POSTER: What is still wrong with security warnings: a mental models approach

Cristian Bravo-Lillo
cbravo@cmu.edu

Lorrie Cranor
lorrie@cs.cmu.edu

Julie Downs
downs@cmu.edu

Saranga Komanduri
sarangak@andrew.cmu.edu

1. INTRODUCTION

Warnings are a form of communication specifically designed to protect people from harm [9]. There is evidence that people do not read computer warnings [4] [8], do not understand them [3], or simply do not heed them [7], even when the situation is clearly hazardous. Most of this evidence comes from studying users' responses to potential phishing threats, and a variety of explanations have been offered for this behavior: they are unaware of the risks [3], they do not understand the options presented to them [4], or their trust in the system causes them to underweigh the risk [6]. Two sources, Wu et al. [10] and Egelman et al. [4] agree that users may have "wrong mental models" and thus they might be applying incorrect beliefs and inadequate strategies to their problems. Much research has been performed on mental models in different areas [5], yet few studies have applied them to computer security or privacy risk communication [1].

This paper describes a work-in-progress aimed at understanding users' mental models of computer risks, and applying this knowledge to build more effective warnings.

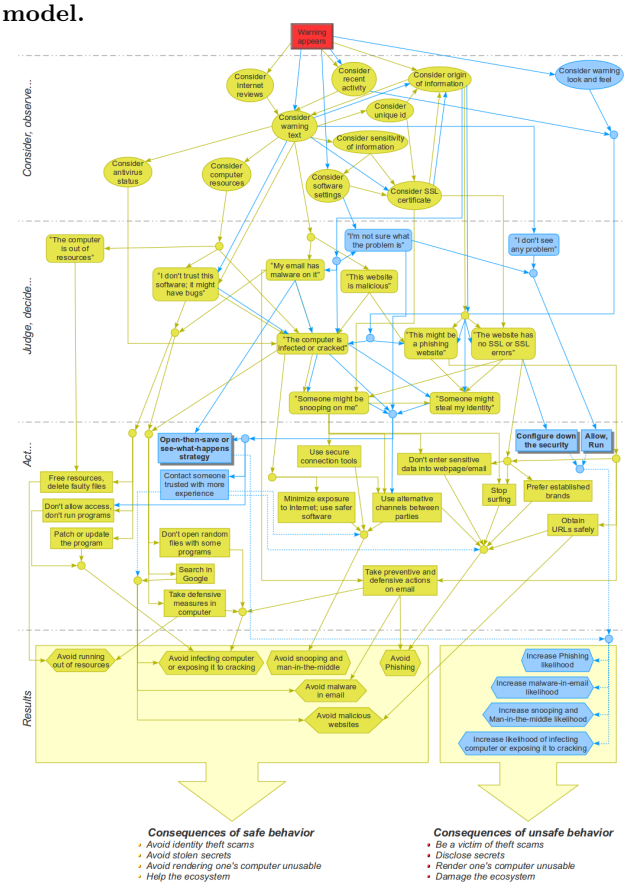
2. METHODS

Our approach is based on the mental model methodology used by Downs et al. to analyze phishing behavior [3] and vaccination patterns [2]. We conducted one-on-one, open-ended interviews with 10 advanced users in security and privacy and 10 novice users. Interviews averaged one hour in length. In these interviews, we showed each participant five selected warning dialogs, and for each dialog we asked them what their advice would be to a very close, non-technically savvy friend. This friend was role-played by the interviewer, who provided context along with each warning screenshot.

A mental model was derived from these interviews as follows. Two researchers independently read the transcripts of the interviews and classified participants' responses into a large number of categories (that we call **codes**). These codes were then compared and the differences discussed until agreement was reached. The same process was followed to determine meaningful relationships between codes.

A large graph was generated showing these relationships (arrows) for both advanced (yellow nodes) and novice users (blue nodes). A greatly reduced version of this graph is shown in Figure 1. The comparison of advanced and novice users' responses to warning dialogs allowed us to determine consistent differences in behavior between advanced and novice users.

Figure 1: Small representation of the created mental model.



3. RESULTS

Table 1 presents the main differences in behavior, with respect to warning dialogs, between our advanced and novice participants. We also found that:

- When faced with a particular warning dialog, advanced users saw fewer potential problems, considered more factors, and acted proactively. In contrast, novice users tended to see more problems, considered fewer variables, and acted reactively.
- Advanced users frequently expect warning dialogs to notify them about particular events (e.g., they *expect* to see a warning for websites using self-signed certifi-

Table 1: Behavioral differences between advanced and novice users in response to warning dialogs

	Advanced users...	Novice users...
1	Maintain a dynamic image of the state[1] of their computer	Often do not know the state of their computer
2	Keep in mind information obtained from expert sources about new and common problems	Are not exposed to reliable information about new and common problems
3	Often read warning texts	Often ignore warning texts
4	Can predict the consequences of their actions	Cannot predict the consequences of their actions
5	Are aware of the risks, and usually estimate their likelihood accurately	Are not aware of the risks, and usually underestimate their likelihood
6	Assess the safety of an action **before** engaging in it	Can only assess the safety of an action **after** engaging in it

cates.)

- Novice users tend to consider irrelevant data, are unable to use data to separate one problem or another, and are unaware of the different consequences of actions such as 'opening' or 'saving' a file.

- Our novice participants seemed to have heard about different problems (e.g., viruses, malware, phishing, and identity theft) and tried to relate those terms with the problems they faced, even when they were not related.

The literature we reviewed suggests many different guidelines for improving warning dialog design. Many of these are concerned with the visual layout and the quality of the message being delivered. We produced complementary guidelines that focus on what users need to know, but do not already, as suggested by Morgan et al. for risk communications [5]. Warning dialogs should:

1. Inform about the state of the computer when needed.

2. Display dynamically relevant information from expert sources.

3. Include pre-attentive features such as color, shape, closure, contrast, and others.

4. Inform clearly about the consequences of the actions offered through buttons and links.

5. Inform about the risks involved and their probability of occurrence (when relevant).

6. Warn about the safety of actions before letting users engage in them.

These guidelines follow from the differences between advanced and novice users discusssed previously. They correspond directly to the rows in Table 1.

4. WORK IN PROGRESS

We are currently conducting a massive online study, using SurveyGizmo and Amazon's Mechanical Turk, to confirm these results with a larger sample size. In this study we control for the contextual information presented along with each warning dialog. We are also using a set of original warning dialogs, and a set of redesigned dialogs obtained from the application of two sets of guidelines: a compilation of those found in literature, and the ones above, derived from our mental model research. We expect to obtain relevant insights about how *understanding* and *incentive to comply* vary depending on context and warning design.

5. REFERENCES

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

[2] J. S. Downs, W. B. de Bruin, and B. Fischhoff. Parents' vaccination comprehension and decisions. *Vaccine*, 26(12):1595 – 1607, 2008.

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

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

[5] G. M. Morgan, B. Fischhoff, A. Bostrom, and C. J. Atman. *Risk Communication: A Mental Models Approach*. Cambridge University Press, 2001.

[6] C. Nodder. Users and trust: a microsoft case study. In L. F. Cranor and S. L. Garfinkel, editors, *Security and Usability: Designing secure systems that people can use*, Theory in practice, chapter 29, pages 589–606. O'Reilly Media, Inc., Sebastopol, CA, USA, first edition, 2005.

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

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

[9] M. S. Wogalter. Purposes and scope of warnings. In M. S. Wogalter, editor, *Handbook of warnings*, Human Factors and Ergonomics, chapter 1, pages 3–9. Lawrence Erlbaum Associates, Mahwah, New Jersey, first edition, 2006.

[10] M. Wu, R. C. Miller, and S. L. Garfinkel. Do security toolbars actually prevent phishing attacks? In R. E. Grinter, T. Rodden, P. M. Aoki, E. Cutrell, R. Jeffries, and G. M. Olson, editors, *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 601–610. ACM, Apr. 2006.