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

Sascha Fahl
Ruhr-University Bochum
sascha.fahl@rub.de

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

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

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

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

Copyright is held by the author/owner. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee. Poster presented at the 14th Symposium on Usable Privacy and Security (SOUPS 2018).

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 as a factor for risk of loss in order to refine the study methodology on habituation research in that direction. 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.

Introduction

Dialog windows are part of operating systems' as well as application software security measures such as the User Account Control (UAC) mechanism in Windows 10 [22], Android permission dialogs [2], or browser warnings about insecure TLS connections or malware-infested websites.

While security dialogs are essential to overall information security, users tend to perceive them as rather annoying and ignore them by clicking-through, even if risks are present [4, 7, 20]. It is therefore an important goal for usable security research to design security dialogs which prevent such habituation effects which occur from frequently "clicking through" security warnings [1, 6].

In previous work, increasing adherence to warnings [13] and user reactions to security dialogs [12, 19] have received some research attention. In particular, Bravo-Lillo et al. [9] have researched the habituation effect on system security dialogs. While their paper provided valuable insights into the design of habituation-resistant dialogs, we argue that an important real-world factor has been ignored. In the user's everyday life, ignoring a valid system security dialog might infect the computer with malware or other kinds of unwanted software. However, in previous studies, falsely clicking-through a security dialog had barely any consequences for the user [9].

In this paper, we aim at addressing this specific limitation of previous work concerning the habituation effect. We perform a slightly varied replication study, proposing a risk model which substitutes real world risks like data loss with consequences of monetary loss. Consequently, adhering to a system security dialog earns participants a bonus, while ignoring the instructions resets the bonus back to zero.

The study results (1) confirm previous findings that both habituation and attractors influence the rate of (non-)compliant decisions. Furthermore, we show that (2) monetary incentives have a significant influence in decreasing the non-compliant answers to dialogs. This effect might be related to the exact modeling of the bonus, as our results show that (3) a higher amount of money gained per click shows a greater effect than a small bonus. In addition, we found a small but significant effect on the extent of a participant's first loss and their subsequent behaviour, indicating that (4) bad experiences shape a person's attention, at least for a short period of time.

Background

Habituation Effect Research

Habituation is a form of learning which describes "a decrease in the strength of a naturally elicited behavior that occurs through repeated presentations of the eliciting stimulus" [8]. It was coined by Humphrey [16] and Harris [15] and expanded by Thompson and Spencer, who presented nine characteristics to classify habituation [24].

Several studies have focused on the habituation effect of dialog windows [7, 10, 12].

Research on Monetary Incentives

Studies by Beresford et al. [5] and Grossklags et al. [14] have shown that people are willing to sell private information such as monthly income or the number of sex partners they have had.

Regarding the impact of immediate monetary loss, various research in psychology has demonstrated that there is a strong connection between the timely and spatially distance between a wrong decision and its consequences on learning from an incident [3, 18, 23, 25, 26]. The most prominent among these research results is the Construal Level Theory, which describes the connection between psychological distance and mental abstraction of an event [21].

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 [21], litera-

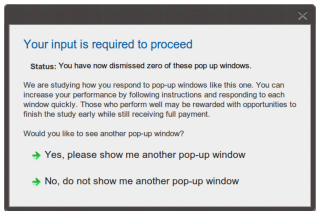

Figure 1: A security dialog from the study by Bravo-Lillo et al. [9]

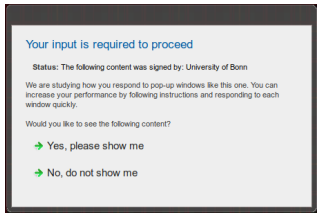

Figure 2: Our new security dialog with adjusted status message, answer options, and description text. Note that messages can also be signed by “Unknown”.

ture from behavioural economics indicates that monetary incentives are a valid methodological tool to model a more arbitrary risk [11, 17].

Study Design

Our work replicates a study by Bravo-Lillo et al. [9] about the habituation effect of warning dialogs. In the original study, participants were asked to dismiss a series of warnings, as quickly as possible, while adhering to their message. After either 1, 2, or 20 of these dialogs, the message switched and compliance to the change was measured. In order to counter the habituation effect, some participant groups had to interact with the dialog in a specific way before the answer options were enabled. Of these habituation-inhibiting *attractors*, the mechanism of swiping over the most important part of the message turned out as being resistant to habituation as well as imposing only a small time overhead for the user.

In our replication study, we slightly change the scenario in which warnings are to be answered. Now, every dialog precedes a content window showing an inspirational quote. Participants are tasked to reject messages with an unknown signature and accept others. Every correct decision to view a message yields a small amount of bonus money which participants were told to receive after finishing the study. Accepting a message with an unknown signature represents a security breach and leads to a loss of the accumulated bonus money. Refusing a message regardless of signature has no negative consequences, since it is always a safe decision to refuse content in our scenario.

We conducted the variation replication experiment on Amazon Mechanical Turk (MTurk) in a between-groups design, where groups were split based on three variables: *Habituation Period*, i.e. the number of same-answer dialogs a user has to answer to build up habituation, *Attractor*, the interaction gateway which was either none or the swipe mechanism [9], and *Bonus Increase per Dialog*, which is either none, 2.5 Cent per correct decision, or 10 Cent per correct decision. The maximum bonus amount was 1 USD. After a participant answered all dialogs in the habituation period, the following dialog is to be rejected in every case, and after that, additional dialogs which are to be rejected are inserted randomly based on the length of the habituation period. In summary, all participants had to answer 41 dialogs after their habituation period. The currently accumulated bonus was always displayed in the interface. Warning dialogs were spawned at random positions within the browser window, however the following content messages always appeared in the center of the screen. 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 task was listed on Amazon MTurk with the same description and properties as in the original study. 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. 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. Regardless of their performance, all participants received this bonus task.

Results

From our initial set of 1,800 MTurkers 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”

Dependent Variables
Habituation Period
Attractor
Bonus Increase
Independent Variables
Compliance at first change
Overall compliance rate
Compliance after first loss

Table 1: Dependent and independent variables in our study

on every dialog. We retained 1,236 valid participants, for whom we report results. 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.

Using linear regression, we found out that both the habituation with 3 dialogs and the habituation with 20 dialogs are responsible for significantly increasing the ratio of compliant clicks, compared to the baseline, with no habituation dialogs. The group with 20 dialogs shows a larger increase than the 3 dialogs group.

Our regression analysis 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, we took participants' self-reported data into account. In the exit survey, we asked participants 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: $\chi^2 = 13.3$, $p = 0.002$).

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.

We hypothesize that a higher amount of accumulated bonus at the point of the first loss increases subsequent compliance. To test this assumption, we correlated the amount of bonus money a participant had accumulated at the time of their first loss with the number of uninterrupted, subsequent non-losses using Kendall's τ and found $\tau = 0.13$, $p < 0.001$. This means that there is a slight, significant positive correlation between losing more money and stronger adherence in the future.

Conclusion

In this work, we conducted and evaluated an extended replication of a study on the habituation effect in the context of system security dialogs. Though the previous work by Bravo-Lillo et al. gave important insights, we created a more realistic experimental design. We additionaly modelled the risk of a malware infection that could result from making a wrong decision.

There are two important takeaways from our study. Firstly, we were not able to fully replicate the results of Bravo-Lillo et al. The 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.

The second important takeaway is that monetary incentives can serve as an effective and promising approach for modeling risk in warning studies, since 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 modelled by Bravo-Lillo et al. and us with habituation effects in the wild.

REFERENCES

1. Devdatta Akhawe and Adrienne Porter Felt. 2013. Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness.. In *USENIX security symposium*, Vol. 13.
2. Android Open Source Project. 2017. Working with System Permissions. (2017). `https://developer.android.com/training/permissions/index.html`
3. Justin Aronfreed and Arthur Reber. 1965. Internalized behavioral suppression and the timing of social punishment. *Journal of Personality and Social Psychology* 1 (1965), 3–16. Issue 1.
4. G. Susanne Bahr and Richard A. Ford. 2011. How and why pop-ups don't work: Pop-up prompted eye movements, user affect and decision making. *Computers in Human Behavior* 27, 2 (2011), 776 – 783. DOI: `http://dx.doi.org/10.1016/j.chb.2010.10.030`
5. Alastair R Beresford, Dorothea Kübler, and Sören Preibusch. 2012. Unwillingness to pay for privacy: A field experiment. *Economics Letters* 117, 1 (2012), 25–27.
6. Rainer Böhme and Jens Grossklags. 2011. The Security Cost of Cheap User Interaction. In *Proceedings of the 2011 New Security Paradigms Workshop (NSPW '11)*. ACM, New York, NY, USA, 67–82. DOI: `http://dx.doi.org/10.1145/2073276.2073284`
7. Rainer Böhme and Stefan Köpsell. 2010. Trained to accept?: a field experiment on consent dialogs. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2403–2406.
8. Mark E. Bouton. 2007. *Learning and Behavior: A Contemporary Synthesis*. Sinauer.
9. Cristian Bravo-Lillo, Lorrie Cranor, Saranga Komanduri, Stuart Schechter, and Manya Sleeper. 2014. Harder to ignore? Revisiting pop-up fatigue and approaches to prevent it. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*. 105–111.
10. José Carlos Brustoloni and Ricardo Villamarín-Salomón. 2007. Improving security decisions with polymorphic and audited dialogs. In *Proceedings of the 3rd symposium on Usable privacy and security*. ACM, 76–85.
11. Daniela Di Cagno, Arianna Galliera, Werner Güth, Francesca Marzo, and Noemi Pace. 2017. (Sub) Optimality and (non) optimal satisficing in risky decision experiments. *Theory and Decision* 83, 2 (2017), 195–243. DOI: `http://dx.doi.org/10.1007/s11238-017-9591-2`
12. Serge Egelman, Lorrie Faith Cranor, and Jason Hong. 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*. ACM, 1065–1074.
13. Adrienne Porter Felt, Alex Ainslie, Robert W Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettes, Helen Harris, and Jeff Grimes. 2015. Improving SSL Warnings: Comprehension and Adherence. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI'15)* (2015), 2893–2902. DOI: `http://dx.doi.org/10.1145/2702123.2702442`

14. Jens Grossklags and Alessandro Acquisti. 2007. When 25 Cents is Too Much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information.. In *WEIS*.

15. J. D. Harris. 1943. Habituatory response decrement in the intact organism. *Psychological Bulletin* 40 (1943), 285–422. Issue 6. DOI: http://dx.doi.org/10.1037/h0053918

16. G. Humphrey. 1930. Extinction and negative adaptation. *Psychological Review* 37 (1930), 361–363. Issue 4.

17. Daniel Kahneman and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society* (1979), 263–291.

18. Gideon Keren and Peter Roelofsma. 1995. Immediacy and Certainty in Intertemporal Choice. *Organizational Behavior and Human Decision Processes* 63, 3 (1995), 287 – 297. DOI:http://dx.doi.org/https://doi.org/10.1006/obhd.1995.1080

19. Soyun Kim and Michael S. Wogalter. 2009. Habituation, Dishabituation, and Recovery Effects in Visual Warnings. *Human Factors and Ergonomics Society Annual Meeting Proceedings* 53, 20 (2009), 1612–1616. DOI:http://dx.doi.org/10.1518/107118109X12524444080675

20. Kat Krol, Matthew Moroz, and M Angela Sasse. 2012. 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* (2012), 1–8.

21. Arie W. Kruglanski and E. Tory Higgins. 2013. *Social psychology: Handbook of basic principles*. Guilford Publications.

22. Microsoft and Brian Lich. 2017. Windows IT Center: User Account Control. (2017). https://docs.microsoft.com/en-us/windows/access-protection/user-account-control/user-account-control-overview

23. Drazen Prelec and George Loewenstein. 1991. Decision Making over Time and Under Uncertainty: A Common Approach. *Manage. Sci.* 37, 7 (July 1991), 770–786. DOI: http://dx.doi.org/10.1287/mnsc.37.7.770

24. R F Thompson and W a Spencer. 1966. Habituation: a model phenomenon for the study of neuronal substrates of behavior. *Psychological review* 73, 1 (1966), 16–43. DOI: http://dx.doi.org/10.1037/h0022681

25. Richard H. Walters and Lillian Demkow. 1963. Timing of Punishment as a Determinant of Response Inhibition. *Child Development* 34, 1 (1963), 207–214. http://www.jstor.org/stable/1126841

26. Ryan West. 2008. The Psychology of Security. *Commun. ACM* 51, 4 (April 2008), 34–40. DOI: http://dx.doi.org/10.1145/1330311.1330320