Improving Computer Security Dialogs

Cristian Bravo-Lillo[1], Lorrie Faith Cranor[1,3], Julie Downs[2], Saranga Komanduri[3], Manya Sleeper[3]

[1]Engineering and Public Policy, [2]Social and Decision Sciences, [3]Computer Science,
Carnegie Mellon University, Pennsylvania, USA
{cbravo, lorrie, downs, msleeper, sarangak}@cmu.edu

Abstract. Security dialogs warn users about security threats on their computers; however, people often ignore these important communications. This paper explores the links between warning dialog design and user understanding of, motivation to respond to, and actual response to computer security warnings. We measured these variables through a 733-participant online study that tested a set of four existing computer security warnings and two redesigned versions of each across low- and high-risk conditions. In some cases our redesigned warnings significantly increased participants' understanding and motivation to take the safest action; however, we were not able to show that participants' responses were differentiated between low and high risk conditions. We also observed that motivation seemed to be a more important predictor of taking the safest action than understanding. However, other factors that may contribute to this behavior warrant further investigation.

Keywords: security warning dialog, usable security

1 Introduction

Warnings are communications designed to protect people from harm [1]. These harms may be immediate, as in the case when road signs warn about sharp turns, or they may be in the future, as in the case of health notices on cigarette boxes. In the case of computer security warnings, the harms arise from immediate and future threats to personal information (e.g., financial data) or property (e.g., computers). However, despite this threat of harm, people often do not read or understand computer security warnings [2-4] and frequently fail to heed them [5], even when the situation is hazardous. There is a lack of empirical evidence about the factors that influence response to computer warnings [6].

This paper uses the results of a 733-participant online study based on a set of existing and redesigned warnings to examine the links between warning design, user understanding of risk, motivation to respond to the risk, and decision to take the least risky action. In this paper, we focus on computer security dialogs, a subset of security warnings, which are warnings that offer users a choice of at least two options.

1.1 Warnings research

In the warnings literature, response to a warning is often evaluated in terms of 'compliance' – performing an action when instructed to do so [7]. Much of the prior research on computer security compliance behaviors focused on phishing attacks or

web browser certificates. In one study, over two-thirds of participants, in a laboratory setting, dismissed anti-phishing warnings because of the websites' look-and-feel and participants' incorrect mental models [8]. Similarly, in an online survey, between 30% and 60% of participants said they would ignore each of the tested browser warnings and continue to a potentially dangerous website. In a subsequent laboratory study, redesigned versions of these warnings achieved greater compliance, but, even in the best case, 45% of participants still ignored the warning when it interfered with their primary tasks [4]. In another laboratory study, about half of the participants ignored a full-page warning presented before an authentication page to an online banking website [5]. Although this behavior may be considered rational from an economic perspective, the problem of how to design effective security communications that do not burden users still remains [9].

Previous research shows high levels of warning non-compliance, even after warning redesign, providing only limited insights into the reasons for non-compliance. People might fail to heed warnings for a variety of reasons, including lack of trust in the warnings, lack of awareness of the risks [2], lack of understanding of the warnings [10], and lack of motivation to comply (perhaps because the required effort is larger than the benefit [9]). Potential consequences for lay users include the possibility of becoming a victim of phishing and other types of scams, of downloading a virus and losing information, of disclosing private and sensitive information, or of being exposed to other harmful threats. This study goes beyond prior research to examine two possible causes of non-compliance: lack of understanding and lack of motivation.

Previous work suggests that lack of understanding may contribute to non-compliance. Egelman et al. observed that some participants who encountered web browser phishing warnings after receiving a phishing email still believed the emails were legitimate. The authors describe a "cognitive dissonance" between the spoofed website and the email leading to it [3]. Motieé et al. reported that 77% of all participants to a laboratory study did not understand the purpose of security warnings and consented to a fake security prompt [11]; in same study, 22% of participants with high level of computer expertise did the same.

There are also qualitative theoretical models that apply to how users interact with computer security warnings. The Communication-Human Information Processing (C-HIP) model [12] describes the human processes involved in the internalization of a warning. In the model, a warning travels from a source through a channel to a receiver. The model focuses on a set of sequential stages–attention switch, attention maintenance, comprehension/memory, attitudes/beliefes, and motivation–through which a receiver processes the warning, resulting in a behavior. The Human-In-The-Loop security framework, based on the C-HIP model [13], can be used to systematically identify security issues created by users who fail to properly carry out security-critical tasks. This framework predicts errors in cases where users do not know how to comply or are unmotivated or incapable of complying with warnings [13]. This study was designed to examine parts of this framework; specifically, it investigates the relationship between understanding, motivation, and user response.

1.2 Safe response

While some previous work talks about warning 'compliance', we use the term 'safe

response' instead. Safe response is an objective measure, that is defined as taking the least risky option provided by a computer security warning dialog. For example, one of the warnings used in this study warns about the possibility that an email attachment may infect a user's computer with a virus. The safe response would be not to open the email attachment, as this is the only response that would present no risk to the user. Any other response, such as opening or saving the attachment, would present some level of risk.

Safe response differs from compliance, which is a concept borrowed from research into physical, non-interactive warnings [7]. In the case of security warning dialogs, we feel that safe response is a clearer metric. In computer systems, there are many situations that may be more or less safe, depending on a context known only to a user. Well-designed security warnings tend to address such situations, as any hazards that could be addressed without contextual knowledge should have been blocked without user intervention. A good security warning will assist the user in using her contextual knowledge to make an informed choice between two or more options, one of which is the least risky option, or the 'safe response.'

High levels of safe response are not always necessary. There is a trade-off between usability and level of risk that is based on the specific context. Always making the least risky choice would allow for a completely safe system but would reduce functionality. A warning is useful if it helps a user to use her knowledge of the context to make an informed decision that balances risk and usability. For example, in the attachment warning outlined above, the 'safe response' would be to not open the attachment. However, within a given context the user should consider factors exogenous to the system, determine how risky the context is, and decide if she should open the attachment. If the user is expecting a file, knows the sender, and can tell from the warning text that this is the file she was expecting, then she finds herself in a low-risk context. In this particular context, the safe response is not necessary and she should open the attachment.

We analyze safe response as being a desirable response in high-risk contexts, under the assumption that users should protect themselves against the high risk of a potential threat, and as being an undesirable response in low risk contexts, under the assumption that it is unnecessary for users to block functionality in these situations. Sunshine et al. took a similar approach in their evaluation of user response to web browser certificate warnings on an online banking login page (high risk) and a university library website (low risk) [4].

2 Methodology

We performed an online survey (n=733) to test the effects of warning design on user understanding, motivation, and safe response. Our study used a 3 x 2 design, with three warning design conditions (E: existing warnings, G: redesigned based on warning design guidelines, and M: redesigned based on our previous work on mental models) and two scenario-based context conditions (S_1: low security priming and S_2: high security priming) for a total of six conditions.

2.1 Warning design conditions

We tested five existing warnings from commercially available software, but report on only the four that are security dialogs. The four warnings, referred to as the *Existing set* (E, see Figure 5 in the Appendix), alerted users about problems encrypting an email (W1), a program trying to access the user's address book (W2), an email attachment (W3), and an unknown certificate (W4).

Table 1: Guidelines used to redesign warnings

Guideline	Examples
1. Follow a visually consistent layout	Use one icon; do not use a close button; use command links for options; use a primary text to explain the risk; describe the consequences of each option below each button.
2. Comprehensively describe the risk	Describe the risk; describe consequences of not complying; provide instructions on how to avoid the risk.
3. Be concise, accurate and encouraging	Be brief; avoid technical jargon; provide specific names, locations and values for the objects involved in the risk; do not use strong terms (e.g., abort, kill, fatal)
4. Offer meaningful options	Provide enough information to allow the user to make a decision; option labels should be answers to explicit question asked to the user; if only one option is available, do not show the warning; the safest option should be the default.
5. Present relevant contextual and auditing information	If the warning was triggered by a known application, describe the application; identify agents involved in the communication by name; if user's information is about to be exposed to risk, describe what information and how it will be exposed.

We created a second set of warnings, referred to as the *Guideline-based set* (G, see Figure 5 in Appendix). Each of the warnings in the E set were redesigned by three HCI Master's students who each had at least one year of HCI coursework as well as previous design experience. We asked the students to redesign the existing warnings by following design guidelines that we compiled from the literature [3, 12-19]. A brief summary of these guidelines is shown in Table 1. We did not provide the designers with any other information about our study.

Similarly, we created a third set of warnings, referred to as the *Mental-model-based set* (M, see Figure 4 in Appendix). To create this set we redesigned each warning in the E set based on previous work on mental models of computer security warnings. In our previous work we found differences in the way experts and non-experts respond to these warnings [20]. We tried to design this set of warnings to include information that experts tend to seek out when responding to a warning, such as the results of analyses by anti-virus software. We also applied many of the guidelines used by the HCI students to create set G.

2.2 Contextual scenarios

Users view security warning dialogs within a specific contextual situation,and make a decision based on that situation. To imitate this context in our online survey, we wrote a *Scenario 1* (S_1) and a *Scenario 2* (S_2) for each warning. Each user who saw a particular warning was presented with a scenario along with that warning. S_1 included low security-priming scenarios with activities that most people would not normally

associate with a security threat; whereas, S_2 included activities that involved sensitive or confidential information, or had characteristics of common security attacks. As warnings must consistently be useful in both low- and high-threat contexts we chose to include both low and high security-priming categories to ensure that our results were consistent across scenarios that presented different threat levels. Table 2 contains all scenarios. We incorporated feedback from security experts when creating the scenarios and strove to ensure that scenarios were of similar readability and length.

Table 2: Scenarios created for the study.

	Low risk Scenario 1 (S1)	Scenario 2 (S2) High risk
W1: Encryption warning	Imagine that you are sending a birthday greeting to your friend Rob by email. You click on the 'Send' button and the warning below appears on your screen.	Imagine that you are sending important financial information to your boss by email. Your boss warned you that it is important to keep this information confidential. You click on the 'Send' button and the warning below appears on your screen.
W2: Address book warning	Imagine that you are trying to connect your PDA or smartphone to your computer to synchronize your email. You plug the device into your computer, and the warning below appears on your screen.	Imagine that you are reading your email. You open a message from your friend Rob, and the message invites you to try out a new social network your friend is using. You click on the invitation, and the warning below appears on your screen.
W3: Attachment warning	Imagine you are reading your email. You open an email from a friend, who says that he is sending you a book he thinks you would find interesting. You double-click on the attachment, and the warning below appears on your screen.	Imagine you are reading your email. You open an email that seems to be from one of your friends, but the email does not contain any text, only a document attached. You double-click on the attachment and the warning below appears on your screen.
W4: Certificate warning	Imagine that you want to buy a gift for a very good friend, but you don't have time to go to a store. You look for a site on the Web, and after searching for a few minutes you find a website that seems to be OK. You click on the link to the website, and the warning below appears on your screen.	Imagine that you need to pay a bill, and you are out of checks. A friend suggests you try paying the bill from your bank's website. You have seen your bank statements online before, but you don't know how to pay bills online. You remember you recently received an email from your bank. You open the email, click on a link to enter the bank's website, and the following warning appears.

2.3 High and low risk conditions

Each warning, in combination with each scenario, presented the user with either a high or low level of risk. Throughout this paper, we refer to the level of risk that the participant faced when presented with a specific warning and contextual scenario combination as either Low Risk (LR) or High Risk (HR). Based on our definition of safe response, when warnings are successful, participants in LR conditions should choose not to take the safe response because the safe response requires them to sacrifice functionality. However, participants in HR conditions should choose the safe response because they should prioritize safety over functionality in risky situations.

We had two low-risk conditions: the encryption and address book warnings with S_1 scenarios. In both cases the risk is minimal and taking the least risky action would

prevent the user from completing her primary task. We had six high-risk conditions: all four warnings with S_2 scenarios, and the attachment and certificate warning with S_1 scenarios.[1] In these cases, the level of risk warranted taking the safe response.

A well-designed security dialog should allow participants to differentiate between low- and high-risk conditions. It should create a higher rate of motivation and safe response for high-risk conditions than for low-risk conditions. If the warnings in our study were well designed we would expect to see warnings with the same level of risk in S_1 and S_2 (attachment and certificate warnings) to have similar rates of motivation and safe response. We would also expect to see warnings with low risk in S_1 and high risk in S_2 (encryption and address book warnings) to have higher levels of safe response and motivation in S_2.

Table 3: Number of participants in each condition.

	W1		W2		W3		W4	
Scenario	**S_1**	**S_2**	**S_1**	**S_2**	**S_1**	**S_2**	**S_1**	**S_2**
E	145	124	114	145	125	114	106	125
G	119	106	124	119	145	124	114	145
M	125	114	106	125	119	106	124	119

2.4 Survey design and participant recruiting

Our survey consisted of 69 questions divided into seven sections, starting and ending with demographic questions. Each of the remaining five sections included a randomly selected image of a warning, a randomly selected corresponding scenario (S_1 or S_2), and a set of questions about each warning.

We recruited participants using Amazon's Mechanical Turk service [20], paying each participant who completed the study 50 cents. We required participants to be computer users, over 18 years old, English-speaking and residents of the United States. Participants took an average of 10 min 47 sec to answer the survey (σ = 7 min 9 sec). We discarded 3 responses that took less than 10 seconds. We were left with 733 respondents, about 62% of whom were females and four-fifths of whom were Caucasian. The number of participants in each condition is summarized in Table 3. Participants ranged in age from 18 to 75, with a mean age of 32.9 (σ = 11.58). We also collected information about usage of operating systems, browsers, and email clients to test any correlation with our dependent variables. As described later, we found no consistent relationship between demographics and dependent variables.

We also asked two questions to probe participants' level of technical expertise: whether they had ever taken or taught a course on computer security, and whether they knew any computer languages. If they answered the latter affirmatively, we asked which languages they knew. Participants who answered only HTML were not considered as having programming expertise. We found no significant correlation

1 The content of the attachment and certificate warnings (see Appendix) was suspicious enough to suggest a high-risk situation, even in S_1.

between affirmative answers and any studied variables, so we excluded these questions from our analyses.

Table 4: Questions asked to participants per warning, and the corresponding measured variable.

Dependent variable	Question	Types of answers	Explanation
Under-standing	*What do you think is/are the problem(s)?*	11 common problems plus an *Othe*r open text field	If participants answered at least one of the correct answers and none of the incorrect answers (based on authors' knowledge and interviews with security experts [20]), understanding was recorded as 1, otherwise as 0.
Motivation	*The problem described by this warning is very important.*	5-point Likert response, from *Strongly disagree* to *Strongly agree*	If participants answered *Agree* or *Strongly agree*, motivation was recorded as 1, otherwise as 0.
Safe response	*What would you do in this situation?*	As many clickable options as the warning offered, plus *Ignore this warning* and *Take another action*	If participants answered at least one action considered safe by experts and none of the actions considered unsafe by experts, safe response was recorded as 1, otherwise as 0.

2.5 Hypotheses

To develop our hypotheses, we defined three dependent variables: understanding, motivation and safe response. These variables are described in Table 4. We also defined low- and high-risk conditions consisting of combinations of warnings and scenarios, as given below:

> **Low-risk condition**: W1 with S_1, W2 with S_1.
>
> **High-risk conditions**: W1 with S_2, W2 with S_2, W3 with S_1 or S_2, and W4 with S_1 or S_2.

We hypothesized that understanding would be higher for all conditions in the redesigned warnings than in the existing set. For motivation and safe response we hypothesized that they would be significantly higher in the redesigned warnings for participants in the high-risk conditions but would not be significantly higher for participants in the low-risk condition. We also hypothesized that understanding and motivation would be found to drive safe response. Our hypotheses are enumerated below:

- **H_1**: For all warnings and scenarios, understanding will be significantly higher in the guidelines-based (G) and mental-model-based (M) sets than in the existing set (E).
- **H_2**: For all low-risk scenarios, motivation and safe response will not be significantly higher in the redesigned sets (G and M) than in the existing set (E).
- **H_3**: For all high-risk scenarios, motivation and safe response will be significantly higher in the redesigned sets (G and M) than in the existing set (E).

H_4: Understanding and motivation will be significant predictors of safe response across all warning sets and scenarios, controlling for demographic factors.

3 Analysis

Based on an analysis of the four warnings we found that understanding and motivation were strongly correlated with safe response. However, we were not able to conclude that users could differentiate between low-risk and high-risk conditions, and we did not see a significant increase in motivation and safe response for W1 and W2 in either the high- or low-risk conditions. However, we did find improvements in motivation and safe response for W3 and W4, the two warnings that were only presented in high-risk conditions.

We analyzed our results separately for each warning using logistic regression. Logistic regression is similar to linear regression except that it is used to analyze data with a binary dependent variable. Factors with significant p-values are significant predictors of the dependent variable, controlling for all other factors (see Tables 4 and 5 in Appendix). We used a significance level of $\alpha = .05$ for all analyses.

3.1 Understanding

In general, our redesigned sets of warnings (G and M) failed to increase understanding over existing warnings. We observed significant increases in understanding in only 3 out of 16 conditions, and in two cases related to W2 we observed significant decreases in understanding. Figure 1 shows our results for understanding. Statistical data are given in Table 5 in the Appendix.

We expected to see increased levels of understanding for the G and M sets versus the E set (H_1). While this occurred in a few conditions, understanding did not increase in the majority of cases (see Table 5 in Appendix). Because understanding increased in more conditions in which participants were shown S_1 than S_2, we tested the possibility that participants spent less time on the scenarios by comparing the mean time that participants took to answer each warning section. However, we found no significant differences between times for the two sets of scenarios.

In the S_1 scenario for the address book warning (W2), the understanding rate was significantly lower for the G and M sets than in the E set. To help explain this lower level of understanding we looked at the specific problems that users thought the warning presented. We found that a higher percentage of respondents believed that the warning was related to a website in the G and M sets than in the E set, which was a "wrong" answer. The misunderstanding was potentially due to a reference to ABC.exe (the program accessing the computer) that only appeared in the redesigned warnings. We speculate that respondents may have mistaken ABC.exe for a website. We mandated in our guidelines that a program prompting a warning should be identified to users, to help them better decide how to respond, but the implementation of this recommendation could have resulted in confusion.

The redesigned warnings (G and M) were also less likely to prompt two 'right' answers than the existing (E) warning. For the G and M versions of the address book warning in the S_1 scenario, participants were less likely to respond that they did not

trust the software being run or that there was no problem than when shown the E version of the warning. Participants may not have considered ABC.exe to be software, or perhaps they considered the redesigned warnings more threatening than the existing warning. Additional testing is necessary to determine which aspects of the warnings lead to misunderstanding.

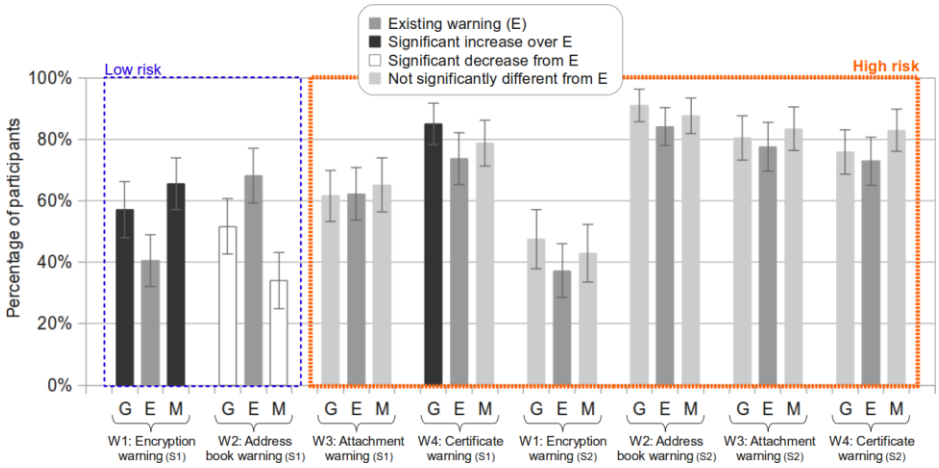

Figure 1: Percentage of participants who showed understanding of the problem that triggered the studied warnings, in the low- and high-risk conditions. G, E, and M correspond to the different sets of warnings. The top bars represent confidence intervals at the 95% level.

These results provide very limited, if any, support for H_1. It should be noted, however, that many warning-scenario combinations had a high initial level of understanding, from which it may be difficult to introduce improvements.

3.2 Motivation

Our redesigned warning sets (G and M) had some success at increasing levels of motivation in the high-risk condition for W3 and W4, but did not show evidence of allowing participants to differentiate between low- and high-risk conditions. Figure 2 shows our results for motivation. Statistical data are given in Table 6 in the Appendix.

If the redesigned warnings allowed participants to differentiate between high- and low-risk contexts and respond appropriately, there would be no change in motivation levels between G/M and E in the low-risk condition, but there would be an increase in motivation levels for the redesigned warnings in the high-risk condition. We were not able to conclude that the redesigned warnings allowed users to differentiate between low- and high-risk contexts. For the encryption warning and address book warning (W1 and W2), which were shown in both high- and low-risk contexts, there was no significant improvement in motivation in the majority of cases in either context.

In the low-risk context we expected motivation not to be significantly higher for the redesigned warnings (G and M) than the existing warnings (E). This held for three out of four cases, providing support for H_2. However, for these results to be

meaningful, we needed to see a corresponding increase in motivation for these same warnings (W1 and W2) in a high-risk context, proving that participants could differentiate between the levels of risk with the redesigned warning set and respond appropriately. However, we found that in all four high-risk cases for W1 and W2 there was no significant difference between the E set and each of the G and M sets for motivation. This indicates that the lack of improvement in the low risk case may have represented a lack of improvement overall, rather than participants' abilities to differentiate between risk levels. Thus, while these results support H_2, they are inconclusive.

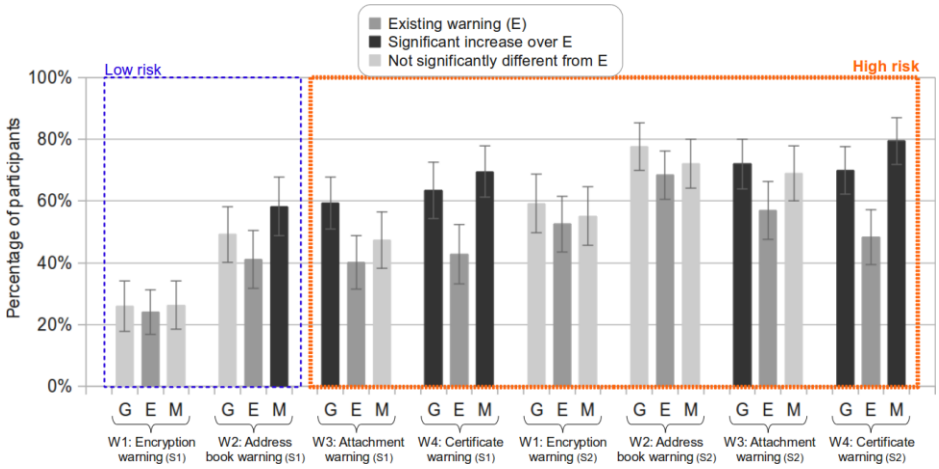

Figure 2: Motivation, as measured by the percentage of participants who agree or strongly agree that the problem described by the warning is very important, in the low- and high-risk conditions. G, E, and M correspond to the different sets of warnings. The top bars represent confidence intervals at the 95% level.

Although there was no evidence that the redesigned warnings allowed participants to differentiate between low- and high-risk contexts, we did find some evidence that the redesigns improved motivation in the high-risk context (H_3). For the attachment and certificate warnings (W3 and W4), which were only shown in high-risk contexts, we found that the redesigned warnings significantly increased motivation in all but one case. As previously described, we expected to see similar results for W1 and W2 in the high-risk context, but did not see any significant differences between G/M and E for W1 and W2.

3.3 Safe response

We found that the redesigned warnings were successful at increasing safe response in the majority of the high risk conditions. However, as was the case with motivation, we were not able to conclude that the redesigned warnings allowed participants to differentiate between high- and low-risk conditions and respond appropriately. Figure 3 shows our results for safe response. Statistical data are given in Table 6 in the Appendix.

As described previously, safe response measures the proportion of participants who pick the option that presents the least risk. We expected participants' rates of safe response to significantly increase for the high-risk conditions for our redesigned warnings and to remain the same for the low-risk conditions. In the low-risk conditions the redesigned warnings should not push participants to pick a safe response that would prevent them from completing the desired task. For the two warnings that we presented in both the high- and low-risk conditions, W1 and W2, we found that, as expected, in three out of four cases, the level of safe response was not higher for the G and M sets than for the E set. However, for these two warnings we also found that, in three out of four cases, the level of safe response did not increase in the high-risk condition for G and M compared to E, indicating that the lack of improvement in the low-risk condition may have been due to an overall lack of improvement rather than participants' ability to differentiate between risk levels. So, although we found some evidence for H_3, our overall results for safe response for warnings W1 and W2 were inconclusive.

We did, however, find a significant increase in safe response levels for the redesigned warnings (G and M) over the existing set (E) for the two warnings that were presented in only the high risk condition, W3 and W4. For these warnings, rates of safe response significantly increased in seven out of eight cases.[2] This result provides some support for H_3.

3.4 Correlation between variables

We hypothesized that understanding and motivation would be predictors of safe response (H_4). We found significant correlation between safe response and understanding, motivation, and other variables (see Table 6 in Appendix), supporting H_4. The higher logistic regression coefficients show that safe response is strongly tied to motivation and also linked, although slightly less strongly, to understanding. Although these results do not prove that understanding and motivation drive safe response, they provide some indication that the variables are strongly related.

Motivation and understanding were significantly correlated with each other for all warnings. Motivation was also significantly correlated with safe response for all four warnings for all warning sets. Understanding was also significantly correlated with safe response for all except the encryption warning (W1). Based on the regression coefficients, motivation was more strongly correlated with safe response for all of the warnings in which both factors were significant, except for the address book warning (W2).

Outside of motivation and understanding, we also found interactions between age and being a user of Microsoft Internet Explorer for the address book (W2) and the certificate (W4) warnings. This was expected, as these users have likely encountered these warnings before. In the address book warning, users of Internet Explorer were more likely to pick the safest response, while in the certificate warning (W4), the opposite relation held.

---

2 We performed a qualitative analysis of participants' open comments at the end of each warning to test the possibility that these higher levels of safe response were due to the novelty of redesigned warnings. We found no evidence of such behavior.

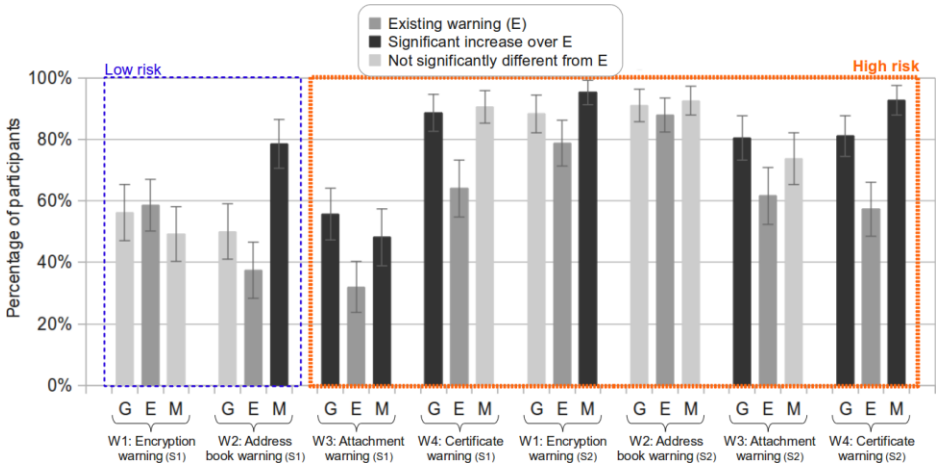

Figure 3: Percentage of participants who took the safest option, in the low- and high-risk conditions. G, E, and M correspond to the different sets of warnings. The top bars represent confidence intervals at the 95% level.

4 Discussion

One of the primary goals of this study was to show differentiated results for low- and high-risk conditions to demonstrate that our redesigned warnings improved participants' abilities to make appropriate security choices in each of the conditions. However, our results did not show differentiated motivation and safe response improvements for the low- and high-risk conditions. For both of the warnings that were presented in low- and high-risk conditions (W1 and W2) we found that in the majority of cases motivation and safe response did not significantly increase for the redesigned warnings in both conditions. It is likely that the redesigned warnings were not more effective than existing warnings and were not able to increase motivation or safe response in either case. It is also possible that the high security-priming scenarios that were used to prompt the high-risk condition were poorly designed and did not prompt a high-risk response. However, this is less likely as 3 out of 8 had significantly higher levels of motivation and safe response for the high-risk condition. Further research is needed to better determine how users respond to high- and low-risk conditions and how to consistently design better security warning dialogs.

One of our redesigned warnings, the M version of the address book warning (W2), turned out to be particularly ineffective. It decreased participants' understanding, increased user motivation and safe response in the low-risk condition, and did not increase motivation or safe response in the high-risk condition. One potential explanation for this unexpected behavior is the amount of information that version contained: the existing version had 44 words and 4 options, and the guidelines-based version had 40 words and 3 options, while the mental-model-based version had 163 words and 6 options. The extra text included the results of an anti-virus scan, and an explanation of the consequences for each option. The large amount of information may

have undermined participants' abilities to understand (or motivation to read) the redesigned warning, or some element of the added text might have confused them.

Although our redesigned warnings appear not to help participants differentiate between high- and low-risk conditions, we were able to demonstrate that it is possible to use a relatively simple redesign process to improve some security warning dialogs for high-risk conditions. Beyond the importance of testing whether participants could differentiate between high- and low-risk conditions, it was also important to show that our results were applicable across different types of contextual scenarios. To do so, we presented participants with low and high security-priming contexts (S_1 and S_2). Further work is necessary to determine which aspects of the redesigns contributed to the successful increases in motivation and safe response and which aspects were not successful at increasing understanding, motivation and safe response.

4.1 Limitations

Our study had a variety of limitations, some of which we hope to improve upon in future work. First, the study is based on self-reported survey data, and as such it may not reflect what users would do when confronted with warnings during their regular computer use. Also, literature suggests that habituation should be considered when studying warnings [12]. To the best of the authors' knowledge, repeated, long-term exposure to computer warnings has not been studied, in part because of the difficulties in setting up adequate experimental designs. However, a deeper look at the answers of our participants show that only a small proportion of them reported that they ignored our warnings, either because they had seen them before or for other reasons. If our participants had been habituated to our set of existing warnings, we would expect to have seen a higher number of people ignoring them. Another factor that might have affected participants' response is the novelty of redesigned warnings. Although we found no evidence in this direction, this remains a limitation of our study.

Another confounding factor might be the possible learning process that takes place after repeated exposures to the same set of questions with different warnings. A technical limitation of the software we used to implement the survey[3] prevented us from tracking the random order in which participants saw our warnings. Although randomization might counter-balance learning effects, we acknowledge that this does not necessarily cancel out the effects. One improvement to the experimental design would be to show a single warning to each participant. We decided to show five warnings instead of one to reduce the number of participants needed for the study.

Our redesigned sets utilized different layouts of options, longer and more descriptive texts for each option, information about context, and the results of analysis by other tools. However, our experimental design did not allow us to isolate the impact of each of these design changes. In future work we expect to better isolate specific factors.

4.2 Conclusion

By comparing existing computer security warnings with two sets of warnings that we

3 SurveyGizmo, available at http://www.surveygizmo.com

created, we explored relationships between the design of the warning, understanding of the problem underlying a warning, the belief that the problem is important (motivation), the tendency to pick the safest option (safe response), and demographic factors. We found that design changes can lead to improvements in understanding, motivation, and tendency to pick the safest option in some cases, but further work is needed to isolate the impact of various design factors. However, we were unable to help participants differentiate between the appropriate option in high- and low-risk conditions. We also found that although understanding and motivation are strongly tied to each other, motivation is a slightly more important factor than understanding when it comes to increasing safe response to warnings.

Warning designers should keep in mind that both the level of importance that users attribute to a warning and the understanding of the problem underlying a warning contribute to user response. To be successful, warnings should both motivate a user to respond, and help users understand the risk, in that order. Future work should look at exactly how much each of these factors, and other factors, contribute to increasing safe response to warnings.

4.3 Acknowledgements

This research was funded in part by NSF grant CNS0831428. The last author wishes to thank the ARCS Foundation for their support.

References

[1] Wogalter, M.S.: Purposes and scope of warnings. In Wogalter, M.S., ed.: Handbook of warnings. Human Factors and Ergonomics. First edn. Lawrence Erlbaum Associates, Mahwah, New Jersey (2006) 3–9

[2] Downs, J.S., Holbrook, M.B., Cranor, L.F.: Decision strategies and susceptibility to phishing. In Cranor, L.F., ed.: Proceedings of the 2nd Symposium on Usable Privacy and Security, (SOUPS). Volume 149 of ACM International Conference Proceeding Series., ACM (July 2006) 79–90

[3] Egelman, S., Cranor, L.F., Hong, J.I.: You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In Czerwinski, M., Lund, A.M., Tan, D.S., eds.: Proceedings of the 2008 Conference on Human Factors in Computing Systems (CHI), ACM (April 2008) 1065–1074

[4] Sunshine, J., Egelman, S., Almuhimedi, H., Atri, N., Cranor, L.F.: Crying wolf: An empirical study of ssl warning effectiveness. In: Proceedings of the 18th Usenix Security Symposium, Usenix Security Symposium (August 2009)

[5] Schechter, S.E., Dhamija, R., Ozment, A., Fischer, I.: The emperor's new security indicators. In: SP '07: Proceedings of the 2007 IEEE Symposium on Security and Privacy, Washington, DC, USA, IEEE Computer Society (2007) 51–65

[6] Camp, L.J.: Mental models of privacy and security. Technology and Society Magazine, IEEE **28**(3) (Fall 2009) 37–46

[7] Meyer, J.: Responses to dynamic warnings. In Wogalter, M.S., ed.: Handbook of warnings. Human Factors and Ergonomics. First edn. Lawrence Erlbaum Associates, Mahwah, New Jersey (2006) 221–229

[8] Wu, M., Miller, R.C., Garfinkel, S.L.: Do security toolbars actually prevent phishing attacks? In Grinter, R.E., Rodden, T., Aoki, P.M., Cutrell, E., Jeffries, R., Olson, G.M., eds.: Proceedings of the Conference on Human Factors in Computing Systems (CHI), ACM (April 2006) 601–610
[9] Herley, C.: So long, and no thanks for the externalities: the rational rejection of security advice by users. Proceedings of the 2009 workshop on new security paradigms workshop, NSPW '09, pages 133–144. New York, NY, USA, 2009. ACM.
[10] Downs, J.S., Holbrook, M.B., Cranor, L.F.: Behavioral response to phishing risk. In Cranor, L.F., ed.: Proceedings of the Anti-Phishing Working Groups 2nd Annual eCrime Researchers Summit 2007. Volume 269 of ACM International Conference Proceeding Series., ACM (October 2007) 37–44
[11] Motieé, S., Hawkey, K., Beznosov, K.: Do windows users follow the principle of least privilege?: investigating user account control practices. In Proceedings of the Sixth Symposium on Usable Privacy and Security, New York, NY, USA, 2010, 1–13.
[12] Wogalter, M.S.: Communication-human information processing model. In Wogalter, M.S., ed.: Handbook of warnings. Human Factors and Ergonomics. First edn. Lawrence Erlbaum Associates, Mahwah, New Jersey (2006) 51–61
[13] Cranor, L.F.: A framework for reasoning about the human in the loop. In Churchill, E.F., Dhamija, R., eds.: Usability, Psychology, and Security, USENIX Association (April 2008)
[14] Apple Inc.: Apple human interface guidelines. Online document available at http://developer.apple.com (2010)
[15] Benson, C., Elman, A., Nickell, S., Robertson, C.Z.: Gnome human interface guidelines 2.2.1. Online document available at http://library.gnome.org (2010) Last visit on Apr/08/2010.
[16] Microsoft Corporation: Windows user experience interaction guidelines. Online document available at http://msdn.microsoft.com (2010) Last visit on Apr/08/2010.
[17] Egelman, S.: Trust me: Design Patterns for Constructing Trustworthy Trust Indicators. PhD thesis, School of Computer Science, Carnegie Mellon University (2009) Available as technical Report CMU-ISR-09-110.
[18] Nodder, C.: Users and trust: a microsoft case study. In Cranor, L.F., Garfinkel, S.L., eds.: Security and Usability: Designing secure systems that people can use. Theory in practice. First edn. O'Reilly Media, Inc., Sebastopol, CA, USA (2005) 589–606
[19] Ross, B.: Firefox and the worry-free web. In Cranor, L.F., Garfinkel, S.L., eds.: Security and Usability: Designing secure systems that people can use. Theory in practice. First edn. O'Reilly Media, Inc., Sebastopol, CA, USA (2005) 577–587
[20] Bravo-Lillo, C., Cranor, L., Downs, J., Komanduri, S.: Bridging the gap in computer security warnings: a mental model approach (to appear). Security and Privacy Magazine, IEEE (2011)
[21] Ross, J., Irani, L., Silberman, M.S., Zaldivar, A., Tomlinson, B.: Who are the crowdworkers? : shifting demographics in mechanical turk. In: CHI EA '10: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems, New York, NY, USA, ACM (2010) 2863–2872.

Appendix

Table 5: Comparison of percentage of participants that showed understanding (top), motivation (middle) and safe response (bottom) between warning sets. Black cells show significant increases over existing set, dark grey show significant decreases from existing set, and light gray cells show non-significant differences from existing set. c is coefficient, SE is standard error, z is z-value, and p is p-value.

	Scenario 1 (S1)		Scenario 2 (S2)	
W1: Encryption warning	**E (41%) < G (57%)** c=0.668; SE=0.260; z=2.569; p=.010	**E (41%) < M (66%)** c=1.025; SE=0.260; z=3.945; p<.001	E (37%) ~ G (48%) c=0.423; SE=0.274; z=1.542; p=.123	E (37%) ~ M (43%) c=0.238; SE=0.273; z=0.871; p=.384
W2: Address book warning	**E (68%) > G (52%)** c=-0.696; SE=0.278; z=-2.508; p=.012	**E (68%) > M (34%)** c=-1.428; SE=0.294; z=-4.859; p<.001	E (84%) ~ G (91%) c=0.648; SE=0.408; z=1.590; p=.112	E (84%) ~ M (88%) c=0.291; SE=0.364; z=0.799; p=.424
W3: Attachment warning	E (62%) ~ G (62%) c=-0.027; SE=0.258; z=-0.105; p=.916	E (62%) ~ M (65%) c=0.125; SE=0.273; z=-0.458; p=.647	E (78%) ~ G (81%) c=0.178; SE=0.328; z=0.541; p=.588	E (78%) ~ M (83%) c=0.380; SE=0.352; z=1.079; p=.280
W4: Certificate warning	**E (74%) < G (85%)** c=0.703; SE=0.352; z=2.00; p=.046	E (74%) ~ M (79%) c=0.279; SE=0.318; z=0.878; p=.380	E (73%) ~ G (76%) c=0.157; SE=0.288; z=0.547; p=.585	E (73%) ~ M (83%) c=0.596; SE=0.324; z=1.840; p=.066

	Scenario 1 (S1)		Scenario 2 (S2)	
W1: Encryption warning	E (24%) ~ G (26%) c=0.098; SE=0.296; z=0.330; p=.741	E (24%) ~ M (26%) c=0.115; SE=0.289; z=0.399; p=.690	E (53%) ~ G (59%) c=0.271; SE=0.272; z=0.996; p=.319	E (53%) ~ M (55%) c=0.105; SE=0.268; z=0.390; p=.696
W2: Address book warning	E (41%) ~ G (49%) c=0.325; SE=0.296; z=1.207; p=.227	**E (41%) < M (58%)** c=0.692; SE=0.280; z=2.470; p=.014	E (68%) ~ G (78%) c=0.474; SE=0.294; z=1.613; p=.107	E (68%) ~ M (72%) c=0.178; SE=0.275; z=0.647; p=.518
W3: Attachment warning	**E (40%) < G (59%)** c=0.779; SE=0.256; z=3.049; p=.002	E (40%) ~ M (47%) c=0.291; SE=0.264; z=1.102; p=.270	**E (57%) < G (72%)** c=0.664; SE=0.283; z=2.344; p=.019	E (57%) ~ M (69%) c=0.515; SE=0.289; z=1.782; p=.075
W4: Certificate warning	**E (43%) < G (64%)** c=0.849; SE=0.283; z=3.00; p=.003	**E (43%) < M (69%)** c=1.117; SE=0.282; z=3.956; p<.001	**E (48%) < G (70%)** c=0.909; SE=0.262; z=3.472; p=.001	**E (48%) < M (79%)** c=1.419; SE=0.296; z=4.795; p<.001

	Scenario 1 (S1)		Scenario 2 (S2)	
W1: Encryption warning	E (59%) ~ G (56%) c=-0.098; SE=0.259; z=-0.378; p=0.7054	E (59%) ~ M (49%) c=-0.382; SE=0.253; z=-1.513; p=0.1303	E (79%) ~ G (88%) c=0.712; SE=0.381; z=1.870; p=.061	**E (79%) < M (95%)** c=1.702; SE=0.510; z=3.334; p=.001
W2: Address book warning	E (37%) < G (50%) c=0.516; SE=0.272; z=1.898; p=0.0576	**E (37%) < M (79%)** c=1.819; SE=0.313; z=5.819; p<5.9e-09	E (88%) ~ G (91%) c=0.333; SE=0.425; z=0.783; p=.434	E (88%) ~ M (93%) c=0.541; SE=0.437; z=1.237; p=.216
W3: Attachment warning	**E (32%) < G (56%)** c=0.982; SE=0.261; z=3.761; p<.0001	**E (32%) < M (48%)** c=0.684; SE=0.271; z=2.523; p=.012	**E (62%) < G (81%)** c=0.942; SE=0.306; z=3.081; p=.002	E (62%) ~ M (74%) c=0.559; SE=0.300; z=1.865; p=.062
W4: Certificate warning	**E (64%) < G (89%)** c=1.490; SE=0.369; z=4.040; p<.0001	**E (64%) < M (91%)** c=1.696; SE=0.377; z=4.495; p<.0001	**E (57%) < G (81%)** c=1.166; SE=0.288; z=4.053; p<.001	**E (57%) < M (93%)** c=2.268; SE=0.410; z=5.530; p<.001

Table 6: Logistic regression coefficients of interactions between variables (H_4), per warning. Dark cells show significant, positive values, and grey cells show significant negative values.

	W1: Encryption warning	W2: Address book warning	W3: Attachment warning	W4: Certificate warning
Understanding	c=0.2693; se=0.3743; z=0.720; p=0.4718	**c=1.7099; se=0.4317; z=3.961; p=7.46e-05**	**c=0.7567; se=0.1911; z=3.959; p=7.53e-05**	**c=0.4945; se=0.2277; z=2.172; p=0.0298**
Motivation	**c=0.9021; se=0.3670; z=2.458; p=0.0140**	**c=1.4442; se=0.4158; z=3.473; p=0.00051**	**c=1.6107; se=0.1751; z=9.195; p<2e-16**	**c=1.6113; se=0.2125; z=7.582; p=3.41e-14**
Gender	c=-0.4687; se=0.3420; z=-1.370; p=0.1706	c=-0.0371; se=0.4255; z=-0.087; p=0.9304	c=-0.2056; se=0.1799; z=-1.143; p=0.2532	c=-0.1445; se=0.2096; z=-0.690; p=0.4904
Age	c=-0.0045; se=0.0223; z=-0.204; p=0.8384	c=0.0724; se=0.0451; z=1.606; p=0.1083	**c=0.0347; se=0.0150; z=2.313; p=0.0207**	c=0.0003; se=0.0156; z=0.025; p=0.9802
Use of Internet Explorer	c=0.6684; se=1.0370; z=0.645; p=0.5192	**c=2.8604; se=1.4160; z=2.020; p=0.0433**	c=-0.2554; se=0.5716; z=-0.447; p=0.6549	**c=-1.7783; se=0.6596; z=-2.696; p=0.0070**
Use of Internet Explorer – Age	c=-0.0024; se=0.0296; z=-0.082; p=0.9347	c=-0.0837; se=0.0501; z=-1.670; p=0.0948	c=-0.0035; se=0.0177; z=-0.200; p=0.8414	**c=0.0536; se=0.0205; z=2.608; p=0.0091**

Figure 4: Mental-model-based (M) set of warnings.

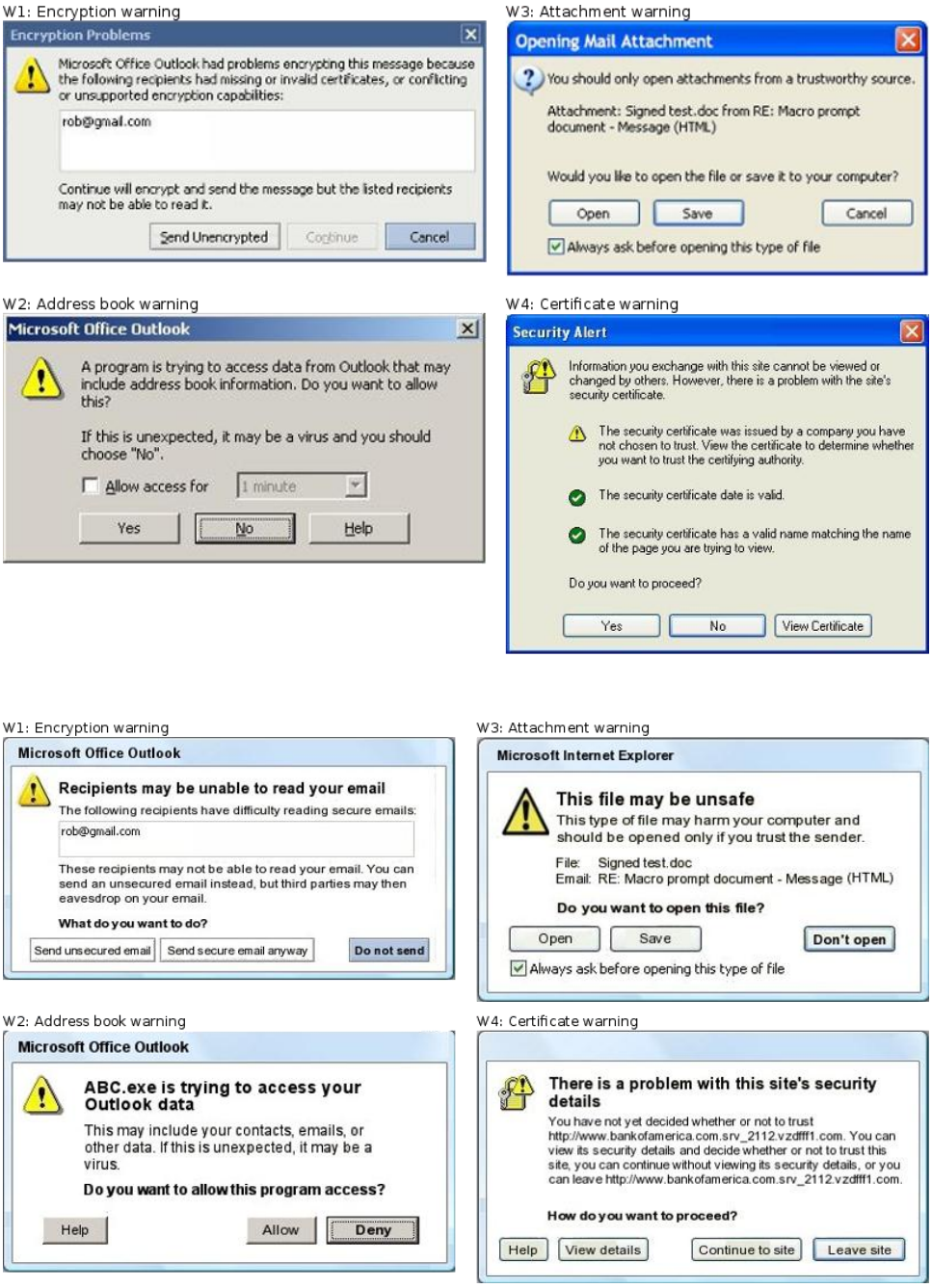

Figure 5: Existing (E, top) set and Guidelines-based (G, bottom) sets of warnings.