually sit in front of their computer in a familiar, comfortable, and quiet environment.

This artificiality results in a high degree of intrusiveness compared to traditional behavioral methods or even other NeuroIS methods (Riedl et al. 2014).

The second tradeoff in ecological validity is specific to studying habituation to warnings. As explained in the discussion section, habituation is a neurobiological phenomenon that relies on the number of presentations of a stimulus in a given amount of time (Rankin et al. 2009). Since security warnings are relatively infrequent, displaying many warning presentations in a single experimental session, even if done in a proportionally accurate way (see Figure 16), may result in a pattern of habituation that is not representative of habituation to warnings in real life. This limitation applies to any laboratory experiment of habituation, regardless of whether it uses neuroimaging or traditional behavioral methods.[2] However, this problem is exacerbated by the requirement of the fMRI method for repeated trials of stimuli to ensure a reliable measure of the response (Turner et al. 2013). Repeated trials result in a high number of warnings being displayed in a single laboratory session.

Given the ecological validity limitations of fMRI and NeuroIS in general, Dimoka et al. (2012) recommended that researchers "replicate [NeuroIS experiments] in a more traditional setting and compare the corresponding behavioral responses to test for external validity" (p. 682), adding that "the richness provided by multiple sources of measures can be used to enhance the ecological validity of IS studies" (p. 695). Accordingly, one of our objectives for Experiment 2 was to conduct a realistic field experiment that we could compare against our fMRI experiment. Although the methods, dependent variables, and experimental designs of Experiments 1 and 2 were quite different, the results corroborate each other in terms of (1) the overall pattern of habituation, (2) the effectiveness of polymorphic warnings in reducing habituation, and (3) the effect of recovery between exposure to warnings (see Table 8). We therefore conclude that the results of Experiment 1 are reasonably accurate, despite the issues of ecological validity discussed above.

References

Anderson, B. B., Jenkins, J., Vance, A., Kirwan, C. B., and Eargle, D. 2016. "Your Memory Is Working Against You: How Eye Tracking and Memory Explain Susceptibility to Phishing," *Decision Support Systems* (25:4), pp. 3-13.

Çevik, M. Ö. 2014. "Habituation, Sensitization, and Pavlovian Conditioning," *Frontiers in Integrative Neuroscience* (8).

Christoffersen, G. R. J. 1997. "Habituation: Events in the History of Its Characterization and Linkage to Synaptic Depression. A New Proposed Kinetic Criterion for Its Identification," *Progress in Neurobiology* (53:1), pp. 45-66.

Colombo, J., and Mitchell, D. W. 2009. "Infant Visual Habituation," *Neurobiology of Learning and Memory* (92:2), pp. 225-234.

Cox, R. W., and Jesmanowicz, A. 1999. "Real-Time 3D Image Registration for Functional MRI," *Magnetic Resonance in Medicine* (42), pp. 1014-1018.

Crano, W., Brewer, M., and Lac, A. 2015. *Principles and Methods of Social Research* (3rd ed.), New York: Routledge.

Dimoka, A. 2012. "How to Conduct a Functional Magnetic Resonance (fMRI) Study in Social Science Research," *MIS Quarterly* (36:3), pp. 811-840.

Dimoka, A., Banker, R. D., Benbasat, I., Davis, F. D., Dennis, A. R., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P. H., Pavlou, P. A., Müller-Putz, G., Riedl, R., vom Brocke, J., and Weber, B. 2012. "On the Use of Neurophysiological Tools in IS Research: Developing a Research Agenda for NeuroIS," *MIS Quarterly* (36:3), pp. 679-702.

Dimoka, A., Pavlou, P. A., and Davis, F. D. 2011. "Research Commentary—NeuroIS: The Potential of Cognitive Neuroscience for Information Systems Research," *Information Systems Research* (22:4), pp. 687-702.

Goense, J. B. M., and Logothetis, N. K. 2008. "Neurophysiology of the Bold fMRI Signal in Awake Monkeys," *Current Biology* (18:9), pp. 631-640.

Grill-Spector, K., Henson, R., and Martin, A. 2006. "Repetition and the Brain: Neural Models of Stimulus-Specific Effects," *Trends in Cognitive Sciences* (10:1), pp. 14-23.

Groves, P. M., and Thompson, R. F. 1970. "Habituation: A Dual-Process Theory," *Psychological Review* (77), pp. 419-450.

Hannula, D. E., Althoff, R. R., Warren, D. E., Riggs, L., Cohen, N. J., and Ryan, J. D. 2010. "Worth a Glance: Using Eye Movements to Investigate the Cognitive Neuroscience of Memory," *Frontiers in Human Neuroscience* (4).

Hawco, C., and Lepage, M. 2014. "Overlapping Patterns of Neural Activity for Different Forms of Novelty in fMRI," *Frontiers in Human Neuroscience* (8).

Heisz, J. J., and Shore, D. I. 2008. "More Efficient Scanning for Familiar Faces," *Journal of Vision* (8:9).

Hsu, Y. F., Hamalainen, J. A., and Waszak, F. 2014. "Repetition Suppression Comprises Both Attention-Independent and Attention-Dependent Processes," *Neuroimage* (98), pp. 168-175.

Kaliukhovich, D., and Vogels, R. 2010. "Stimulus Repetition Probability Does Not Affect Adaptation in Macaque Inferior Temporal Cortex," paper presented at the 40th Annual Meeting of the Society for Neuroscience Abstract Viewer and Itinerary Planner, San Diego.

Mayrhauser, L., Bergmann, J., Crone, J., and Kronbichler, M. 2014. "Neural Repetition Suppression: Evidence for Perceptual Expectation in Object-Selective Regions," *Frontiers in Human Neuroscience* (8).

---

[2]We thank an anonymous reviewer and the associate editor for these insights.

Ramaswami, M. 2014. "Network Plasticity in Adaptive Filtering and Behavioral Habituation," *Neuron* (82:6), pp. 1216-1229.

Rankin, C. H. 2009. "Introduction to Special Issue of Neurobiology of Learning and Memory on Habituation," *Neurobiology of Learning and Memory* (92:2), pp. 125-126.

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., and Thompson, R. F. 2009. "Habituation Revisited: An Updated and Revised Description of the Behavioral Characteristics of Habituation," *Neurobiology of Learning and Memory* (92:2), pp. 135-138.

Riedl, R., Banker, R. D., Benbasat, I., Davis, F. D., Dennis, A. R., Dimoka, A., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P., Müller-Putz, G., Pavlou, P. A., Straub, D. W., vom Brocke, J., and Weber, B. 2010. "On the Foundations of NeuroIS: Reflections on the Gmunden Retreat 2009," *Communications of the Association for Information Systems* (27), Article 15, pp. 243-264.

Riedl, R., Davis, F. D., and Hevner, A. R. 2014. "Towards a NeuroIS Research Methodology: Intensifying the Discussion on Methods, Tools, and Measurement," *Journal of the Association for Information Systems* (15:10), pp. i-xxxv.

Riedl, R., and Léger, P. M. 2016. *Fundamentals of NeuroIS: Information Systems and the Brain*, Berlin: Springer.

Rosenholtz, R., Li, Y., and Nakano, L. 2007. "Measuring Visual Clutter," *Journal of Vision* (7:2), pp. 1-22.

Ryan, J. D., Althoff, R. R., Whitlow, S., and Cohen, N. J. 2000. "Amnesia Is a Deficit in Relational Memory," *Psychological Science* (11:6), pp. 454-461.

Schmid, S., Wilson, D. A., and Rankin, C. H. 2014. "Habituation Mechanisms and Their Importance for Cognitive Function," *Frontiers in Integrative Neuroscience* (8).

Sokolov, E. 1963. "Higher Nervous Functions: The Orienting Reflux," *Annual Review of Physiology* (25), pp. 545-580.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., and Egner, T. 2008. "Neural Repetition Suppression Reflects Fulfilled Perceptual Expectations," *Nature Neuroscience* (11:9), pp. 1004-1006.

Summerfield, C., Wyart, V., Johnen, V. M., and de Gardelle, V. 2011. "Human Scalp Electroencephalography Reveals That Repetition Suppression Varies with Expectation," *Frontiers in Human Neuroscience* (5).

Todorovic, A., and de Lange, F. P. 2012. "Repetition Suppression and Expectation Suppression Are Dissociable in Time in Early Auditory Evoked Fields," *Journal of Neuroscience* (32:39), pp. 13389-13395.

Turner, B. O., and Miller, M. B. 2013. "Number of Events and Reliability in fMRI," *Cognitive, Affective, & Behavioral Neuroscience* (13:3), pp. 615-626.

Valentini, E. 2011. "The Role of Perceptual Expectation on Repetition Suppression: A Quest to Dissect the Differential Contribution of Probability of Occurrence and Event Predictability," *Frontiers in Human Neuroscience* (5).

van Turennout, M., Ellmore, T., and Martin, A. 2000. "Long-Lasting Cortical Plasticity in the Object Naming System," *Nature Neuroscience* (3:12), pp. 1329-1334.

Verplanken, B., and Aarts, H. 1999. "Habit, Attitude, and Planned Behaviour: Is Habit an Empty Construct or an Interesting Case of Goal-Directed Automaticity?," *European Review of Social Psychology* (10:1), pp. 101-134.

Verplanken, B., and Orbell, S. 2003. "Reflections on Past Behavior: A Self-Report Index of Habit Strength," *Journal of Applied Social Psychology* (33:6), pp. 1313-1330.

Vidyasagar, R., Stancak, A., and Parkes, L. M. 2010. "A Multimodal Brain Imaging Study of Repetition Suppression in the Human Visual Cortex," *NeuroImage* (49:2), pp. 1612-1621.

vom Brocke, J., Riedl, R., and Léger, P. M. 2013. "Application Strategies for Neuroscience in Information Systems Design Science Research," *Journal of Computer Information Systems* (53:3), pp. 1-13.

Xiong, J. H., Gao, J. H., Lancaster, J. L., and Fox, P. T. 1995. "Clustered Pixels Analysis for Functional MRI Activation Studies of the Human Brain," *Human Brain Mapping* (3:4), pp. 287-301.