Replication: Do We Snooze If We Can't Lose? Modelling Risk with Incentives in Habituation User Studies

Karoline Busse
University of Bonn
busse@cs.uni-bonn.de

Dominik Wermke
Leibniz University Hannover
wermke@sec.uni-hannover.de

Sabrina Amft
University of Bonn
amft@cs.uni-bonn.de

Sascha Fahl
Leibniz University Hannover
fahl@sec.uni-hannover.de

Emanuel von Zezschwitz
University of Bonn
zezschwitz@cs.uni-bonn.de

Matthew Smith
University of Bonn
smith@cs.uni-bonn.de

***Abstract*—Users of computer systems are confronted with security dialogs on a regular basis. As demonstrated by previous research, frequent exposure to these dialogs may lead to habituation (i.e., users tend to ignore them). While these previous studies are vital to gaining insights into the human factor, important real-world aspects have been ignored; most notably, not adhering to security dialogs has barely had a negative impact for user study participants. To address this limitation, we replicate and extend previous work on the habituation effect. Our new study design introduces a monetary component in order to refine the study methodology on habituation research. To evaluate our approach, we conducted an online user study ($n = 1236$) and found a significant effect of monetary loss on the compliance to security dialogs. Overall, this paper contributes to a deeper understanding of the habituation effect in the context of warning dialogs and provides novel insights into the complexity of ecologically valid risk modeling in user studies.**

I. Introduction

Dialog windows are part of the general user experience of interactive computer systems. Specifically, they are used as part of operating systems' security measures, such as the User Account Control (UAC) mechanism in Windows 10 [35], the OS X Gatekeeper feature [5], and Android permission dialogs [4]. Such security dialogs are also part of application programs, such as web browsers and mobile applications. In this regard, they are commonly used to warn users about insecure TLS connections, malware-infected websites, or weak passwords. While security dialogs are essential to overall information security and thus user data privacy, they suffer from inherent limitations. For example, false-positives are a serious problem as they unsettle users [2] even if there is no real risk or threat. On the other hand, users tend to perceive security dialogs as rather annoying and ignore them by clicking through them, even if risks are present [7], [12], [32]. Besides identifying and reducing root causes of false-positives [1], it is an important goal for usable security and privacy research to design security dialogs that prevent such habituation effects and are harder to ignore [2], [11].

Previous work has already proposed to leverage design principles in order to increase adherence to warnings [24], and user reactions to security dialogs have been thoroughly investigated [23], [31]. In particular, Bravo-Lillo et al. [14] have researched the habituation effect on system security dialogs. While their paper provides valuable insights into the design of habituation-resistant dialogs, we argue that an important real-world factor has been ignored by their research. In a user's everyday life, ignoring a valid system security dialog might infect their computer with malware or other unwanted software. However, in previous work, falsely clicking through a security dialog has had barely any consequences for the user [14].

In this paper, we aim to address this specific limitation of previous work concerning the habituation effect. First, we replicate the study of Bravo-Lillo et al. reusing their study infrastructure. Then, we perform a second user study that is an extension of the original study. In the second study, we add a new variable to the design: we propose a risk model that substitutes real-world risks, like data loss, with the consequence of monetary loss. Consequently, adhering to a system security dialog brings the participant bonus money, while ignoring instructions resets the bonus back to zero. These micro-reward systems have been long been used in psychology and behavioural economics research to enhance study designs and model risk of all sorts [11], [22]. For ethical reasons, all participants were paid independently from their performance with the maximum achievable amount of money after the study was finished.

The study results (1) confirm previous findings that both habituation and visual attractors influence the rate of (non-)compliant decisions in the replication as well as in the revised study design. Furthermore, we show that (2) monetary incentives have a significant influence on reducing the number of non-compliant answers to dialogs. This effect may be related to the exact modeling of the bonus, as our results show that (3) a higher amount of money gained per click has a greater effect than a small bonus. In addition, we found a small but

Workshop on Usable Security (USEC) 2019
24 February 2019, San Diego, CA, USA
ISBN 1-891562-57-6
https://dx.doi.org/10.14722/usec.2019.23001
www.ndss-symposium.org

significant effect on the extent to which a participant's first loss impacts their subsequent behavior, indicating that (4) bad experiences shape a person's attention, at least for a short period of time.

Thus, our research provides novel insights into risk modeling in online user studies and contributes to a deeper understanding of the habituation effect in regard to warning dialogs.

As with other studies, this one suffers from some limitations that affect the generalizability of the results. For example, to stay faithful to the study we were replicating, our study elevates the security tasks to be the participants' primary tasks, rather than the secondary tasks they would be in practice. We also incentivized participants monetarily, which also isn't a good match for the incentives that users would perceive in practice. These and other limitations are discussed in Section V in greater detail.

II. Background

Here we will discuss background in the following areas: research on habituation effects and warnings and the psychological effect of monetary incentives. Relevant work by Bravo-Lillo et al. on system security dialogs research is discussed in Section III, "Study Design".

A. Habituation Effect Research

When users are frequently confronted with security dialog decisions, their attention as well as their risk perception declines, and a so-called *habituation effect* arises.

Habituation is a form of learning described as "a decrease in the strength of a naturally elicited behavior that occurs through repeated presentations of the eliciting stimulus" [13]. It was coined by Humphrey [28] and Harris [27] and expanded by Thompson and Spencer, who presented nine characteristics to classify habituationamong them the ability to recover from habituation over time and the impact of weak and strong stimuli on habituation [39].

Several studies have focused on the habituation effect of dialog windows [12], [16], [23]. In 2008, Egelman et al. conducted a study about the effectiveness of phishing warnings by exposing participants of a laboratory study to a spear phishing email that triggered the browser's phishing warning page. A substantial number of users were susceptible to the attack, in part because they were habituated to browser warnings. The authors recommend five improvements for phishing warning design, which included the need to prevent habituation [23].

Brustoloni et al. tried to customize email agent warnings based on the context in which they appear to combat the habituation effect. In addition to warning customization, they also experimented with audited responses where they would tell the user that others could read their answers in a dialog. In a laboratory study, these adaptations proved resilient to habituation, but because of their customized nature and the high amount of extra auditing work needed, they are unrealistic to deploy in many areas of internet life [16].

In 2010, Böhme and Köpsell conducted a field study with users of an online anonymization tool and tested the effect of differently designed consent dialogs. The results indicate that participants were more likely to agree with messages that looked like typical license agreements, which indicates the habituation effect on license agreement dialogs [12].

Anderson, Vance, et al. researched the habituation effect in security with several studies employing functional magnetic resonance imaging (fMRI) of the brain. When observing the reaction to a habituated warning, activity in the virtual processing center of the brain drops. By employing polymorphic warnings, they could mitigate the replication effect [3]. A longitudinal study of this measure over the course of one week confirmed these results [40].

B. Research on Monetary Incentives

Various research in psychology, behavioral economics, and computer science has shown that, for many people, monetary incentives are of equal or even greater value than protecting personal information.

Research by Gehring and Willoughby revealed that monetary losses resulted in higher brain activity than gains, which suggests that assessment of decision situations is particularly sensitive to losses, validating Kahnemann and Tverski's research [29]. The researchers conducted an experiment in which participant had to choose repeatedly between 5 Cents and 25 Cents, not knowing which one would result in a loss or gain towards their study compensation. One second after the decision, the numbers were displayed either on red or green background, indicating either an increase of study compensation by 5 or 25 cents, or decreasing the compensation by the respective amount. During the experiment, the brain's medial frontal cortex was monitored and its activity evaluated [25].

Della Libera et al. conducted another psychological experiment, researching if a higher monetary reward leads to increased motivation [21]. Participants were shown pictures with different shapes and symbols. Beforehand, they were instructed to focus on a specific feature of the picture. They were told that performance would be measured and monetarily rewarded. Although the reward level had no direct influence on the participant's performance in the subsequent task, the attention devoted to the asked features was much greater when the participants thought they would be rewarded highly [21].

Work from the field of behavioral economics by Kahneman and Tverski researches the choice between two offers under various configurations. If one option would earn a participant a certain amount of money for sure and the other option would earn them more than double the amount, but only with a 50% chance, participants in general preferred the option with the guaranteed money. Kahneman and Tverski call this the *certainty effect*. When it comes to a choice between losses, participants tend to seek the riskier choice, even if it implies higher monetary losses. This mirrors the results of the certainty effect experiment and thus is labeled the *reflection effect* [29].

Another experiment from the field of behavioral economics investigates the monetary value of private information. Beresford et al. let participants choose between two offers from online DVD stores. The stores' programs and services were identical, but one store charged 1 euro less per DVD. However, the cheaper store required its customers to enter their birth date and income before placing their order. Despite the potential revelation of sensitive information, almost all participants chose the cheaper store [10]. This indicates that although people state that privacy is important to them, the barrier for giving data in exchange for money is rather low in practice.

This is confirmed by various other research. A study by Grossklags et al. showed that people tend to agree to sell their

personal data for very low amounts of money. In a quantitative study, the researchers collected data, like the participant's weight, their favorite holiday destination, and the number of sex partners they have had, during an introductory test and later offered them money in exchange for an information disclosure agreement in front of the whole participant group. Most participants agreed to sell for even the smallest available reward of 25 cents [26]. Work by Danezis and Cvrcek, including field studies, confirmed this permissive behavior for location data [19], [20]. These findings indicate that monetary incentives are often regarded as even more valuable than personal data.

Regarding the impact of immediate monetary loss, various research in psychology has demonstrated that, when it comes to learning from mistakes, there is a strong connection between the time and spatial distance of a wrong decision and its consequences [6], [30], [36], [41], [42]. The most prominent among these is construal level theory, which describes the connection between psychological distance and mental abstraction of an event [34]. These findings suggest that while monetary incentives can work as an equivalent for data loss, the exact amount and timing of said loss have a grave impact on the user's learning process.

Based on the insights of related work, we built a model that simulates the risk of data loss with the risk of losing money to assess the habituation effect in a more realistic way. While psychological theory shows that determining an exact monetary equivalent is nearly impossible [34], various literature indicates that monetary incentives are a valid methodological tool to model a more arbitrary risk [22], [29].

III. Study Design

A. Previous Studies

Our work builds mainly upon research by Bravo-Lillo et al. Their work from 2013 focused on the design of *attractors*, "user interface elements designed for attracting users' attention to critical information in a security-decision dialog" [15]. Bravo-Lillo et al. conducted three controlled experiments to understand and improve security dialogs. The third study, and subject of this work, focused on the habituation resistance of selected attractors.

In the first experiments, participants were prompted with a recreated Windows security dialog featuring various attractors, based on features like color or contrast, mouse movement, or typing to activate the dialog's answer options. In a between-groups design, the authors found that all but one of the proposed attractors significantly influence the user decision [15].

After having tested these attractors in a first-contact study, Bravo-Lillo et al. conducted a third study about the habituation effect of these attractors, which can be classified as a microworld study [9]. Users were to dismiss a number of simulated pop-up dialogs with a yes-no decision (see Figure 1). Participants were asked to answer as many dialogs as possible in a given time limit with the "yes" option. After a certain habituation period, the "no" option was enabled and the introductory sentence within the dialog window told the participants to choose it in order to end the study early. The metric used for evaluation was the proportion of users who chose the "no" option the first time it appeared.

Evaluation of the experimental data reveals that if a user interaction (mouse movement or some typing) is required before the dialog can be answered, users are significantly more likely to act on the "no" option the first time it is encouraged. In contrast, passive attractors that only modify colors and contrast are as susceptible to habituation as the control group [15].

One year later, Bravo-Lillo et al. published a follow-up paper extensively studying the habituation effect of certain attractors [14]. A new study was conducted that included different habituation periods for all featured attractors, which enabled the authors to monitor the habituation effect over the number of exposures per user. While the study from 2013 used a fixed habituation period, the new study used four different habituation periods per attractor in a between-groups design. A dialog was either presented 1, 3, or 20 times, or the user had to answer dialogs for 150 seconds before the option to finish the study early appeared. Again, several attractors were tested in a between-groups experiment with workers from Amazon Mechanical Turk (MTurk).

The extended experiment shows that some attractors are more susceptible to habituation than others; again, the more interactive an attractor is, the better it resists habituation. They also evaluated additional effort when interacting with different attractors over a longer period of time and identified methods of keeping the users' attention to dialog contents that are both efficient and resistant to the habituation effect.

The results of Bravo-Lillo et al. are beyond doubt very important insights for the HCI and usable security and privacy communities, but we regard the scenario as too artificial. The authors note this limitation as well, but they focus on artificial habituation over a very short period of time [14]. What also matters in the chosen task of answering system security dialogs is the perceived riskthe consequence of a wrong answer.

B. Use Case and Attractors

For this paper, we first partially replicated the original study by Bravo-Lillo et al. using the most promising conditions; afterward, we conducted a follow-up study to examine the effect of monetary incentives on habituation.

For this we created the following model. In the real world, users are faced with warnings that can be either true- or false-positives. If a user clicks through a false-positive warning, they correctly and safely achieve their desired goal, such as to read a website, to install a piece of software, etc. If they do not click through a false-positive warning, they usually suffer no harm, but they also do not gain the benefit of achieving their primary goal. If, however, a user clicks through a true-positive warning, they are likely to suffer negative consequences, such as to be phished, to install malware on their system, etc. Based on our literature review, we argue that this risk model can be simulated with monetary incentives and losses (cf. Section II). We model this by offering users a small monetary incentive for clicking through a false-positive warning to represent the beneficial aspect of achieving the primary goal. This monetary bonus accumulates over the course of the study. If a user heeds a false-positive warning, the bonus does not change. This should reflect that the user did not achieve their primary goal, but they also did not endanger their system. While we acknowledge that in daily life, not achieving this goal is often perceived as a great immediate loss, we chose to frame the gain as neutral in our experiment, also attributing the study's artificial setting. If, however, the user clicks through a true-positive warning, they

lose the entirety of their bonus accumulated so far. This should represent the negative effect of being phished or installing malware. More details, such as the size of the small bonus payment, are described in the next section. All differences and changes in relation to the original work are listed in Table I.

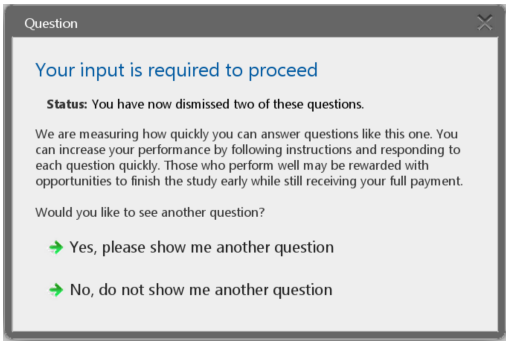
A key element of the study is customized dialog pop-ups mimicking Windows security dialogs. In the original study, the dialogs contained a headline and a status notification that at first displayed a statistic about the number of already dismissed pop-ups but changed to an important instruction on the answer choice during the test period. Below the status field was a short description of the task and in the bottom of the window were the answer options labeled "Yes" and "No". In comparison to the design used by Bravo-Lillo et al. (as depicted in Figure 1b), we removed the upper-right "X" button since our replication data indicates that this non-functioning element confuses users and results in them trying to dismiss the message via this button rather than engaging with the study. In addition, we adjusted the texts and descriptions to our design for our extended replication (cf. "Extended Study" section).

Previous work by Bravo-Lillo et al. has established several *attractors*, mechanisms to direct a user's attention to an important part of the security dialog, referenced as the so called *salient field*. In our experiment, as well as previous studies, the salient field contains the most important information in a dialog window [14].

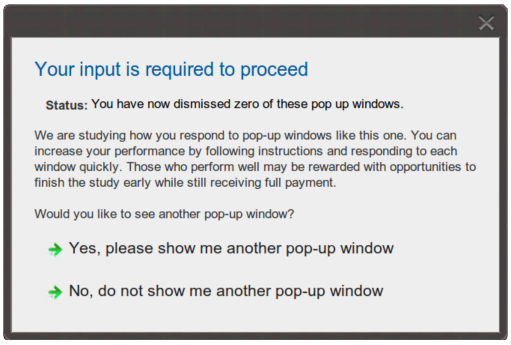
The *Swipe* attractor disables the "Yes" option of the dialog until the user has moved their cursor in a left-to-right fashion over the status message. If the user hovers over the answer options, a tutorial about the mechanism pops up (see also Figure 2. The *Type* attractor requires the user to type the contents of the status field in a text box directly below it. Until the text box contains the exact contents of the status field, the "Yes" option of the dialog remains deactivated. Pasting into the text box is not possible.

In the experiment by Bravo-Lillo et al., results in compliance as well as response time showed that in contrast to other tested attractors, the Type and Swipe attractors led to greatly increased compliance rates while resisting habituation. However, Type also led to significantly increased response times whichin contrast to Swipe did not improve with longer exposures. Therefore, the authors recommend the Swipe attractor as most usable while being resistant to the habituation effect. With regard to the results explained above, we chose the Swipe attractor as our subject of further research, along with the obligatory control group.

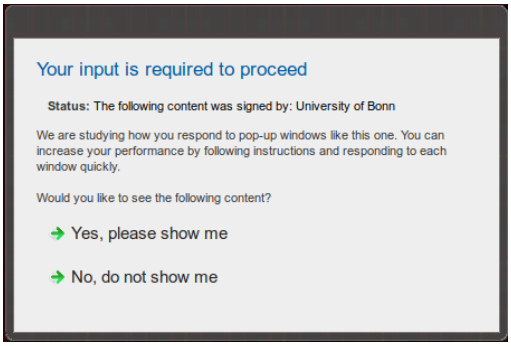
In order to be able to change the original study minimally, we contacted Christian Bravo-Lillo, who was able to share his protocol and study software with us. However, he was unable to share his raw data. We did an exact replication with this study platform and parameters, and afterwards adapted the implementation to our own use cases for the extended study. We implemented a bonus mechanism which awards a small amount of money for each compliant click on the "Yes" option and a loss of the accumulated bonus for an incorrect "No" decision as described above. For this setting, we inserted additional content dialogs showing an inspirational quote which was gate-kept by the warnings. Also, we added the currently accumulated bonus to the status bar of the experiment (cf. Figure 2).



(a) A security dialog used in the first habituation experiment of Bravo-Lillo et al. The answer options may be inactive (greyed out) depending on the attractor and habituation conditions. [15]



(b) A security dialog from the second study by Bravo-Lillo et al. The term "questions" was replaced by "pop up windows", and the window title disappeared, otherwise the content remained the same. [14]



(c) Our new security dialog without the upper-right "X" and with adjusted status message, answer options, and description text.

Fig. 1: Pop-up designs in previous work and in our new study.

C. Replication Study

To have sufficient data for evaluating additional characteristics of Bravo-Lillo et al.'s study design and to confirm previous work, we did an exact re-run of their platform and task with the set of tested attractors limited to *Control* and *Swipe* (see also the previous subsection) and omitted the 150-second habituation condition, since the other conditions were not used in our main experiment either (see also Section III-B).

We found that the timeout of 30 seconds before a participant dropped out of the experiment was barely sufficient for them to carefully read and examine the first dialog, so we removed the inactivity timer from the study design.

D. Extended Study

For our extended study, another online experiment was conducted on Amazon MTurk in a between-groups design, where groups were split by three variables: *habituation period*, *attractor*, and *bonus increase per dialog* (see also Table II). Previous work used only two variables, *habituation period* and *attractor*. In the following paragraphs, the study will be explained in greater detail.

The *habituation period* condition in our experiment is the same as in previous work. We have three different habituation levels that determine the number of dialogs a participant encounters during the habituation phase. Depending on the assigned group, the habituation period spans either one dialog, three dialogs, or twenty dialogs. The condition with a habituation period of 150 seconds has been omitted as explained above [14].

Regarding the *attractor* condition, Bravo-Lillo et al. introduced several mechanisms to direct a user's attention to the important part of a dialog (see also Section III-B). In previous work, five different attractors were tested, but results showed that several of the proposed attractors did not lead to the desired habituation resistance. While Type has been widely regarded as annoying among the participants, it showed increasing compliance rates with the number of dialogs. Swipe resulted in similar compliance trends, but participants quickly became efficient in using the attractor, halving the response time between three exposures and twenty exposures [14]. Because of these results we decided to include Swipe as the only attractor in our study design besides the obligatory control group for the attractor variable.

Our third variable, *bonus increase per dialog*, is newly added to our replication study and models the risk of a malware infection as a consequence of a wrong answer to a security dialog (cf. section II). In two scenarios, participants can accumulate a bonus payment of up to $\$0.50$, the variable determines the increase of said bonus per correct dialog answer: either 2.5 cents or 10 cents. If a "malicious" dialog is answered with "yes", the whole bonus is lost. As in the real world, a click on the "no" option never has any negative consequences (besides not getting access to the content) and leaves the bonus counter untouched. This emulates the risk of a malware infection and the resulting loss of personal data. In order to have a baseline for our slightly altered dialog behavior, we added a third bonus level, which does not include any additional payments or bonus counter.

Participants were tasked with dismissing as many dialogs as possible. The introductory text also explained the bonus mechanic and the risk of losing accumulated money as a result of a wrong decision (see also Appendix A). As seen in Figures 1c and 2, the dialogs were modeled after Windows pop-up dialogs and appear at different, randomized positions within the browser window. In the top section of the window, a status bar with the currently accumulated bonus money and the total time remaining for the task was displayed (cf. Figure 2). If this timer reached zero, nothing happened (as in the original work), but there was a hidden time limit of 30 seconds per dialog that caused an alert after 15 seconds of inactivity. This time limit was the same as in the Bravo-Lillo study. If a participant triggered this time limit, they were dismissed from the study. As stated above, we removed the time limit for the first dialog in order to give the participants enough time to carefully read the dialog.

In contrast to Bravo-Lillo et al.'s work, we chose to imitate the behavior of real-world security dialogs more closely. We designed the security dialogs as gatekeepers to some actual content, which was in our case a message window with an inspirational quote. The dialog's status field, which in the case of previous work usually displayed the number of already dismissed pop-ups, contained either "This message was signed by: University of Bonn" or "This message was signed by: Unknown", which is closer to the Windows UAC dialogs the original work by Bravo-Lillo et al. intended to mimic [14], [15], [35]. The answer options were changed to "Yes, please show me" and "No, do not show me". If a participant chose the "no" option, instead of the quote, a similar window with the message "You chose not to view the message" appeared. Although the dialog itself did not instruct the participants which messages they should accept, the introductory task briefing clearly stated that only the messages signed by University of Bonn should be accepted and others dismissed (cf. Appendix A).

A comparison between the dialog design of the previous work and our altered version is depicted in Figure 1.

During the habituation period, a series of either one, three, or twenty dialogs that encouraged the "yes" option were shown. Where previous work disabled the "no" answer option during the habituation period, we chose to leave both answer options open to make the setting more realistic. We chose the bonus increase of 2.5 cents per dialog such that the groups with long habituation periods had the chance to accumulate the maximum bonus of $0.50\$$ by the end of the habituation phase. The condition of earning 10 cents bonus per correct dialog was added to test if behavior differed in the "one dialog" and "three dialogs" habituation groups. Afterward, the *test phase* started.

In accordance with the study by Bravo-Lillo et al., the first dialog of the test phase had to be answered with "no" in all participant groups. In previous work, participants could finish the study at this point if they chose the right answer. If they failed to do so, the same dialog was displayed again until the participant clicked the "no" answer or interacted with the study for five more minutes [15].

In our extended study, the test period consisted of a fixed amount of 41 dialogs. Additional dialogs that were to be answered with "no" were inserted in the test period depending on the participant's habituation condition, in order to keep the probability of a true warning roughly in line with the frequency in the habituation phase (see also Table I). For a habituation phase of one dialog, a "no" dialog appeared in the test phase with a probability of 0.5. For a habituation phase of three dialogs, a "no" dialog appeared with a probability of 0.25. We opted for random instead of round robin assignment of the warnings, as participants could otherwise have easily deduced a pattern in the one and three exposure scenarios. For a habituation phase of twenty dialogs, additional "no" dialogs were inserted in position 21 and 41 of the habituation period. Please note again that the first dialog in the test phase was to be answered with "no" in all conditions.

A summary of the differences between Bravo-Lillo et al.'s study and our study can be found in Table I.

We emulate a greater risk for the participants, which involved the accumulation of a monetary bonus whenever a correctly signed message was accepted and the complete

Feature	Original Study	Our Replication
Dialog design	Term “pop up windows”, focus on dismissing quickly	Gatekeeping scenario for signed content
Status Message	Contains number of dismissed dialogs or finish early instruction	Contains signing information for following content
“No” answer option	Deactivated during habituation phase	Always enabled
Attractors	Control, ANSI, Type, Swipe, ACD	Control, Swipe
Habituation Phase	1, 3, 20 dialogs, 150 s	1, 3, 20 dialogs
Test Phase Lengths	Between 1 dialog and 300 s	41 dialogs
Test Phase Length	Depends on compliance	Fixed
Additional dialogs to answer with “No”	None	Inserted depending on habituation condition
Compliance reward	Finishing the study early	Additional payment of up to 0.50$

TABLE I: Comparison of the original study and our replication.

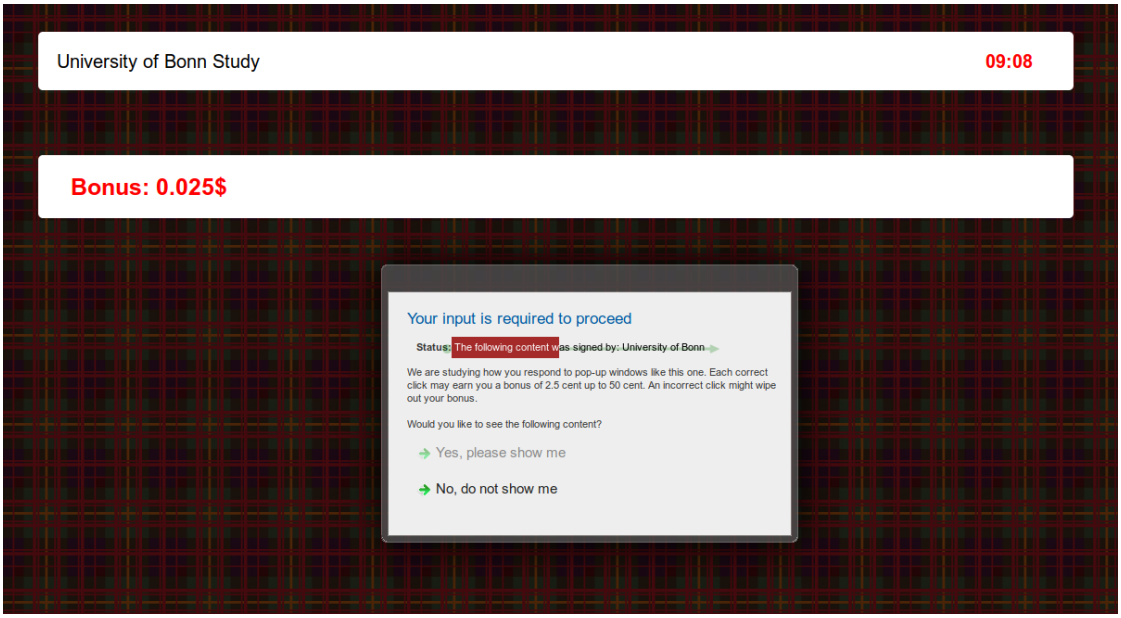


Fig. 2: Screenshot of the study interface. The timer in the top right corner shows the remaining time for the whole study task, an inactivity warning would appear next to it. Currently accumulated bonus is always displayed in the top left part of the screen. The dialog shown uses the *swipe* attractor. Note that the No-Option is always enabled and choosing it is not penalized.

loss of said bonus if a message signed by “Unknown” was accepted. A correct answer gave the participant either 2.5 or 10 cents, depending on their assigned conditions. This bonus was accumulated but did not increase further after reaching a total bonus amount of $0.50. If a “malicious” message was accepted by a click on “Yes, please show me”, the entire bonus was set to zero. Note that, while in the real world the inability to finish a task because of a security risk would probably perceived as a loss, in our study, a click on “No, do not show me” never comes with any consequences, since declining a display or installation is always safe. A click on the “no” answer was always possible in our study and did not require compliance with the attractor. The longer test period, in comparison to previous work, made it possible to finish the study with the maximum bonus, despite having made a mistake in the first dialog of the test period.

E. Procedure

The task was listed on Amazon MTurk with the same description and properties as in the original study, the only difference being the compensation. Where Bravo-Lillo et al. offered their participants a payment of $1 upfront, we needed to disguise the guaranteed payout of the bonus; therefore, the task was listed with a compensation of $0.50. This is a limitation of the MTurk platform. Despite telling the participants that they would receive bonus payment dependant on their performance (cf. Appendix A), all participants who successfully completed our study task were offered another task with a compensation of $0.50 in which they only needed to confirm the collection of their bonus. That is, regardless of their performance, all participants received this bonus task.

After a participant had successfully finished the task, they were redirected to a post-study survey asking about their attention during the task and their perception of the attractor they experienced. Survey questions were replicated from Bravo-Lillo et al.’s work and slightly adapted to our altered study design. The contents of the post-study survey can be found in Appendix B.

F. Ethical Considerations

As researchers utilizing crowdwork for scientific studies, we have the collective obligation to treat the workers in a best possible way, which not only encompasses protecting their privacy, but also paying them an adequate wage for their work. When replicating previous MTurk studies, the compensation is an important steering factor for attracting a study population. We therefore chose to keep the original compensation of $1 for all workers. This was not an easy decision, but in the light of

good scientific practice, we regarded the replication importance in this case as higher.

While our institution does not have an IRB, we took the following steps to ensure that participants were treated ethically: We adhered the IRB-reviewed protocol by Bravo-Lillo et al. as closely as possible, making only the changes detailed in Section III. We stored participant data in password-protected encrypted cloud storage. We anonymized all study data immediately after collection, using MTurk identifiers only to pay the participants, then discarding them.

Participants consented to the use of their study data in a consent form, and could withdraw participation at any time during the study. We used deception when participants collected the bonus: After the study, we debriefed them and paid out to full bonus to all participants, to make sure that participants received equal payment for equal work. To ensure that participants did not share this information, we only paid the bonus after the study had been completed by all, in a separate task. After experiencing the very low rate of workers who collected their bonus money in a separate task (see Section III-G), we would not recommend this procedure for future work; instead we recommend paying the bonus directly.

G. Participants

We invited 1,800 MTurk members to participate in our study. We required the same attributes as in the original study; we asked for their location to be the United States with a HIT (task) approval rate of at least 95%. We paid a base rate of $0.50 regardless of performance. For our analysis, a total of 564 participants were removed from the set, 504 who failed to answer all dialogs because of timeouts and 60 who answered "No" on every dialog, as they had not adhered to the study instructions. We retained 1,236 valid participants, for whom we report results. The average study participation time was 358 seconds (sd=113 seconds). All participants had access to the bonus collection task, where they could get another $0.50 regardless of their bonus condition and their performance in the main experiment. The task was only visible to workers who had participated in the main experiment (realized with a special MTurk qualification) and listed for the duration of two weeks. While we thought this was a good idea, a low return rate of 12% showed that a different bonus distribution method might be more suitable for this kind of experiment.

Our participants were predominantly female (58%), and their mean age was 35 years (sd=9.62). They were mostly White/Caucasian (78%), and 90% had a college-level or higher education. This fits well within the usual MTurk population as described by Buhrmester et al. [17], which is generally skewed to be more female, slightly older, and more educated than the general US public.

We report replication results and new results separately.

IV. Results

A. Replication

Firstly, we replicated Bravo-Lillo et. al's study, focusing on only two attractor conditions. We replicated the control as well as the swipe attractors, using one, three, or twenty exposures, and we measured response times to the final habituation dialog as well as compliance on first click.

Unfortunately, we were not able to get hold of the data used in the experiment by Bravo-Lillo et al. that we use as the baseline for our extended study; only the data reported in their paper was available to us. To assess how closely we were able to replicate their experiment, we firstly investigated whether participant response times to their final habituation dialog were comparable. We present $1 - 0.1$ confidence intervals of the data we obtained in the replication part of our experiment and determine whether the median data points reported by Bravo-Lillo et al. lie within these confidence intervals.

Habit.	Attr.	Median	C.I.	No×Yes
1 exp.	C	9.34 (10)	[8.42, 10.57]	28×71 (50×56)
	S	32.38 (39)	[29.92, 40.63]	51×42 (61×45)
3 exp.	C	3.04 (3.4)	[2.43, 3.67]	23×81 (43×64)
	S	7.56 (6.9)	[6.86, 8.65]	41×52 (65×48)
20 exp.	C	2.04 (1.2)	[1.96, 2.12]	14×80 (24×90)
	S	4.73 (3.9)	[4.47, 4.96]	32×55 (59×48)

TABLE II: Replication: Comparison of different habituation levels ("Habit.") and Attractors ("Attr.", "C" for control, "S" for swipe) for median response times to final habituation dialog (labeled "Median"), 90% confidence interval for the median ("C.I."), number of respondents who chose "No" (i.e. who complied) vs. number of respondents who chose "Yes"; we report the results by Bravo-Lillo et al. in parentheses.

As visible in Figure 3, we come close, but we are not able to replicate the exact times. As reported in Table II, only the median times for one and three exposures from Bravo-Lillo's study, both for the control group and the group that used the swipe attractor, are within the $1 - 0.1$ confidence intervals of our own measurements. The medians for twenty exposures are somewhat further away from our own measurements.

To assess compliance on first click in our study as compared to compliance in Bravo-Lillos et al.'s study, we report compliance and non-compliance counts in Table II; clicking "no" in this case would have been compliant. Our control group is less compliant than in the original study across all exposure conditions; except for the twenty-exposure group, the difference is statistically significant (Chi-Square-test, $X = \{7.75, 8.01, 1.31\}, p = \{0.005*, 0.005*, 0.253\}$; * indicates significance at p=0.05, Bonferroni-Holm corrected for multiple testing). The swipe groups are also less compliant; however, this difference is only significant for twenty exposures (Chi-Square-test, $X = \{.15, 3.69, 6.49\}, p = \{0.7, 0.05, 0.01*\}$; * indicates significance at p=0.05, Bon-Ferroni-Holm corrected for multiple testing).

Bravo-Lillo et al. calculated odds ratios and used them to calculate likelihood ratios. Even though we used their original numbers as input, we were not able to replicate their test with the same result. We, therefore, omit reporting a similar value for our replication study.

Generally, while the timings in our sample differed from the original study, we were able to replicate the overall effect that the swipe attractor performs significantly better than the control; also, the effect that different exposures habituate differently holds.

B. Main Study

In the following subsections, we apply statistical inference to analyze our results in detail.

To analyze the effect of the various conditions on compliance, we performed the following linear regression. Firstly, we transformed the series of compliant and non-compliant choices

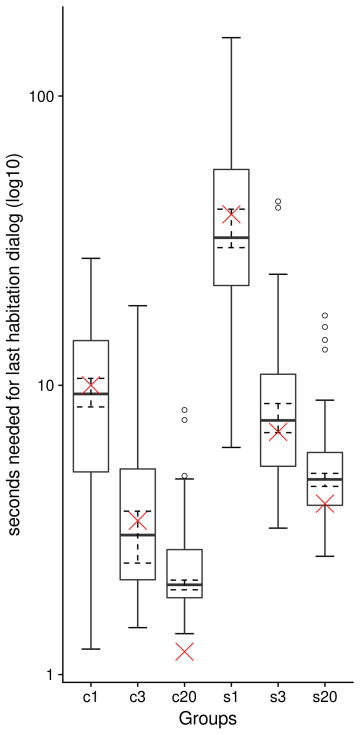


Fig. 3: Time spent on last habituation dialogue by condition. Red crosses indicate reported median times from Bravo-Lillo et al. Dashed error bars indicate quantiles with a 90% confidence interval.

made by each participant into the fraction of non-compliant choices divided by all possible choices, which is our dependent variable. As we measured this on a per-participant basis, not on a per-click basis, we do not have to use a mixed model or random intercepts.

For the regression analysis, we considered a set of candidate models and selected the model with the lowest Akaike Information Criterion (AIC) [18]. The included factors were the three levels of bonuses and the three levels of habituation as well as the two possible attractors. We considered all possible combinations of all interactions as optional variables and all possible combinations of the optional variables. We report the factors in Table III. The outcome of our regression is reported in Table IV. Each row measures change in the analyzed compliance outcome related to change from the *baseline* value for a given factor to a different value for that factor. Linear regressions measure change in the absolute value of the outcome; baseline factors by construction have coef=0. In each row, we also provide a 95% confidence interval and a p-value indicating statistical significance.

For the regression, we set the control, combined with the shortest training time as well as the control attractor, as the baseline. All baseline values are given in Table III.

C. What influences compliance?

1) Habituation: Both the habituation with three dialogs and the habituation with twenty dialogs are responsible for significantly increasing the ratio of compliant clicks compared to the baseline with no habituation dialogs. The habituation group with twenty dialogs shows a larger increase than the three dialogs group. This can likely be explained by participants in these groups getting accustomed to quickly accepting dialog choices.

Factor	Description	Baseline
Habituation	Number of dialogs	1 dialog
Attractor	Attention mechanism	Control
Bonus	Monetary reward component	No bonus

TABLE III: Required factors and their baseline values used in the linear regression models. Categorical factors are individually compared to the baseline. Final models were selected by minimum AIC; candidates are defined using all possible combinations of any two required factors; all required factors are included in every candidate.

Factor	Estimate	C.I.	p-value
Habit. 3 dialogs	0.03	[-0.04, 0.1]	0.424
Habit. 20 dialogs	0.16	[0.09, 0.23]	<0.001*
Bonus $0.025	-0.12	[-0.18, -0.07]	<0.001*
Bonus $0.10	-0.14	[-0.2, -0.09]	<0.001*
Swipe Attractor	-0.03	[-0.1, 0.04]	0.42
Habit. 3 : Swipe	-0.01	[-0.11, 0.1]	0.889
Habit. 20 : Swipe	-0.15	[-0.26, -0.03]	0.017*

TABLE IV: Results of the final linear regression model examining the ratio of non-compliant clicks (payout lost for groups with bonus) to all possible choices for participants. Statistically significant factors indicated with *. See Table III for details.

2) Bonus: Our regression shows that the level of bonus matters: compared to the baseline of not paying out any bonus, paying out a small bonus of 2.5 or 10 cents per correct click lowered the non-compliance ratio significantly.

In addition to performing these tests, we took participants' self-reported data into account. In the exit survey, we asked participants in a bonus condition if they paid more attention because of the monetary incentive. Of all participants, 359 replied with a strong "Yes, very", 322 with a "Yes, a little", and 103 with "No". We compare the amounts of total bonus payout at the end of the study across these three groups, finding significant differences in the distribution of values between the survey answers (Kruskal-Wallis: $\chi^2 = 13.33$, $p = 0.002$). However, the median bonus payout for "No" was 50 cents ($\mu = 38.54$), 50 cents ($\mu = 37.31$) for "Yes, a little", and 50 cents ($\mu = 42.56$) for "Yes, very"; the means differed slightly, with the mean payout for "Yes, very" being the highest.

3) Attractor: The swipe attractor was responsible for a significant decrease in non-adherence over the control. This result was found in Bravo-Lillo's first contact study as well, and it was stable in our study too.

4) Does the extent of first loss impact compliance?: We hypothesize that a higher amount of accumulated bonus at the point of the first loss increases subsequent compliance. To test this assumption, on the set of participants that had the chance to obtain a bonus payment, we correlate the amount a participant had accumulated at the time of their first loss with the number of non-losses after the first loss using Kendall's τ. We find that $\tau = 0.13$ with $p = 0.0005$. This means that there is a slight, significant positive correlation between losing more money and adhering to the warnings better in the future. While significant, this correlation is small. However, we find this promising since the intuitive interpretation that risk perception rises after an incident seems to hold even for small losses.

On the same set of bonus participants, we also correlate

Habituation	Bonus		
	$0	$0.025	$0.1
1 dialog	4.5 (2)	3 (2)	16 (2)
3 dialogs	32 (4)	23 (4)	36 (4)
20 dialogs	21 (21)	- (21)	- (21)

TABLE V: Median click index for first loss. "-" indicates that the median is "no loss". Numbers in parentheses are medians of first loss, excluding all participants who did not encounter a loss at all.

the extent of the first loss with the accumulated bonus payout at the end of the study. We find the correlation between the accumulated bonus at first loss and a payout in the end to be non-significant (Kendall's $\tau = 0.05$, $p = 0.19$). To test whether the amount of non-loss clicks until the first loss is different across all three bonus conditions, we use the Kruskal-Wallis test. We find that the difference in the distribution of values among the bonus conditions is highly significant ($\chi^2 = 18.95$, $p < 0.001$).

Median click indices at first loss are reported in Table V. When looking at participants who encounter a loss, the vast majority encounter the first loss at the first possible decision time, or directly after the habituation phase ends. Across all participants, however, in the bonus conditions for the longest habituation phase, the majority of participants do not lose any money. On the one hand, these participants are habituated the most strongly; on the other hand, they encounter the least risk-bearing warnings. Generally and surprisingly, participants' attention seems to be better with no bonus than with a low bonus. Thus, a high bonus seems to work better in keeping participants attentive than either of the other conditions.

V. Limitations

As with any user study, our results should be interpreted in context. We drew participants from MTurk, which is a common source of participants for social sciences, human-computer interaction research, and usable security and privacy research [33]. Studies have shown that participants from the MTurk population are more diverse than those drawn from traditional university participant pools [8] and produce results similar to those recruited from other sources, including nationally representative samples [37]. Additionally, the purpose of our study was to replicate and extend the work by Bravo-Lillo et al. [14], [15], who also used MTurk to recruit participants.

To facilitate the bonus payment while keeping the total payment for our extended study the same as in Bravo-Lillo's study, we had to initially advertise the study as only paying $0.50 and only in the description would participants learn that they could actually earn up to $1. This limitation means that we cannot draw direct comparisons between our replication study and our extended study. The other option of advertising the extended study with a payout of $1 and then adding the bonus could also have effected recruitment. We therefore chose to keep the total value of the study constant. For participants who read the study instructions, this was closest to the original study by Bravo-Lillo et al.

By today's standards, the original study is not free from criticism and this also affects its replication. First and foremost, the study is missing a primary task in which the warnings are embedded. We didn't want to move away too far from Bravo-Lillo et al.'s original design, but tried to mitigate this a little by adding content messages in between warnings. The study is still not to be considered realistic, and thus, the results are not directly applicable to real world scenarios.

The bonus system itself can also not be mapped on real-world data loss directly. Since for ethical reasons we could not simulate data loss in a realistic fashion, we had to rely on a symbolic proxy for data loss. Therefore, we chose a capped, incremental bonus for complying with the study instructions and a complete loss of the accumulated bonus in case of a wrong decision. Of course, the kind of loss does not match the loss of data after a malware infection, since factors like time between the loss and subsequent behaviour cannot be adequately addressed in a study context [34].

Our study focuses on habituation and study design effects. It does not address how false-positive warnings can be reduced themselves. While false positives are a concern in the wild and have been proven to greatly shape user trust in and behaviour towards warnings [38], we consider them as safe answer choices in our study since we didn't want to increase the complexity any further. This again was a compromise which we solved towards sticking closely to the original study. Adding a primary task could solve this issue in future work.

Showing participants a high number of security dialogs does not directly model habituation, which usually occurs over a longer period of time. However, this is currently the only known study setup for researching habituation without conducting a long-term field study. In addition, we wanted our study design to be comparable to Bravo-Lillo et al.'s; therefore we chose the same frequency of security dialogs during the study. Nonetheless, the experimental environment remains artificial and should only be used for initial exploration of concepts. We suggest long-term studies on the habituation effect, preferably as an in vivo study on the participants' private devices, to study the habituation effect of warnings in greater detail.

When comparing our replication results to the original study, we were able to only use aggregated data from the paper, as the authors no longer had access to their raw data.

Finally, while we showed that monetary incentives have a significant effect on warning message adherence, we cannot make any claim on how this translates to other real-world risks. One could consider this as a general mismatch between the incentives to participants in this study with the incentives experienced by real users. As stated above, field studies are required to examine this relationship in the future.

VI. Discussion and Future Work

A. Result Replication

We were not able to fully replicate the results of the original study by Bravo-Lillo et al. For high habituation levels with twenty dialogs, we failed to reproduce the median response time. Compared to the original study, our participants needed more time. Regarding the swipe attractor, this also indicates that swipe might not be as fast as originally assumed. Bravo-Lillo et al. argue that when people get used to its mechanic, the time overhead needed to use the attractor diminishes, but our results indicate that the offset might remain rather high. This means that the swipe attractor, although showing good and replicable results, needs further investigation in a longer study in order to thoroughly investigate its mental and time-related load for users. Additionally, the compliance rates could

not be replicated consistently. We observed less compliance in the groups without an attractor compared to Bravo-Lillo et al., except for the twenty-dialog condition. For the swipe attractor, the compliance rates for low-habituation groups (one and three dialogs) could be replicated.

These inconsistent results need to be examined in future replication studies, since they bring to question the viability of crowd-working studies for security tasks in general. However, when looking at a specific attractor, the replication results were similar. This might be the result of receiving more diverse sample groups than we first thought, indicating that there may be no such thing as a "static" MTurk population. This could be due to worker fluctuation, changes in payment standards (we paid the same as the original study from 2013 but did not correct for inflation), or more experienced workers in regard to scientific studies, since MTurk is an established recruiting tool for quantitative research.

B. Monetary Incentives

A major contribution of this paper is the investigation of monetary incentives to emulate risks in online studies. We showed significant effects of monetary incentives on security dialog compliance, which indicates that monetary incentives can have a notable impact on user behavior and, thus, measured performance. We noted remarkably different behavior between the groups who had monetary risks compared to our control, who had to imagine the risk of ignoring a warning without any real consequences, as is the case in virtually all IT warning studies to date.

While this is an important finding which we hope will set future directions of warning studies, our study only represents the first step. While we have shown that our participants perceived an actual risk because of the implemented monetary bonus system, we do not yet have any information on how this relates to more realistic online risks. Previous work on warning studies has shown that often users do not fully understand the risks they subject themselves to when ignoring warnings of different types. For instance, some users believe that because they are running MacOS or Linux that they are not in danger of a man-in-the-middle attack [38].

Using monetary incentives to model risk in studies thus gives us a great opportunity to introduce risk perception by conducting studies with and without such incentives and to compare participant behavior to real-world behavior. The calibration of the amounts is, however, an open research issue. To study the relation between perceived risk of monetary loss and risk of phishing, malware, or data loss, further studies in both qualitative and quantitative domains are needed, e.g., mental model studies about data loss in a cloud-dominated internet or large-scale studies for fine-tuning the risk perception between various kinds of virtual threats and corresponding monetary loss.

C. Risk Modeling

By modeling risks, we can address previously unexplored aspects of the habituation study design, like behavior after a bad experience, and gain novel insights into risk modeling as a part of experimental methodologies. We also find that the monetary incentive do not prevent loss at all, except under long habituation conditions (cf. Table V), but it significantly shapes subsequent behavior.

While previous research has shown that monetary incentives are not fit to model concrete real-world losses [20], [26], the introduction of the monetary component has nonetheless improved the study design regarding the habituation effect. While we show that monetary incentives improve the study methodology, many open questions about risk modeling remain. Since working with actual malware infections in user studies is highly unethical, the monetary replacement might be the best option to emulate said risk. Thus, future work in this domain is needed in order to improve study methodology.

VII. Conclusion

In this work, we conducted and evaluated an extension of a study conducted by Bravo-Lillo et al. on the habituation effect in the context of system security dialogs. Though previous work by Bravo-Lillo et al. gave important insights, we created a more realistic experimental design by adding the risk of monetary loss in order to model the risk of a malware infection that could result from making a wrong decision on a system security pop-up.

There are two important takeaways from our study. Firstly, we were not able to fully replicate the results of Bravo-Lillo et al. This highlights the importance of replication studies, which are still underrepresented in the field of usable security and privacy research. Further replication studies are needed to understand the reasons for the differences found in our work compared to previous work. One possible factor could be the MTurk population not being robust against resampling over the timespan of several years. We think this is particularly important to study, since MTurk is a very popular platform for conducting such studies, and we rely on results that have usually not been replicated or confirmed.

The second important takeaway is that monetary incentives look like an effective and promising approach for modeling risk in warning studies. Our results show that the monetary component influences user compliance significantly, and that a monetary loss influences subsequent compliant behavior. This has both the potential to improve future warning studies as well as to provide a potential measure against which to compare perceived risk in different situations.

While this work already extends current methodology on habituation studies, future work regarding the exact modeling of monetary incentives for security risks is still needed. Furthermore, field studies will be required to compare the habituation effects modeled by Bravo-Lillo et al. and us with habituation effects in the wild. This extremely challenging task would be of significant benefit to the community.

Acknowledgments

We would like to thank Christian Bravo-Lillo for supplying us with the original study material. We would like to thank all authors of the original paper for their help and assistance in recreating the study conditions. Big thanks go to Yasemin Acar, who helped with guidance and feedback throughout the whole process and gave invaluable support. Many thanks to our shepherd Lujo Bauer for the detailed and supportive feedback. Also, thanks to rehwanne for providing us with infrastructure when our own was not available.

REFERENCES

[1] M. E. Acer, E. Stark, A. Porter Felt, S. Fahl, R. Bhargava, B. Dev, M. Braithwaite, R. Sleevi, and P. Tabriz. Where the Wild Warnings Are: Root Causes of Chrome HTTPS Certificate Errors. In *ACM CCS*, 2017.

[2] D. Akhawe and A. P. Felt. Alice in warningland: A large-scale field study of browser security warning effectiveness. In *USENIX security symposium*, volume 13, 2013.

[3] B. B. Anderson, C. B. Kirwan, J. L. Jenkins, D. Eargle, S. Howard, and A. Vance. How polymorphic warnings reduce habituation in the brain: Insights from an fmri study. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2883–2892. ACM, 2015.

[4] Android Open Source Project. Working with System Permissions, 2017.

[5] Apple, Inc. OS X: About Gatekeper, 2016.

[6] J. Aronfreed and A. Reber. Internalized behavioral suppression and the timing of social punishment. *Journal of Personality and Social Psychology*, 1:3–16, 1965.

[7] G. S. Bahr and R. A. Ford. How and why pop-ups don't work: Pop-up prompted eye movements, user affect and decision making. *Computers in Human Behavior*, 27(2):776 – 783, 2011.

[8] T. S. Behrend, D. J. Sharek, A. W. Meade, and E. N. Wiebe. The viability of crowdsourcing for survey research. *Behavior Research Methods*, 43(3):800, Mar 2011.

[9] N. Ben-Asher, J. Meyer, Y. Parmet, S. Moeller, and R. Englert. An experimental microworld for evaluating the tradeoffs between usability and security. In *Symposium on Usable Privacy and Security (SOUPS)*, 2010.

[10] A. R. Beresford, D. Kübler, and S. Preibusch. Unwillingness to pay for privacy: A field experiment. *Economics Letters*, 117(1):25–27, 2012.

[11] R. Böhme and J. Grossklags. The security cost of cheap user interaction. In *Proceedings of the 2011 New Security Paradigms Workshop*, NSPW '11, pages 67–82, New York, NY, USA, 2011. ACM.

[12] R. Böhme and S. Köpsell. Trained to accept?: a field experiment on consent dialogs. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2403–2406. ACM, 2010.

[13] M. E. Bouton. *Learning and Behavior: A Contemporary Synthesis*. Sinauer, 2007.

[14] C. Bravo-Lillo, L. Cranor, S. Komanduri, S. Schechter, and M. Sleeper. Harder to ignore? revisiting pop-up fatigue and approaches to prevent it. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*, pages 105–111, 2014.

[15] C. Bravo-Lillo, S. Komanduri, L. F. Cranor, R. W. Reeder, M. Sleeper, J. Downs, and S. Schechter. Your attention please: designing security-decision uis to make genuine risks harder to ignore. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, page 6. ACM, 2013.

[16] J. C. Brustoloni and R. Villamarín-Salomón. Improving security decisions with polymorphic and audited dialogs. In *Proceedings of the 3rd symposium on Usable privacy and security*, pages 76–85. ACM, 2007.

[17] M. Buhrmester, T. Kwang, and S. D. Gosling. Amazon's mechanical turk. *Perspectives on Psychological Science*, 6(1):3–5, 2011. PMID: 26162106.

[18] K. P. Burnham. Multimodel Inference: Understanding AIC and BIC in Model Selection. *smr*, 33(2):261–304, 2004.

[19] D. Cvrcek, M. Kumpost, V. Matyas, and G. Danezis. A study on the value of location privacy. In *Proceedings of the 5th ACM Workshop on Privacy in Electronic Society*, WPES '06, pages 109–118, New York, NY, USA, 2006. ACM.

[20] G. Danezis, S. Lewis, and R. Anderson. How much is location privacy worth. In *In Proceedings of the Workshop on the Economics of Information Security Series (WEIS*, 2005.

[21] C. Della Libera and L. Chelazzi. Visual selective attention and the effects of monetary rewards. *Psychological science*, 17(3):222–227, 2006.

[22] D. Di Cagno, A. Galliera, W. Güth, F. Marzo, and N. Pace. (Sub) Optimality and (non) optimal satisficing in risky decision experiments. *Theory and Decision*, 83(2):195–243, 2017.

[23] S. Egelman, L. F. Cranor, and J. Hong. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1065–1074. ACM, 2008.

[24] A. P. Felt, A. Ainslie, R. W. Reeder, S. Consolvo, S. Thyagaraja, A. Bettes, H. Harris, and J. Grimes. Improving SSL Warnings: Comprehension and Adherence. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI'15)*, pages 2893–2902, 2015.

[25] W. J. Gehring and A. R. Willoughby. The medial frontal cortex and the rapid processing of monetary gains and losses. *Science*, 295(5563):2279–2282, 2002.

[26] J. Grossklags and A. Acquisti. When 25 cents is too much: An experiment on willingness-to-sell and willingness-to-protect personal information. In *WEIS*, 2007.

[27] J. D. Harris. Habituatory response decrement in the intact organism. *Psychological Bulletin*, 40:285–422, 1943.

[28] G. Humphrey. Extinction and negative adaptation. *Psychological Review*, 37:361–363, 1930.

[29] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, pages 263–291, 1979.

[30] G. Keren and P. Roelofsma. Immediacy and certainty in intertemporal choice. *Organizational Behavior and Human Decision Processes*, 63(3):287 – 297, 1995.

[31] S. Kim and M. S. Wogalter. Habituation, Dishabituation, and Recovery Effects in Visual Warnings. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 53(20):1612–1616, 2009.

[32] K. Krol, M. Moroz, and M. A. Sasse. Don't Work. Can't Work? Why It's Time to Rethink Security Warnings. *Risk and security of internet and systems (CRiSIS), 2012 7th International conference*, pages 1–8, 2012.

[33] K. Krol, M. Moroz, and M. A. Sasse. Don't work. can't work? why it's time to rethink security warnings. In *2012 7th International Conference on Risks and Security of Internet and Systems (CRiSIS)*, pages 1–8, Oct 2012.

[34] A. W. Kruglanski and E. T. Higgins. *Social psychology: Handbook of basic principles*. Guilford Publications, 2013.

[35] Microsoft and B. Lich. Windows IT Center: User Account Control, 2017.

[36] D. Prelec and G. Loewenstein. Decision making over time and under uncertainty: A common approach. *Manage. Sci.*, 37(7):770–786, July 1991.

[37] D. J. Simons and C. F. Chabris. Common (mis)beliefs about memory: A replication and comparison of telephone and Mechanical Turk survey methods". *PLOS ONE*, 7(12), 2012.

[38] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F. Cranor. Crying wolf: An empirical study of ssl warning effectiveness. In *USENIX security symposium*, pages 399–416, 2009.

[39] R. F. Thompson and W. a. Spencer. Habituation: a model phenomenon for the study of neuronal substrates of behavior. *Psychological review*, 73(1):16–43, 1966.

[40] A. Vance, B. Kirwan, D. Bjornn, J. Jenkins, and B. B. Anderson. What do we really know about how habituation to warnings occurs over time?: A longitudinal fmri study of habituation and polymorphic warnings. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2215–2227. ACM, 2017.

[41] R. H. Walters and L. Demkow. Timing of punishment as a determinant of response inhibition. *Child Development*, 34(1):207–214, 1963.

[42] R. West. The psychology of security. *Commun. ACM*, 51(4):34–40, Apr. 2008.

APPENDIX

A. Replication Study

University of Bonn Habituation Study

In the following page you will see a timer on the screen, and a number of consecutive dialogs (pop-up windows) asking

you to click 'Yes' or 'No'. Your task is to respond to as many dialogs as you can before the timer goes off. You can increase your performance by following instructions and responding to each question quickly. Some dialogs may require you to wait or perform an action before the 'Yes' button is activated.

Those who perform well may be rewarded with opportunities to finish the study early while still receiving their full payment. After finishing the task, you will have to answer a short survey.

When you are ready to begin, please click on the URL below.

B. Extended Study with Bonus

University of Bonn Habituation Study

In the following page you will see a timer on the screen, and a number of consecutive dialogs (pop-up windows) asking you to click 'Yes' or 'No'. Please accept dialogs signed by University of Bonn and refuse others. Your task is to respond to as many dialogs as you can before the timer goes off. You can increase your performance by following instructions and responding to each question quickly. Some dialogs may require you to wait or perform an action before the 'Yes' button is activated.

For every correctly accepted dialog, you will receive a small bonus which adds up to a maximum of 50 cent. An incorrect click may wipe out your entire bonus. Those who perform well will be rewarded with an invitation to a separate MTurk task for collecting the accumulated bonus. After finishing the task, you will have to answer a short survey.

When you are ready to begin, please click on the URL below.

For our new study design, the post-study survey consisted of the following questions, which are in exact replication of Bravo-Lillo et al.'s work [14]. For the test groups without a bonus system, the third answer option in Question 2 as well as Questions 9, 10, and 11 were omitted.

After answering these questions, the participants received their MTurk qualification code in order to finish the task.

1. The image below corresponds to one of the dialogs you saw during this study. Please type in the contents of the Status: field in the most recently shown dialog, to the best of your memory. If you have no memory, please type "none": _____ [Screenshot]

2. What did the last status message you saw communicate?

- ☐ That I should press "yes" to view the message
- ☐ That I could press "no" to dismiss the message
- ☐ [The amount of bonus money I would get for the next click]
- ☐ The amount of money I will be paid for this study
- ☐ The quality of my performance
- ☐ I'm not sure

3. How many times did you see this message?

- ☐ Just once
- ☐ Between 1 and 10
- ☐ Between 10 and 20
- ☐ Between 20 and 50
- ☐ 50 or more
- ☐ I don't know

4. Was it easy to detect the correct answers?

- ☐ Yes
- ☐ No

5. If your messages were highlighted, please describe how useful you thought they are to emphasize the content of pop ups.

_____

6. Overall, how annoying was this task?

Not annoying at all ☐ ☐ ☐ ☐ ☐ Very annoying

7. Did you suspect that the study may require you to answer questions about the content of the status field?

- ☐ Definitely
- ☐ Somewhat
- ☐ Maybe a little
- ☐ Definitely not

8. During most of the dialogs you saw, did you intentionally read the text in the field labeled "Status"?

- ☐ I ignored it
- ☐ I tried to read it a little
- ☐ I read every word

9. Did the possibility to accumulate a monetary bonus motivate you to read the dialogs more carefully?

- ☐ Yes, it made me read them very carefully
- ☐ Yes, but only a little
- ☐ No, it didn't make me read them more carefully

10. Did you lose your accumulated bonus at least once during the task?

- ☐ Yes
- ☐ No
- ☐ I'm not sure

11. If you answered "yes" to the previous question: Did this influence your behaviour?

- ☐ Yes, I was much more careful afterwards
- ☐ Yes, I was a little more careful afterwards
- ☐ No
- ☐ I'm not sure

12. Please let us know what, if anything, was not working with the dialogs that popped up on your browser?

_____

13. Do you know any programming language?

- ☐ Yes
- ☐ No

14. If you chose "Yes" in the last question: Which programming languages do you know?

_____

15. What is your gender?

☐ Female
☐ Male
☐ None of the above
☐ Decline to answer

16. What is your age?
[dropdown]

17. What is your race/ethnicity?

☐ Asian/Pacific Islander
☐ Black/African-American
☐ White/Caucasian
☐ Hispanic
☐ Native American/Alaska Native
☐ Other/Multi-Racial
☐ Decline to answer

18. What is your current occupation?

☐ Adminsitrative Support (eg., secretary, assistant)
☐ Art, Writing and Journalism (eg, author, reporter, sculptor)
☐ Business, Management and Financial (eg, manager, accountant, banker)
☐ Education (eg, teacher, professor)
☐ Legal (eg, lawyer, law clerk)
☐ Medical (eg, doctor, nurse, dentist)
☐ Science, Engineering and IT professional (eg., researcher, programmer, IT consultant)
☐ Service (eg., retail clerks, server)
☐ Skilled Labor (eg., electrician, plumber, carpenter)
☐ Student
☐ Other Professional
☐ Not Currently Working/Currently Unemployed
☐ Retired
☐ Other
☐ Decline to answer

19. If you chose "Other" in the last question: What is your current occupation?

_____

20. What is the highest level of education you have completed?

☐ Some high school
☐ High school/GED
☐ Some college
☐ Associate's degree
☐ Bachelor's degree
☐ Master's degree
☐ Doctorate degree
☐ Law degree
☐ Medical degree
☐ Trade or other technical school degree
☐ Decline to answer