A Brick Wall, a Locked Door, and a Bandit: A Physical Security Metaphor For Firewall Warnings

Fahimeh Raja
University of British Columbia
Vancouver, Canada V6T 1Z4
fahimehr@ece.ubc.ca

Kirstie Hawkey
Dalhousie University
Halifax, Canada B3H 4R2
Hawkey@cs.dal.ca

Steven Hsu
University of British Columbia
Vancouver, Canada V6T 1Z4
h.steven@alumni.ubc.ca

Kai-Le Clement Wang
University of British Columbia
Vancouver, Canada V6T 1Z4
w.kaile@alumni.ubc.ca

Konstantin Beznosov
University of British Columbia
Vancouver, Canada V6T 1Z4
Beznosov@ece.ubc.ca

ABSTRACT

We used an iterative process to design firewall warnings in which the functionality of a personal firewall is visualized based on a physical security metaphor. We performed a study to determine the degree to which our proposed warnings are understandable for users, and the degree to which they convey the risks and encourage safe behavior as compared to text warnings based on those from a popular personal firewall. The evaluation results show that our warnings facilitate the comprehension of warning information, better communicate the risk, and increase the likelihood of safe behavior. Moreover, they provide participants with a better understanding of both the functionality of a personal firewall and the consequences of their actions.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Evaluation/Methodology*; D.4.6 [**Software**]: Security and Protection—*Information flow controls*

General Terms

Human Factors, Security

Keywords

Usable security, personal firewall, warning, mental model, physical security metaphor

1. INTRODUCTION

A personal firewall is security software that checks the traffic flowing between a *personal computer* and the network(s); and based on its configuration, it allows or blocks the traffic elements. Intended to be used by non-experts, personal firewalls are becoming commonplace; they are recognized as "the first line of defense" for personal computers [20, 40]. However, the protection provided by them depends strongly on their correct configuration [18]. Therefore, the usability of personal firewalls is key to their effectiveness.

Our prior usability study of the Microsoft Windows Vista Firewall [33] revealed that the lack of an accurate mental model about the firewall's system model was one of the root causes of errors when configuring the firewall. Our next study examined users' knowledge and perceptions of personal firewalls and their experiences with them [34]. One interesting finding was that most participants' interactions with their personal firewalls was limited to responding to firewall warnings that ask them to allow or block a connection. Therefore, a correct response to these warnings is key to the correct configuration, and thus, the effectiveness of personal firewalls. However, we also found that users do not make informed decisions in response to the warnings. Therefore, it is crucial to design firewall warnings that are understandable for users and that properly communicate the risk to them so that they can make informed decisions.

There is evidence that the communication of risks to home computer users has been unsuccessful in the field of computer security. Prior research has found that users do not heed risk communications [36, 26], that they do not read security warning texts [14, 41], and that they ignore them [14, 41, 34]. Several reasons are identified for this behavior [7]: users do not understand the messages [13, 34, 26]; they do not understand the options provided to them for responding to the warning [14]; and they are unaware of the risks [13], underestimate the risks [28, 34], or have an incorrect mental model of the risks [47, 14].

Risk communication should convey the basic facts relevant to the warning recipient's decision [3]. In the warning science literature, one successful technique for characterizing and designing risk communication is to employ the mental models approach, which is a decision-analytic framework [25]. With this approach, the design of risk communication is based on the recipients' mental model(s). The goal of the framework is to help people make decisions by providing risk communication that improves the recipients' mental models in one of three ways: 1) adding missing knowledge, 2) restructuring the person's knowledge when it inappropriately focussed (i.e., too general or too narrow), and 3) removing misconceptions [6].

Symposium on Usable Privacy and Security (SOUPS) 2011, July 20–22, 2011, Pittsburgh, PA USA

The mental models approach has been successfully applied in such areas as medical [21] and environmental [35] risk communications, but not in computer security. Risk communications in computer security have been based on experts' mental models, which are not good models for typical users. An expert's mental model of security is different from that of a non-expert [2, 8]. This difference could lead to ineffective risk communications to non-experts. Similarly, Asgharpour et al. [2] proposed that risk communication methods such as security warnings should be designed based on non-expert mental models and metaphors from the real world, emphasizing that:

> *the purpose of risk communication is not conveying the perfect truth, but rather prompting the users to take an appropriate action to defend their system against a certain threat. While mitigation of a risk requires knowledge of the general nature of the risk, efficacy of the risk communication requires communication that is aligned with the mental model of the target group* [2, p. 368].

While employing a mental models approach has been previously proposed for computer security warnings [2], it was not evaluated; the goal of our research is to do so. We first needed to consider which metaphor of security would best align with the end user's mental model(s) of security and would be most appropriate to deploy within the firewall warning's interface to convey the risks to them. Interviews by Wash [43] revealed that one of the most common mental models of security is the physical security and burglar mental model. Liu et al. [23] also performed a quantitative analysis to evaluate the five mental models proposed by Camp [8] for computer security: physical security, medical infections, criminal behavior, warefare, and economic failure. Their results show that for 70% of risks, non-expert users have physical and criminal mental models; for firewalls, the physical mental model was the closest to both expert and non-expert users' mental models. This suggests that the physical security metaphor could be appropriate for risk communication to non-expert users in computer security, particularly for firewalls.

In this paper, we present our iterative design of a firewall warning using a physical security metaphor, and we present our study of the effectiveness of this approach. In the warnings, the functionality of a personal firewall is visualized based on a physical security metaphor, which includes the metaphor of a firewall, a fireproof wall that "separates the parts of a building most likely to have a fire from the rest of a structure" [19, p. 2]. The goals of our study were to determine the degree to which the warnings are understandable for our participants and the degree to which they convey the risks and encourage safe behavior. We used an open-ended test to evaluate the initial clarity of the warnings, and we used Likert-type scales, followed by an interview, to evaluate participants' risk perceptions. We also used the self-reported likelihood of choosing any action as the intention for performing that action.

We compared our warnings with warnings based on those from the Comodo personal firewall. The Comodo firewall is the most popular personal firewall [32, 42], and is the top one in online reviews not only for its protection [42, 31, 10, 16, 30], but also for its "warning features that make it easy for novices to understand how to respond to those warnings" [42]. Our results show that our proposed warnings facilitate comprehension of warning information. They also better communicated the risk; with our warnings, participants had a better estimation of the level of hazard, likelihood of damage or loss, and the severity of potential damage or loss. Participants could also better describe the potential consequences of their intended actions. More importantly, our warnings increased the likelihood of safe behavior in response to the warnings. These findings suggest that our use of a physical security metaphor in the warnings has altered the participants mental model(s) of the functionality of a personal firewall as it relates to their security and risk. Our warnings were also preferred by the majority of participants.

It should be noted, that in our approach we attempted to influence the development of a users' mental model(s) through visual metaphors. We did not employ the standard HCI formal conceptual design approach as proposed by Norman [29] and employed in the design of various systems (e.g., [9]). We did select a plausible user mental model of security as a starting point based on related work on end user mental models of security [8, 23, 43] and our own prior research investigating personal firewall use and related usability issues [33, 34]. As we developed the visual metaphor to employ in the warning, we used an iterative user-centric process to refine it; however, we did not conduct a systematic analytical evaluation of the candidate mental models and their representations as described in Anderson et al. [1]. This process is recommended in order to ensure that the model can cover a sufficient number of features, that those not covered can be done so by some way of extension, and that the model does not have too much conceptual baggage.

Next, we present related work. In Sections 3 and 4, we present our prototype design and study methodology. Our results and a discussion of our findings are presented in Sections 5 and 6. We conclude in Section 7 with a summary of our contributions and directions for future work.

2. RELATED WORK

In this section, we first present prior research on usability of personal firewalls, and then on security warnings.

2.1 Usability of Personal firewalls

Prior research has considered the usability of personal firewalls. Johnston et al. [20] performed a heuristic evaluation of the Windows XP firewall and proposed improving the visibility of the system features, the system status, and the learnability of the interface. Berson [4] described design principles used by ZoneAlarm labs to create personal firewalls that are usable for consumers. He advocated human-centered design and offered guidelines to increase the usability of security products, including personal firewalls. He recommended that the designers of security products know the audience of their products, speak their language, and eliminate complexity. Herzog and Shahmehri [18] defined use and misuse cases for personal firewalls and performed a cognitive walk through of 13 personal firewalls to examine the behavior of the firewalls for those scenarios. Their results highlight the need to convey the firewall design model to users. While these studies are informative, they are based on evaluation by experts and the findings were not validated through studies with target users of those firewalls.

Hazari [17] performed an exploratory study of users' perceptions of the factors that could affect the selection of a

Figure 1: Comodo's original warnings; S: Safe, U: Unrecognized, M: Malicious.

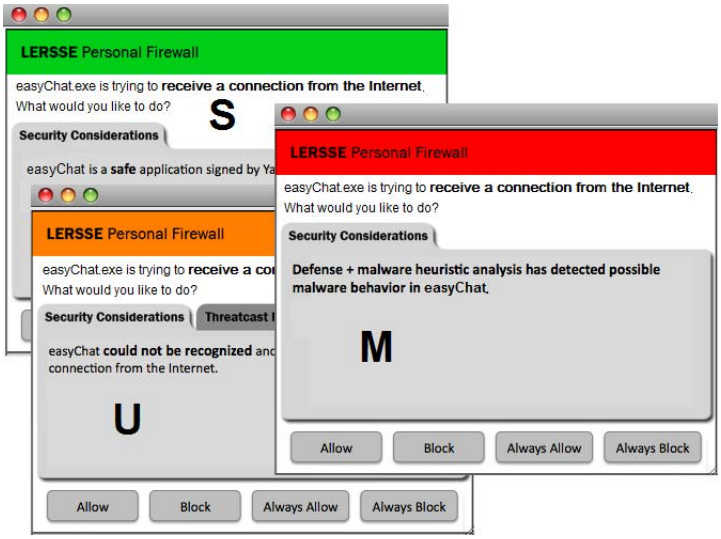

Figure 2: Warnings designed based on Comodo's warnings; S: Safe, U: Unrecognized, M: Malicious.

personal firewall in an organization. His Q-sort analysis [38] showed that ease-of-use is of high priority for users, but he did not describe what users meant by ease-of-use. Stoll et al. [40] used a spatial extension of the desktop metaphor to visually show system-level information for a personal firewall-like tool. Their goal was to present technical information in an understandable way so that non-expert users could make informed decisions. They performed a user study to evaluate the usability of their proposed approach and found that their participants made fewer mistakes in the configuration of their interface than with traditional firewalls.

In our earlier work [33], we performed a usability analysis of the Vista firewall. The results of our laboratory study showed that revealing the hidden network context helps users develop a more complete mental model of the firewall and a better understanding of its configuration. We performed another study to examine users' knowledge, perceptions, and experience of personal firewalls [34]. One interesting finding of this study was that the interaction of most of the participants with their personal firewall was limited to responding to firewall warnings that ask them to allow or block a connection. Very rarely did participants encounter the main interface of their personal firewall. Therefore, it is important to design firewall warnings that are understandable for users and to properly communicate risk to them.

2.2 Security Warnings

We are unaware of any related work specifically investigating the effectiveness of personal firewall warnings. However, there has been a great deal of security warning research.

Zurko et al. [49] evaluated the usability of Lotus Notes security warnings. They found that 59% of their participants allowed unsigned content to run due to a lack of awareness of the risks. Based on the results, they suggest educating users or including more information in security-related warnings.

Downs et al. [13] performed interviews with 20 non-experts to understand their reactions when encountering phishing sites. Their participants had little awareness of phishing, and the security warnings were not meaningful for many of them. They suggest describing the intuition behind recommended actions in a non-technical way.

Cranor [11] proposed the human-in-the-loop framework that provides a systematic approach for identifying reasons for human failures in interaction with security applications. Her framework can be used to find out why a particular security warning is ineffective.

Egelman et al. [14] performed a lab study to examine the effectiveness of active and passive phishing warnings. They found that active warnings are more effective than passive ones. They suggest that security warnings should interrupt the users' primary task, convey the recommended action clearly, fail safely if the user ignores or misunderstands the warning, and prevent habituation.

Sunshine et al. [41] performed a survey of around 400 Internet users to investigate their understanding of, and reactions to, SSL warnings. Their survey shows that risk perception is correlated with decisions to obey or ignore security warnings and that the participants who understood security warnings perceived a different level of risk associated with the warnings. They designed two new warnings and compared them with three existing SSL warnings in a lab study with 100 students. Based on the results, they suggest improving the design of warnings by using appropriate colors and text, and by decreasing warning frequency.

Motiee et al. [26] performed a lab study to investigate the motives, understanding, behaviour, and challenges users face when responding to User Account Control (UAC) prompts in Windows Vista and Windows 7. They found that 69% of the participants do not use the UAC approach correctly. They suggested using educational prompts to convey the purpose of UAC, as well as reducing the number of prompts that occur when a user initiates an action.

While prior research informed our study of firewall warnings, none evaluated the effectiveness of the mental model approach for computer security warnings. The goal of this paper is to mitigate this gap.

3. PROTOTYPE INTERFACE DESIGN

We designed two sets of firewall warnings. The first is our proposed one, P-warnings, which is based on a graphical depiction of physical security and firewall metaphors to facilitate risk communication to end users through the support of their physical security mental model. The second is one based on the Comodo personal firewall warnings, which we call C-warnings.

The Comodo personal firewall includes a database, which is a classification of many known executable files. It categorizes different applications in three categories based on the level of risk: safe (window with "S" in it in Fig. 1, or just Fig. 1.S, for short), unrecognized (Fig. 1.U), or malicious (Fig. 1.M). Using this database, the firewall provides "security considerations" in its warnings to help users make informed decisions in response to the warnings. This is in line with recommendations for designing security warnings [39, 14]. Based on this classification we designed six different interfaces, three for P-warnings: P-safe (Fig. 3.S), P-unrecognized (Fig. 3.U), P-malicious (Fig. 3.M), and three for C-warnings: C-safe (Fig. 2.S), C-unrecognized (Fig. 2.U), C-malicious (Fig. 2.M).

To design C-warnings, we made several changes in Comodo's warnings. As security warnings should be jargon-free [11], we removed three pieces of technical information (i.e., protocol (TCP or UDP), remote IP address, and port), as well as the "More Options" link from Comodo's warnings. We also removed the recommended action from the warnings (although it is recommended in the usable security literature [39, 14]) in order to eliminate a potentially confounding effect of this parameter on users' intentions for the purposes of our study. Comodo's warnings also show how other people who are using the Comodo personal firewall have responded to the warning for a specific unrecognized application. Prior research [12, 5] shows that relying on a community consensus can be effective for users without expertise to make an appropriate security decision. We kept this feature in our warnings; however, in a similar fashion to Comodo's design, it was not in the first window of the warnings, but hidden under a second tab called "Threatcast Rating." We did not want this feature to have a confounding effect on participants' risk perception and their intended action.

For P-warnings, we mimicked Comodo's layout. We applied an iterative process for the design of the warnings. Through Internet searches and by asking friends and family, we determined common metaphors for physical security, including locks, keys, doors, walls, money safes, policemen, and stop signs. We then performed a series of two formative studies. The first one was performed with 10 participants on paper prototypes of the warnings. In this study, we had three different designs for conveying the functionality of a firewall through a physical security metaphor (See Appendix A). We also had several icons for each action (i.e., Allow, Block, or Remember my answer). Based on the participants' feedback, we selected the most appropriate design, and redesigned the warnings (See Appendix B). We performed another formative study with 15 participants; and then based on the findings of the second study, we finalized the design of the warnings. We hired a graphical interface designer to develop the interfaces for us. This ensured consistency in presentation of both P-warnings and C-warnings.

In the final P-warnings we used a brick wall and a metal door to resemble a physical firewall and a fire exit, which are the actual metaphors for a computer firewall. We added a lock on the door which was recognized as the most familiar and understandable metaphor for controlling access in physical security by our formative study participants. We put a computer behind the wall to show that the wall is preventing access to the computer. For the actions (Allow, Block, Always Allow, Always Block),[1] we added icons corresponding to the lock on the metal door as follows. For Allow, the lock is unlocked to depict a single access through the door; for Block, the lock is locked to depict denied access; for Always Allow, the key is shown to depict granting permanent access; and for Always Block, the lock does not have a keyhole to depict permanently blocking access.

We used a figure of a person who wants to go through the door to represent the application, which wants to make a connection through the firewall. We also used a speech bubble to show the name of the application and its developer. To indicate different levels of security risk, we used different colors and different figures with different speech bubble quotes. For the safe interface, we used a figure wearing a green shirt with a smile on his face, and with a quote representing the application developer (i.e., Yahoo! in Fig. 3.S). Our formative study revealed that this figure is friendly, trustworthy, and gives a positive impression, thereby encouraging the user to unlock the lock and grant access. For the unrecognized interface, we used a black silhouette of the figure with a question mark as its head to show that the application is unidentified (Fig. 3.U). For the malicious interface, we used a figure dressed in a prisoner's uniform, carrying a knife and a thief's bag (Fig. 3.M). According to our formative study participants, "the message that the interface conveys is very clear: It is very dangerous!"

4. METHODOLOGY

As suggested by the human-in-the-loop framework [11], there are three steps in processing warning information: (1) communication delivery (attention switch and maintenance), (2) communication processing (comprehension and knowledge acquisition), and (3) application (knowledge retention and transfer). The focus of our research is on step (2), communication processing.[2] It is an important step; if the user does not understand the warning, he can not make an informed decision in response to that warning.

Our research questions in this study were:

- Do our participants understand what the warnings mean when they encounter them for the first time?
- What are the participants' misunderstandings or sources of confusion about the warnings?
- Are our proposed warnings more understandable for participants than those based on Comodo's warnings?

[1]Our formative study revealed that participants preferred "Always Allow" and "Always Block" to "Remember my answer."

[2]We are not concerned with communication delivery, because firewall warnings are "active warnings," that is, they interrupt users' primary tasks and force them to pay attention to the warnings and decide whether to allow or block the connection before they can proceed with their primary task [11].

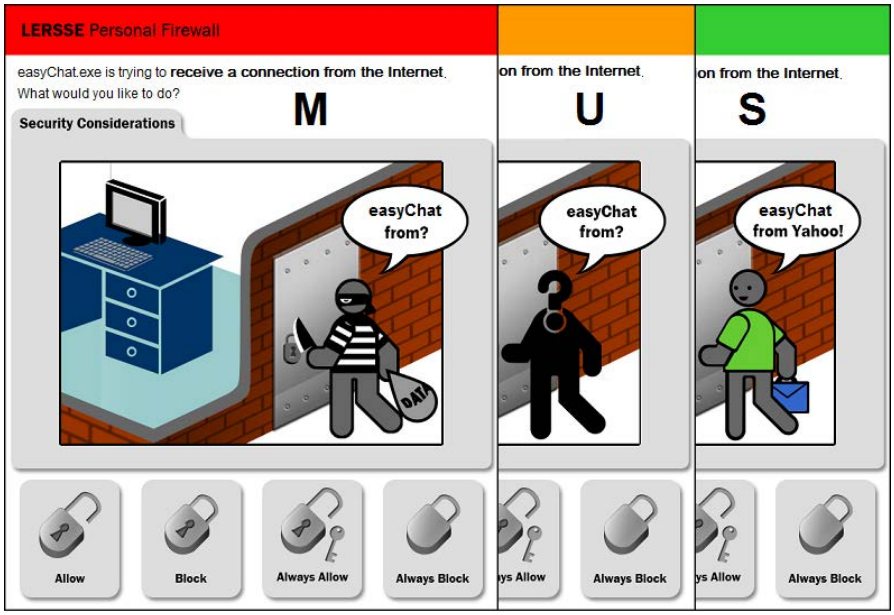

Figure 3: Warnings based on a physical security metaphor; S: Safe, U: Unrecognized, M: Malicious.

- How do the warnings affect the participants' intention to act?
- Which kind of warnings would the participants prefer to have for their personal firewall? Why?

To answer these questions we performed a user study. There are two commonly applied approaches in warning science literature for evaluating comprehension and the degree of initial clarity of the warning: open-ended and multiple choice tests [48, 22]. As recommended by Leonard [22], we used an open-ended test because it provides more information about sources of confusion and the types of errors that people make [46]. This information might assist in any subsequent redesign work. One issue with open-ended tests is that the evaluation of the responses is less clear-cut and, thus, more difficult. There is usually some subjective judgment about the correctness of open-ended responses; therefore, the responses should be evaluated by more than one evaluator [48, 22]. In our study, we used three evaluators to increase the reliability of our evaluations.

As it is common in warning science, we used "risk perception" to measure intention. Risk perception is considered to be the most important factor of intention [48]; and it is defined as the "perceived chance of injury, damage, or loss" [24]. There are multiple known contributing variables to risk perception, including the hazardousness of the situation, the likelihood of damage or loss, and the severity of potential damage or loss [48]. We used the most common approach for evaluating warnings on each dimension, which is using Likert-type scales ranging from 0 to 7, followed by an interview for clarifications [48, 45]. We also used the self-reported likelihood of choosing any action as the behavioral intention for performing that action.

It should be noted that intention is not actual behavior. Evaluating users' behavior in response to warnings requires the study to be conducted in a real context [48]. That is not always possible in the case of security warnings [15], and especially firewall warnings. Direct observation of users' behavior in response to firewall warnings is time and labor consuming because users' interaction with these warnings is infrequent and sporadic [34]. Laboratory studies also may not be generalizable to real world situations; this is because providing a believable risk situation, which is actually safe, is very challenging [37]. On the other hand, allowing insecure situations to occur in laboratory studies to simulate the real context has serious ethical concerns [48]. Given our focus on the communication processing phase of warning processing, we did not design our study to evaluate users' behavior in response to the warnings, but rather to examine whether or not the changes that we propose for firewall warnings will increase the ability of participants to understand them. We also wanted to determine the degree to which our proposed warnings convey the risk and might impact participants' intended actions, as compared to current firewall warnings.

4.1 Study Design

We performed a within-subjects study to compare our proposed warnings based on a physical security metaphor (P-warnings) with the ones designed based on Comodo's warnings (C-warnings). We felt this was more appropriate than a between-subjects design, because we had concerns about controlling the individual differences between our participants that might impact the results. A between-subjects design might allow a more precise comparison of each pair of interfaces; however, as discussed by Cranor [11], each individual "brings to the situation a set of personal variables, intentions, and capabilities" that impact the warning information processing.

To reduce the learning effects and presentation sequence effects introduced by a within-subjects design, we counterbalanced the presentation order of the warnings. We had two conditions; in the first condition (P-C) all participants saw P-warnings first, while in the second condition (C-P), they saw C-warnings first. We also counterbalanced the presen-

P-C	C-P
PsCs – PuCu – PmCm	CsPs – CuPu – CmPm
PsCs – PmCm – PuCu	CsPs – CmPm – CmPm
PuCu – PsCs – PmCm	CuPu – CsPs – CmPm
PuCu – PmCm – PsCs	CuPu – CmPm – CsPs
PmCm – PsCs – PuCu	CmPm – CsPs – CuPu
PmCm – PuCu – PsCs	CmPm – CuPu – CsPs

Table 1: Presentation order of the warnings. P: our proposed warnings based on a Physical security metaphor, C: warnings designed based on Comodo's warnings, s: safe, u: unrecognized, m: malicious.

tation order of safe, unrecognized, and malicious interfaces within each condition.

In our initial study design, we presented all safe, unrecognized, and malicious interfaces of one warning design (P or C) before the other one (C or P). However, during our pilot study, we realized that presenting one complete set of warnings (P or C) to the participants primes them about the existence of three levels of risk for the other set (C or P). Thus, they were more careful when specifying their perceived risk level for the second set of warnings. Moreover, the only difference (if any) they would express between safe, unrecognized, and malicious interfaces in each set was the level of risk. Seeing one interface from a specific set did not affect their understanding of the later interfaces from the same set; they maintained their understanding of one interface when they saw the other two in the same set of warnings.

We modified our study design in order to reduce the bias from risk priming that we observed in the pilot stduy. We showed one interface (safe, unrecognized, or malicious) from one set (P or C) and then showed the corresponding interface (safe, unrecognized, or malicious) from the other set (C or P). We examined the participants' understanding for the first interface they saw from each set. Based on this revised study design, we had 12 presentation orders of the interfaces (see Table 1.[3]) We randomly assigned participants to each order.

4.2 Study Protocol

Each participant completed a one-hour session in a meeting room in our department. There were three researchers in each session of the study. We used voice recording software to augment the researchers' notes. In the study, the participants first completed a consent form and background questionnaire, which included an assessment of their security knowledge and experience.

We assessed the participants' security knowledge and experience with the following six tasks taken from the "Security Center" of Windows Vista [44]: (1) installing updates; (2) scanning for viruses, spyware, and other potentially unwanted software; (3) changing the security settings of web browsers; (4) deleting browsing history and cookies; (5) setting different security controls for different users; and (6) managing browser plug-ins. We chose these tasks because they are common security tasks that a home computer user might perform on any operating system and with any web browser. We asked the participants to describe what they knew about the tasks and their importance, and to specify how often they performed those tasks. To determine whether participants had any special computer security knowledge or expertise, we asked them if they had attended any computer security course, workshop, or conference, or have had any computer security-related jobs.

We then described a scenario to participants. This is recommended in the warning science literature to provide a context for participants when examining their understanding and comprehension of warnings [48]. We asked them to imagine that they had urgently needed to use a chat application that gives them the ability to video conference with four other people in different locations. For this purpose, they had done research on the Internet and found an application called easyChat. They downloaded and installed the application and sent it to the others so that they could begin the video conference. But, when they wanted to use the application, they received a warning from their security software. We used this scenario because chat application traffic must be allowed through personal firewalls in order to connect to the Internet, and also because chat applications are commonly used by home users of computers. We then presented the participants with the interfaces using one of the orders in Table 1.

After presenting each interface, we had a brief interview with the participants to examine their understanding and also possible misunderstandings of the interface and its elements. We asked them what their reaction would be to the warning. Then we asked them to fill out a questionnaire and specify their perceived level of hazard, likelihood of damage or loss, and severity of potential damage or loss on a scale of 0 to 7. If any response was greater than 0, we asked the participant to describe the potential hazards. We then asked the participants to specify the probability with which they would choose one of the options provided for them to respond to the warning (Allow, Block, Always Allow, Always Block). We then had a brief interview with them about the reasoning for their answers. We repeated the same procedure for the other interfaces. At the end of the experimental session, we opened all the interfaces and asked the participants about their preferred warning design and their reasoning. This study was approved as a minimal risk study by the University of British Columbia's Research Ethics Board.

4.3 Participants

We recruited 61 participants from both the university and the general community. We excluded one participant because of his inconsistent responses about his demographics in the screening email and in the questionnaire used during the study. To recruit participants, we sent out messages to email lists of several departments in the university, including Computer Science, Electrical and Computer Engineering, Mining Engineering, History, and Psychology. We also posted messages to our university's online classified advertisements, as well as two public online classified advertisement sites, Craigslist and Kijiji. Moreover, we posted and handed out flyers at the university and local public places, such as public libraries. To ensure diversity, we screened interested participants by email. We asked their age, gender, last educational degree and major, whether or not they

[3]The ideal study design is to randomly order the 6 possible conditions; however, this design results in 6!=720 orders, which needs at least 720 participants to be fully counterbalanced.

Condition		P-C	C-P	Total
Group Size (N)		30	30	60
Age	Mean	31.4	31.2	31.3
	Range	18-67	19-68	18-68
Gender	Female	16	14	30
	Male	14	16	30
Student	Yes	16	17	33
	No	14	13	27
Educational Level	High school	5	6	11
	Bachelor	13	15	28
	Master	10	7	17
	PhD	2	2	4
Background	Comp. Sci./Eng.	6	5	11
	Science	5	2	7
	Engineering	7	9	16
	Art	8	12	20
	Business	4	2	6
Primary OS	Windows XP	9	9	18
	Windows Vista	6	9	15
	Windows 7	7	6	13
	Mac	7	5	12
	Linux	1	1	2
Security Level	High	2	3	5
	Medium	8	9	17
	Low	20	18	38

Table 2: Participants' demographics.

were students, and their occupation (if not a student). All participants were given a $15 honorarium.

Table 2 shows the demographics of the participants. They had a wide range of occupations (e.g., professor, librarian, physician, diamond trader). All except one, who used a computer weekly, were daily users of computers, but their expertise varied. Over half (18/30 (P-C), 20/30 (C-P)) considered themselves regular or advanced users of basic programs (e.g., web browsers, email), while the rest considered themselves more advanced (e.g., able to configure operating systems). Based on the responses to our background questionnaire, we classified our participants' security knowledge and experience as high, medium, and low. The categorization was done to understand our participants' comprehension of the warnings in relation to their level of security knowledge and expertise. It should be noted that participants in group H are not security experts who practice security as their primary task (i.e., security practitioners), but their security knowledge and expertise is higher than the average computer users. None of our participants had attended any computer security courses, workshops, or conferences; none had been employed in any computer security-related jobs.

We should note that it is statistically likely that at least one of the participants had some form of colourblindness, which may have had an impact on their ability to discern the colours of our warnings. However, we did not collect any data about this, so cannot report whether this played any role in the understandability of our warnings.

4.4 Data Analysis

As mentioned previously, our analysis was conducted by three evaluators to help mitigate the subjective bias. We used a card sorting approach [27] to analyze the data. We first wrote our participants' responses to the interview questions on index cards. Then, we iteratively sorted the index cards for each question into multiple piles so that cards representing similar responses were in the same pile. We then associated a theme with each pile, that represented participants' understandings, misunderstandings, risk perceptions, and preferences for the warnings. Throughout the process, evaluators discussed the sorting of the cards and the theme associations until consensus was achieved.

5. RESULTS

Results include (1) our participants' understandings of the warnings, (2) the effect of the warnings on their risk perception and intended action in response to the warnings, and (3) their preferences for the warnings.

5.1 Warning Understanding

We report our participants' initial understanding of the warnings for the two first interfaces viewed, one P-warning and one C-warning. When we asked our participants to describe what they understood from the warnings, most of them started by repeating the text at the top of the warnings. Further assessment of their comments revealed that P-warnings made them more aware of the protection provided by the firewall.

With P-warnings, most of the participants (48/60 (80%): 23/30 in (P-C), 25/30 in (C-P)) said that the warning was generated by security software that was preventing access to their computer: "It is from your software, which is kind of a barrier" (P34). They also explained what would happen if they chose each of the options provided for them in response to the warnings. As P8 noted: "It is a door, and you control the lock. Nice ... so your computer is presumably safe in a locked up space. I think this is what your firewall does, now there is this new software [easyChat] which is trying to access the computer through the door, and you have the control of the lock, so you can either allow or block it. Cool ... I like it. It tells you the whole story."

With C-warnings, 34 (57%) participants talked about the prevention provided by the firewall. However, it should be noted that most of these participants had seen P-warnings first (23 in (P-C), 11 in (C-P)). They mentioned that the warning conveys a similar message as P-warnings and just emphasized different presentations of the message (17 participants), or different levels of risk that they convey (16 participants). From the remaining 26 participants, most of them (18) read the text in C-warnings and emphasized the terms written in bold. The rest (8) either had a misunderstanding about the warnings; or they said they did not understand the warnings, why they would get the warning, or what would happen if they clicked on *Allow* or *Block*.

Some participants misunderstood the warnings. For P-warnings, 9 participants (6 in (P-C), 3 in (C-P)) thought that the warning was generated by the chat application, asking them if they wanted to chat with someone from Yahoo!, someone unknown, or someone malicious (for safe, unrecognized, or malicious applications respectively). As P54 mentioned, "Somebody from Yahoo! is connecting to me, how does this guy from Yahoo know me?" Three participants also thought that the warning was about their wireless connectivity, asking them whether or not they wanted to have a connection to the Internet. For C-warnings, aside from those who did not understand what the warning message was and those who only read the text in the warnings, we had one participant who thought that the message was from easyChat. Another thought there was a problem in his Internet connection: "It says it needs to connect to the Internet, but it can't. Maybe I need to refresh my Internet, or turn the

router off and on" (P49). We also had one participant who thought that the warning was for the security of his Internet connection, "Now if I allow it, my Internet is safe, and I can check my emails securely."

In addition to the participants' understanding of the warnings as a whole, we assessed their understanding of each element of the warnings as well. For P-warnings, we found that the question mark on the top of the figure for unrecognized applications, and also the quote "from?" was not clear for 5 participants: 2 thought that the figure had a question, and that was why he wanted to connect to them; the other 3 stated that they did not know what that quote and question mark meant. For C-warnings, three technical terms were not understandable for some of the participants. Three showed confusion about the term "signed" in the warning for safe applications, as P24 mentioned: "It says it can steal your Yahoo! password to sign in to my Yahoo! ID." Two mentioned that they did not understand "could not be recognized" for unrecognized applications: "Unrecognized by whom?" Five participants also mentioned that they did now know what "malware" means.

5.2 Risk Perception and Intended Action

To examine if our within-subjects study design had affected our results, we performed a preliminary analysis of the data. To see if there was an effect of the presentation order of the warnings we conducted a two-way ANOVA with two between-subjects factors: warning type order (P-C, C-P) and threat order (SUM, SMU, USM, UMS, MSU, MUS). Our analysis did not reveal a significant main effect for either warning type order or threat order on participants' perceived level of hazard, likelihood of damage or loss, severity of potential damage or loss, or reported likelihood of allowing or blocking the program (in all cases, $p>.05$ for safe, unrecognized, and malicious applications). There were also no significant interactions between the two factors (in all cases, $p>.05$ for safe, unrecognized, and malicious applications).

Moreover, we compared our participants' perceived level of hazard, likelihood of damage or loss, severity of the potential damage or loss, and probability of allowing and blocking the program when they saw each interface first (before seeing any interface) and when they saw it last (after seeing all the interfaces). Our results showed no significant difference between the last exposure to each interface and the first exposure (in all the cases $p>.05$ for safe, unrecognized, and malicious applications). The lack of evidence that the presentation order of the warnings impacted our participants' consideration of later warnings suggests that our within-subjects study design was appropriate given our objectives, the diverse population that we wanted to study, and the feedback we received in our pilot testings.

We conducted a one-way repeated measures ANOVA to evaluate the effect of warning type (P-warnings or C-warnings) on the participants' risk perception and intended action. Next, we present our results for the warnings based on the level of risk they convey.

5.2.1 Warnings for Safe Applications

For safe applications, we did not find a significant main effect for warning type on participants' perceived level of hazard ($F(1,59)=.532$, $p>.1$, Partial Eta Squared=.009), likelihood of damage or loss ($F(1,59)=1.712$, $p>.1$, Partial Eta Squared=.028), severity of the potential damage or loss ($F(1,59)=3.445$, $p>.05$, Partial Eta Squared=.055), and their reported likelihood of allowing ($F(1,59)=1.112$, $p>.05$, Partial Eta Squared=.010) or blocking ($F(1,59)=.156$, $p>.1$, Partial Eta Squared=.003) the program. These results suggest no difference between P-warnings and C-warnings for safe applications with respect to participants' risk perception and intended action. For both warnings, their perceived level of risk was appropriately low and most of them were more likely to allow the program (See Table 3).

Our participants mentioned several factors that could affect their low level of risk perception and their intention to allow the safe application. For P-warnings, 33 participants pointed out that the interface does not have any element that conveys risk to them, as P25 said: "This particular image doesn't not give any signal for any danger." Twenty four participants mentioned that the appearance of the figure in the image is friendly; more specifically they noted that the smile on his face is a sign of safety: "A smile would definitely tell you that it's OK" (P54). Some participants (17) noted "Yahoo!" is a trusted corporation, thereby allowing the program; as P22 mentioned "easyChat is from Yahoo! this encourages us to relax." Twelve participants pointed to the green color as an indicator of safety: "the green really signals to go ahead." For C-warnings, the most important factors for most participants was the term "safe" (46), the name of "Yahoo!" corporation (37), and the green color (27).

5.2.2 Warnings for Unrecognized Applications

For unrecognized applications, the participants' risk perception was significantly higher for P-warnings than for C-warnings (See Table 3). Our analysis showed a significant main effect for warning type on participants' perceived level of hazard ($F(1,59)=7.792$, $p<.01$, Partial Eta Squared=.117), likelihood of damage or loss ($F(1,59)=10.014$, $p<.01$, Partial Eta Squared=.145), severity of the potential damage or loss ($F(1,59)=4.882$, $p<.05$, Partial Eta Squared=.076), and their reported likelihood of allowing ($F(1,59)=9.384$, $p<.01$, Partial Eta Squared=.137) the program. These results suggest that with P-warnings, the participants' risk perception was higher, and they were less likely to allow the unrecognized application.

Quotes from participants revealed that they might be more cautious when they encounter P-warnings than C-warnings for unrecognized applications. According to 36 participants, the question mark on the head of the figure in P-warnings is a sign of risk; as P42 mentioned, "the question mark says it all; who is it from? It is alarming, indicating I should be aware." Some participants (17) mentioned that the question mark made them "question the application;" P8 stated that "It is giving you a very good feeling of do you know what is sneaking upon you." Several participants (13) even noted that as the firewall cannot provide information about the application, they would do more research about it, such as searching in the Internet to find reviews about the application or asking friends, family, or others who they think would be more security knowledgeable: "It looks more like a warning. The application is not familiar to the firewall. I might go on the Internet and see if I can find something else; if not, and I am in urgent need, I will ask people who know more about computers than me" (P37). If they could not find more information they would prefer to block the program "to be safe than sorry, because it is something questionable. It is better to block it" (P57).

	Risk Perception	Safe				Unrecognized				Malicious			
		P-warning		C-warning		P-warning		C-warning		P-warning		C-warning	
		μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
	Level of Hazard	1.6	1.6	1.3	1.4	3.9	1.9	3.3	1.8	6.0	1.2	5.1	1.4
	Likelihood of loss	1.7	1.5	1.3	1.4	3.8	1.9	3.1	1.9	6.0	1.1	5.0	1.5
	Severity of loss	1.9	1.9	1.3	1.5	3.5	1.9	3.3	2.1	5.9	1.3	5.1	1.5
P-C	Allow	65.1	38.5	55.6	43.4	50.0	41.2	61.8	32.9	10.3	19.0	19.7	29.9
N=30	Block	9.2	20.3	9.7	21.2	20.9	30.5	26.7	28.5	35.7	41.1	43.5	38.6
	Always Allow	21.0	34.4	34.3	44.6	8.2	21.6	3.8	9.4	6.3	19.7	1.0	2.8
	Always Block	4.7	18.3	0.3	1.3	20.9	38.4	8.0	25.5	47.3	46.8	35.5	40.6
	Level of Hazard	1.4	1.4	1.4	1.5	3.6	1.7	3.0	1.8	6.3	0.9	4.4	1.7
	Likelihood of loss	1.3	1.3	1.2	1.3	3.5	1.5	2.9	1.6	6.1	0.8	4.2	1.7
	Severity of loss	1.8	2.0	1.5	1.6	3.6	1.8	3.0	1.9	6.0	0.9	4.3	1.7
C-P	Allow	66.1	37.0	50.8	37.8	50.0	33.8	63.6	36.3	12.4	22.4	30.0	32.4
N=30	Block	8.1	16.1	10.4	18.7	35.1	31.6	18.7	26.0	47.3	39.2	43.0	38.1
	Always Allow	22.3	36.7	36.2	42.6	2.8	7.2	6.7	19.4	1.2	4.7	4.2	15.0
	Always Block	3.4	9.9	2.7	6.8	12.1	27.0	10.9	27.1	39.1	41.2	22.8	34.6
	Level of Hazard	1.5	1.4	1.3	1.4	3.8	1.8	3.2	1.8	6.1	1.0	4.7	1.6
	Likelihood of loss	1.5	1.4	1.3	1.3	3.6	1.7	3.0	1.8	6.0	1.0	4.6	1.6
	Severity of loss	1.8	1.9	1.4	1.5	3.6	1.8	3.1	2.0	6.0	1.1	4.7	1.6
Total	Allow	65.6	37.4	53.2	40.4	50.0	37.3	62.7	34.4	11.4	20.6	24.8	31.3
N=60	Block	8.7	18.2	10.1	19.8	28.0	31.6	22.7	27.4	41.5	40.3	43.3	38.1
	Always Allow	21.7	35.3	35.3	43.3	5.5	16.2	5.3	15.2	3.8	14.4	2.6	10.8
	Always Block	4.0	14.6	1.5	5.0	16.5	33.2	9.4	26.1	43.2	43.9	29.2	38.0

Table 3: Participants' perceived level of hazard, likelihood of damage or loss, and severity of potential damage or loss on a scale of (0..7); and the probability of choosing Allow, Block, Always Allow, and Always Block. P: warnings based on a physical security metaphor, C: warnings based on the Comodo personal firewall warnings. Cells with grey background show statistically significant results.

5.2.3 *Warnings for Malicious Applications*

For malicious applications, warning type had a significant effect on participants' perceived level of hazard ($F(1,59)=35.998$, $p<.001$, Partial Eta Squared=.379), likelihood of damage or loss ($F(1,59)=37.735$, $p<.001$, Partial Eta Squared=.390), severity of the potential damage or loss ($F(1,59)=28.531$, $p<.001$, Partial Eta Squared=.326), and their reported likelihood of allowing ($F(1,59)=11.302$, $p<.01$, Partial Eta Squared=.161) the program. More interestingly, we found that type of warning had a significant main effect on the probability of always blocking a malicious program ($F(1,59)=8.031$, $p<.01$, Partial Eta Squared=.120). These results show that P-warnings convey more risks to the participants than do C-warnings.

According to 31 participants, P-warnings make the connection with theft in the real world. As P8 noted, "cognitively, I know it is not any different from the other one [C-warnings], but it just hints at you. If you let it in, it is like inviting a burglar in your house." This made them think about consequences of allowing the program (47 participants), such as allowing it to access their computer and steal their information. They found the bandit appearance of the figure and the knife in his hand to be very scary (23 participants); and they also found that it conveyed the risk well (43 participants). In addition to these factors, the red color in the interface contributed to the decision of 26 participants to block the application.

5.3 Warning Preference

Most of our participants (40) preferred P-warnings; some (13) noted P-warnings provide them with a mental model of the functionality of a firewall. As P60 commented, "this one (C-warning) is just a warning of a firewall. [The] brick wall and the locked door is very good. It tells me the theory of the firewall." Some also found P-warnings more intuitive (16), easier (37), and faster (9) to understand. P38 mentioned when "multitasking, [such as] talking to your friends, this [P-warnings] is very effective. It tells you everything at a glance. You make less mistakes." P18 also said, "it just takes my time to read it [C-warnings], and at the end I do not know, OK, what? But I understand this image instantly. I do not need to concentrate on every single word that is written." Some (11) emphasized that P-warnings convey the risk and the consequences of allowing the program more clearly. As P32 said, "The top one [P-Warnings] is clear, it tells you what the risk is, and if it's acceptable, if there's a threat." P51 also noted: "I like that one (P-warnings), especially that bandit's outfit, trying to steal something, my data, cannot get any more clear than that."

In addition, several participants (12) thought P-warnings grabbed their attention better, while 8 mentioned that it is easier to ignore C-warnings. As P18 said, "You do not decide to see or not to see this [P-warnings], but you can choose to read or not to read this one [C-warnings].". Five participants also stated that P-Warnings are more universal, as P19 commented: "This one [P-Warning] is better, especially for old people that cannot see clearly or children that may not understand security, or those who do not know what a firewall is."

Only one third of the participants (20) preferred the C-warnings over our graphical P-warnings. Most of them thought C-warnings were more professional (11) and that they would take them more seriously (4): "It [P-warnings] is kind of childish" (P13). They also found C-warnings more informa-

tive (7) and descriptive (2). Five noted they would understand C-warnings better. Two also mentioned C-warnings are more specific; they thought different people would have different interpretations for P-warnings. P2 also noted that a good feature of C-warnings is that they are text-based, which means that users can search for more details about the warning and the risks on the Internet by directly copying and pasting the text. P20, who had a high level of security knowledge, mentioned: "I already knew how a firewall works, so I don't need this image, but it is definitely helpful for those who know less."

6. DISCUSSION

Our findings suggest that supporting a mental model of physical security in firewall warnings through the employment of a physical security metaphor graphically displayed in the warnings could be a promising approach for improving users' information processing of the warnings, as well as their perceived level of risk and intention for safe behavior. For most participants, our proposed warnings promoted a better understanding of the protection provided by a personal firewall. However, we noticed some misinterpretations of the warnings, as well. Some of the participants thought that the warnings were generated by their chat application because the scenario that we described for them was about using a chat application. This shows that context also plays a role in users' understanding of the warnings. Therefore, context of use should be considered in the design of the warnings; care also should be taken about which metaphors to use for varying contexts.

A second implication of the role of context in the interpretation of the pictorial warnings is for the design of user studies. When evaluating the warnings, the different usage contexts of the warnings, and their possible impacts on the results, should be considered in the study design. It is important to interpret the results based on the context in which participants consider themselves to be in. Considering different contexts of use in our study might have provided us with different misinterpretations of our warnings.

Our results also show that the application of known metaphors, such as the bandit figure, in the warnings are very effective in conveying risk to the user. Our participants could relate the potential risks of the warning to the risks from the physical world; this resulted in them better understanding the consequences of their potential actions. However, we had participants who mentioned they would take our warnings less seriously than the textual warnings. They attributed this to their personality, and they did not considering pictorial representations to be a professional method of risk communication for security warnings.

Another interesting finding of our study was that one third of our participants preferred C-warnings to our proposed warnings. A quick look at their demographics showed that all but one of these participants had a high or medium level of security knowledge and expertise. Furthermore, all the participants with a high level of security knowledge and expertise were in that group. These results indicate that the design of warnings may need to be customizable for different groups of users with different demographics, e.g., their preferences can be obtained at the installation time.

It should be also noted that our warning design was just one of the many possible designs. For example, we evaluated just one possible presentation order of actions (Allow, Block, Always allow, and Always block). Moreover, the relation between the participants demographics and the suitability of the choice of metaphors is not clear cut. Ordinary people might not even use or understand the term "firewall" when it comes to computers, despite our use of the firewall metaphor in the warnings. As a result, there might be a clash for some users between the physical model of burglary and the name of the tool, a"firewall." As well, there could be issues relating to the real nature of the firewall tool, such as its role of restricting outgoing connections as well as incoming ones. The physical security metaphor that we depicted did not address traffic leaving the computer. Further research is required to examine what kinds of metaphors ar appropriate for different user groups and to encompass the full activity of the personal firewall. Our study was a first step in evaluating the effectiveness of using real world metaphors for conveying risk in security warnings.

7. CONCLUSION AND FUTURE WORK

We presented a user study in which we evaluated a novel approach for designing personal firewall warnings. We used an iterative process to visualize the functionality of a personal firewall based on a metaphor of physical security, which included the depiction of a physical firewall. We compared our warnings with textual ones, which were based on the warnings of one of the most popular personal firewalls; these did not employ any metaphors.

Our results show that our proposed warnings facilitate comprehension of warning information. They also better communicated the risk; with our warnings, participants had a better estimation of the level of hazard, likelihood of damage or loss, and severity of potential damage or loss. They could also better describe the potential consequences of their intended actions. More importantly, our warnings increased the likelihood of their safe behavior in response to the warnings. These findings suggest that our use of a physical security metaphor in the warnings has altered the participants mental model(s) of the functionality of a personal firewall as it relates to their security risks. Our proposed warnings were also preferred by the majority of participants.

While the focus of this research is on firewall warnings, we believe that our research will benefit the broader community of usable security, especially those who are working on security warnings.

Further research is required to examine the appropriateness of different metaphors for different groups of users with different demographics. As one reviewer suggested, it might be useful to include an 'Einstein" or 'guru" icon, or perhaps a "technical report" tab, to give the full details for those who want to know or learn more. Of course, this could add clutter or confusion; and the impact of this must be investigated in future user studies. Having the information always available could be beneficial as (1) there is only one warning format for the end user to learn, support, and understand; and (2) the warning would suffice if the users of a computer vary (e.g., a naive user and their local tech support).

Moreover, user studies need be performed to analyze the actual behavior in response to our proposed warnings in a real context. Instead of implementing our design in a working fashion, we evaluated our approach with mock-ups. Furthermore, we asked the users to reflect on the firewall warnings, thus causing the users to think much harder about their meanings than would normally be done in situ. Normally,

when a personal firewall pop-up warning appears, it interrupts something the user actually cares about, and the user tries to think as little as possible about the pop-up. Our user study did not recreate this type of environment, and thus potentially lacks external validity. Also, it cannot answer the question whether the new warning designs require more mental processing than traditional pop-ups. However, our approach did allow us to focus on the comprehension of the warnings and the impact of the physical security metaphor on participants' mental models, which was an appropriate and important first step.

Acknowledgements

We thank the study participants for their time, and the members of the Laboratory for Education and Research in Secure Systems Engineering (LERSSE) at UBC, who provided valuable feedback on the earlier drafts of this paper. Cormac Herley provided feedback in May 2010 on the design of the project. We also thank Angela Sasse for shepherding this paper; along with the anonymous reviewers and Diana Paterson, she provided valuable insights that helped us to improve the final version of this paper. This research has been partially supported by the Canadian NSERC ISSNet Internetworked Systems Security Network Program.

8. REFERENCES

[1] B. Anderson, M. Smyth, R. Knott, M. Bergan, J. Bergan, and J. Alty. Minimising conceptual baggage: Making choices about metaphor. In *People and Computers IX - Proceedings of HCI'94*, pages 179–194, 1994.

[2] F. Asgharpour, D. Liu, and L. J. Camp. Mental models of security risks. In *FC'07/USEC'07: Proceedings of the 11th International Conference on Financial Cryptography and 1st International Conference on Usable Security*, pages 367–377, Berlin, Heidelberg, 2007. Springer-Verlag.

[3] C. J. Atman, A. Bostrom, B. Fischhoff, and M. G. Morgan. Designing risk communications: Completing and correcting mental models of hazardous processes, part i. *Risk Analysis*, 14(5):779–788, 1994.

[4] J. Berson. ZoneAlarm: Creating usable security products for consumers. In L. F. Cranor and S. Garfinkel, editors, *Security and Usability: Designing Secure Systems that People Can Use*, chapter 27, pages 563–575. O'Reilly Media, Inc., 2005.

[5] A. Besmer, J. Watson, and H. R. Lipford. The impact of social navigation on privacy policy configuration. In *SOUPS '10: Proceedings of the Sixth Symposium on Usable Privacy and Security*, pages 1–10, New York, NY, USA, 2010. ACM.

[6] A. Bostrom, C. J. Atman, B. Fischhoff, and M. G. Morgan. Evaluating risk communications: Completing and correcting mental models of hazardous processes, part ii. *Risk Analysis*, 14(5):789–798, 1994.

[7] C. Bravo-Lillo, L. Cranor, J. Downs, and S. Komanduri. Poster: What is still wrong with security warnings: a mental models approach. In *SOUPS '10: Proceedings of the 6th Symposium on Usable Privacy and Security*, New York, NY, USA, 2010. ACM.

[8] L. Camp, F. Asgharpour, D. Liu, and I. Bloomington. Experimental Evaluations of Expert and Non-expert Computer Users? Mental Models of Security Risks. *Proceedings of WEIS 2007*, 2007.

[9] L. Clark and M. A. Sasse. Conceptual design reconsidered: The case of the internet session directory tool. In *Proceedings of HCI on People and Computers XII*, HCI 97, pages 67–84, London, UK, 1997. Springer-Verlag.

[10] Personal firewall software review. http://www.consumersearch.com/firewalls, 2010.

[11] L. F. Cranor. A framework for reasoning about the human in the loop. In *UPSEC'08: Proceedings of the 1st Conference on Usability, Psychology, and Security*, pages 1–15, Berkeley, CA, USA, 2008. USENIX Association.

[12] P. DiGioia and P. Dourish. Social navigation as a model for usable security. In *SOUPS '05*, pages 101–108, Pittsburgh, Pennsylvania, 2005. ACM.

[13] J. S. Downs, M. B. Holbrook, and L. F. Cranor. Decision strategies and susceptibility to phishing. In *SOUPS '06: Proceedings of the Second Symposium on Usable Privacy and Security*, pages 79–90, New York, NY, USA, 2006. ACM.

[14] S. Egelman, L. F. Cranor, and J. Hong. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *CHI '08: Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, pages 1065–1074, New York, NY, USA, 2008. ACM.

[15] S. Egelman, J. King, R. C. Miller, N. Ragouzis, and E. Shehan. Security user studies: methodologies and best practices. In *CHI Extended Abstracts*, pages 2833–2836. ACM, 2007.

[16] Best software firewalls for maximum protection and greater user involvement. http://www.techsupportalert.com/best-free-firewall.htm, 2010.

[17] S. Hazari. Perceptions of end-users on the requirements in personal firewall software: An exploratory study. *The Journal of Supercomputing*, 17(3):47–56, 2005.

[18] A. Herzog and N. Shahmehri. Usability and security of personal firewalls. *New Approaches for Security, Privacy and Trust in Complex Environments*, pages 37–48, 2007.

[19] K. Ingham and S. Forrest. A history and survey of network firewalls. Technical report, University of New Mexico, 2002.

[20] J. Johnston, J. H. P. Eloffa, and L. Labuschagneb. Security and human computer interfaces. *Computers and Security*, 22:675–684, 2003.

[21] H. Jungermann, H. Schutz, and M. Thuring. Mental models in risk assessment: Informing people about drugs. *Risk Analysis*, 8(1):147–155, 1988.

[22] S. Leonard, H. Otani, and M. Wogalter. Comprehension and memory. *Warnings and risk communication*, pages 149–187, 1999.

[23] D. Liu, F. Asgharpour, and L. Camp. Risk Communication in Security Using Mental Models. *Usable Security*, 7, 2008.

[24] J. McKechnie. *Webster's new universal unabridged dictionary*. Dorset & Baber, 1983.

[25] M. Morgan. *Risk communication: A mental models approach*. Cambridge University Press, 2002.

[26] S. Motiee, K. Hawkey, and K. Beznosov. Do windows users follow the principle of least privilege? investigating user account control practices. In *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS)*, pages 1–13, New York, NY, USA, July 14-16 2010. ACM.

[27] J. Nielsen. Card sorting to discover the users' model of the information space. http://www.useit.com/papers/sun/cardsort.html, 1995.

[28] C. Nodder. Users and trust: A microsoft case study. *Security and Usability*, pages 589–606, 2005.

[29] D. A. Norman. *Cognitive Engineering*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1986.

[30] Up-to-date coverage and product reviews of firewall software. http://www.pcmag.com/, 2010.

[31] Comodo firewall is a superb security program. http://www.pcworld.com/article/188008l, 2010.

[32] Popular in firewalls. http://www.pcworld.com/downloads/file/fid,63762-order,4/description.html, 2010.

[33] F. Raja, K. Hawkey, and K. Beznosov. Revealing hidden context: improving mental models of personal firewall users. In *SOUPS '09: Proceedings of the 5th Symposium on Usable Privacy and Security*, pages 1–12, New York, NY, USA, 2009. ACM.

[34] F. Raja, K. Hawkey, P. Jaferian, K. Beznosov, and K. S. Booth. It's Too Complicated, So I Turned It Off! Expectations, Perceptions, and Misconceptions of Personal Firewalls. In *Proceedings of the 3rd ACM Workshop on Assurable & Usable Security Configuration (SafeConfig)*, October 4 2010.

[35] C. Ronnfeldt. Three generations of environment and security research. *Journal of Peace Research*, 34(4):473–482, 1997.

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

[37] A. Sotirakopoulos, K. Hawkey, and K. Beznosov. "I did it because I trusted you": Challenges with the Study Environment Biasing Participant Behaviours. In *SOUPS Usable Security Experiment Reports (USER) Workshop*, 2010.

[38] W. Stephenson. *The study of behavior: Q-technique and its methodology*. University of Chicago Press, 1953.

[39] D. W. Stewart and I. M. Martin. Intended and unintended consequences of warning messages: A review and synthesis of empirical research. *Journal of Public Policy and Marketing*, 13(1):1–19, 1994.

[40] J. Stoll, C. S. Tashman, W. K. Edwards, and K. Spafford. Sesame: informing user security decisions with system visualization. In *CHI '08: Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human factors in Computing Systems*, pages 1045–1054, New York, NY, USA, 2008. ACM.

[41] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F. Cranor. Crying Wolf: An empirical study of SSL warning effectiveness. In *Proceedings of 18th USENIX Security Symposium*, pages 399–432, 2009.

[42] TopTenReviews: 2010 personal firewall software review product comparisons. http://personal-firewall-software-review.toptenreviews.com/, 2010.

[43] R. Wash. Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, SOUPS '10, pages 11:1–11:16, New York, NY, USA, 2010. ACM.

[44] Explore the features: Windows security center. http://www.microsoft.com/windows/windows-vista/features/security-center.aspx, 2010.

[45] M. Wogalter, V. Conzola, and T. Smith-Jackson. Research-based guidelines for warning design and evaluation. *Applied Ergonomics*, 33(3):219–230, 2002.

[46] J. S. Wolff and M. S. Wogalter. Comprehension of pictorial symbols: Effects of context and test method. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40:173–186(14), 1998.

[47] M. Wu, R. C. Miller, and S. L. Garfinkel. Do security toolbars actually prevent phishing attacks? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems(CHI '06)*, pages 601–610, New York, NY, USA, 2006. ACM.

[48] S. Young and D. Lovvoll. Intermediate processing stages: Methodological considerations for research on warnings. *Warnings and risk communication*, pages 27–52, 1999.

[49] M. E. Zurko, C. Kaufman, K. Spanbauer, and C. Bassett. Did you ever have to make up your mind? what notes users do when faced with a security decision. In *ACSAC '02: Proceedings of the 18th Annual Computer Security Applications Conference*, pages 371–381, Washington, DC, USA, 2002. IEEE Computer Society.

APPENDIX

A. FIRST FORMATIVE STUDY WARNINGS

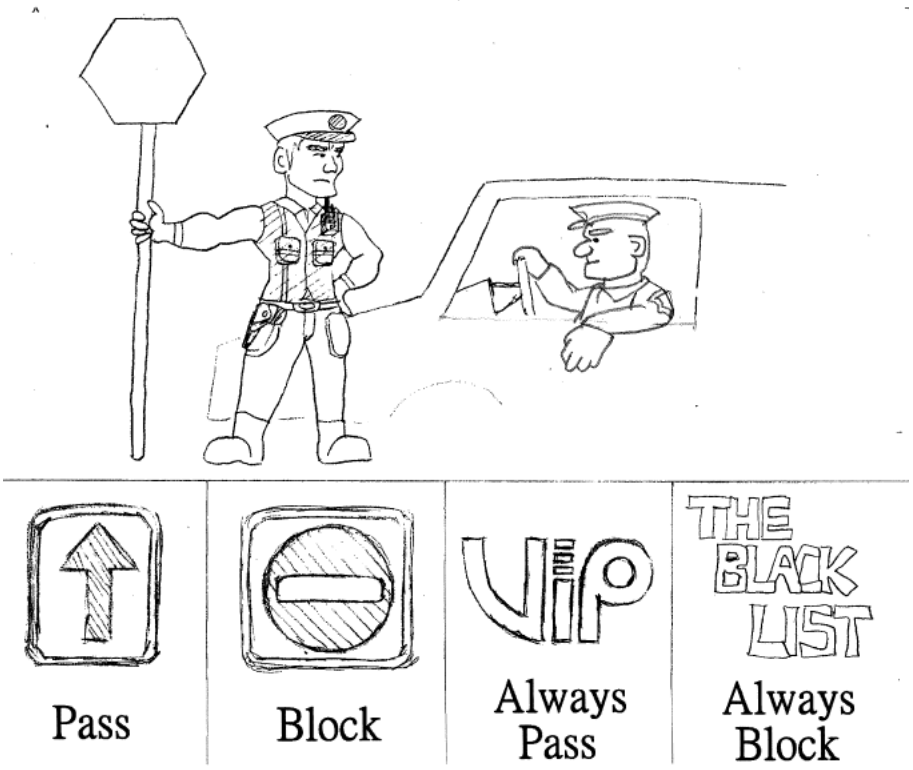

Figure 4: Firewall warning based on physical security by a policeman (for safe applications).

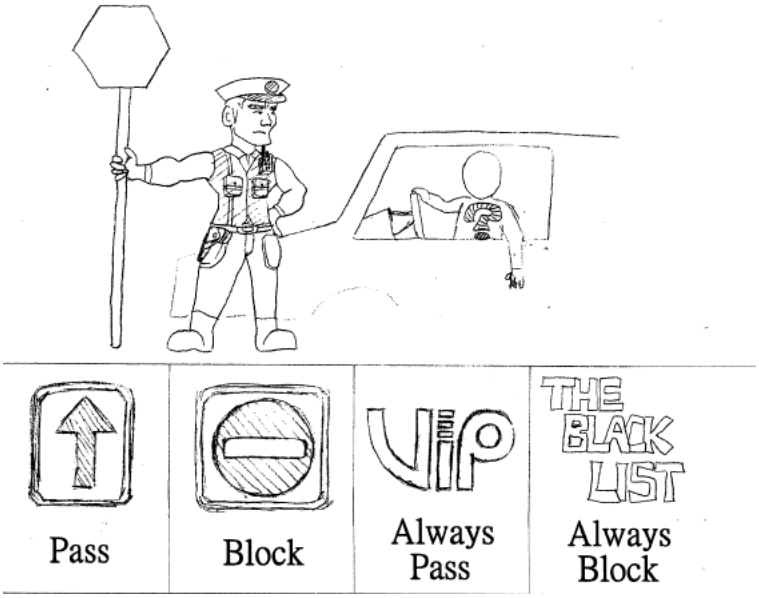

Figure 5: Firewall warning based on physical security by a policeman (for unrecognized applications).

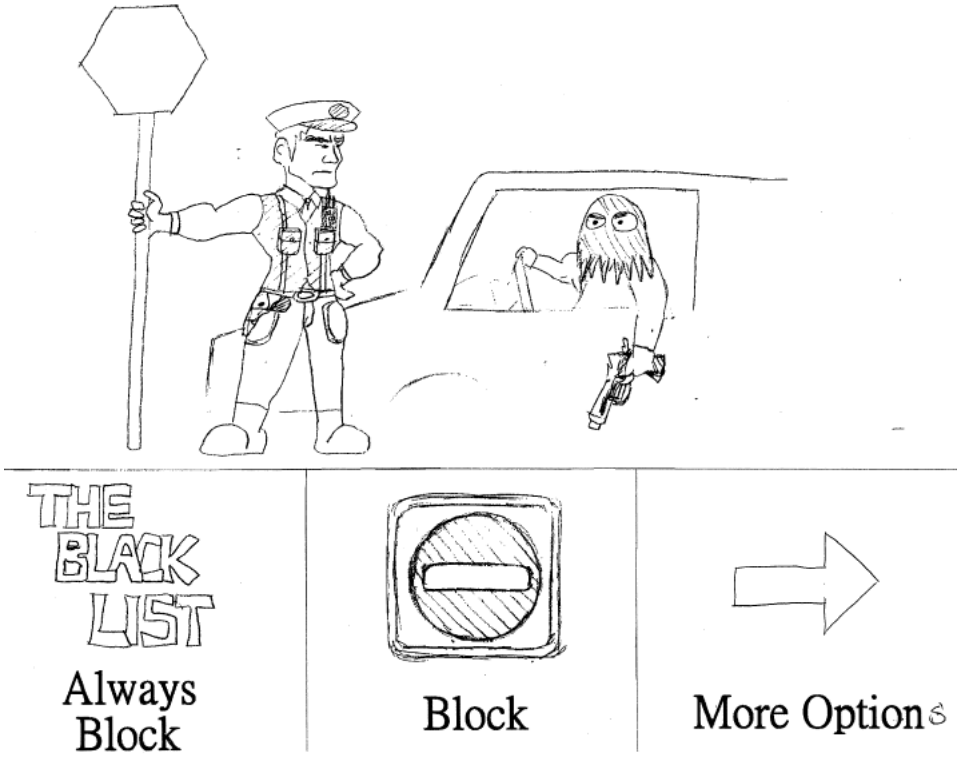

Figure 6: Firewall warning based on physical security by a policeman (for malicious applications).

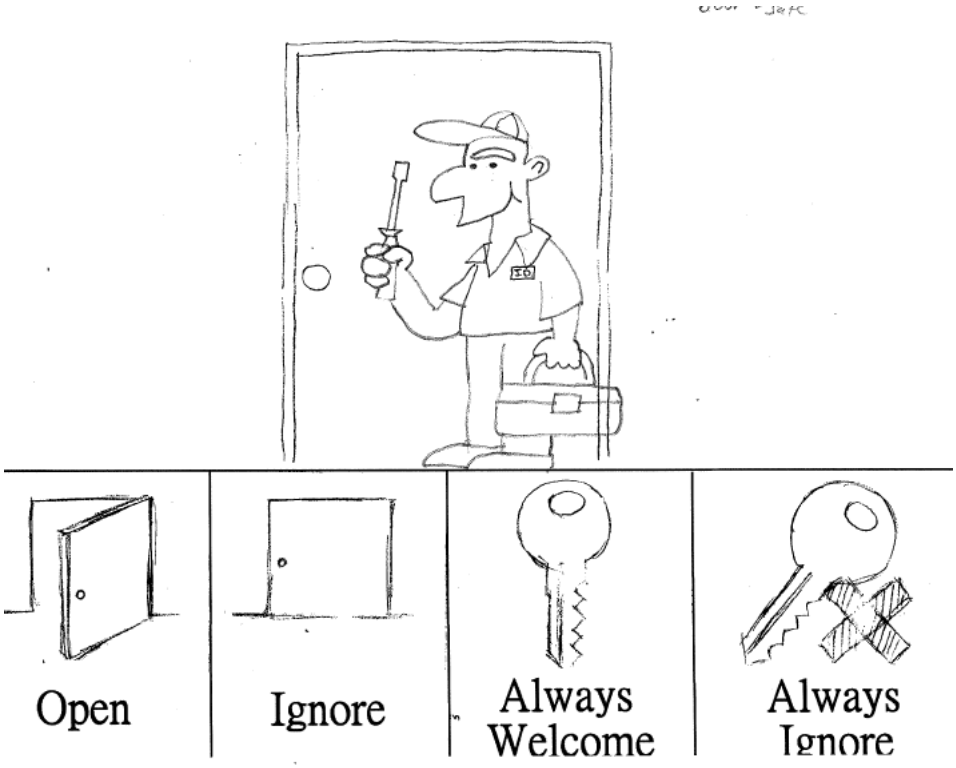

Figure 7: Firewall warning based on physical security by a door (for safe applications).

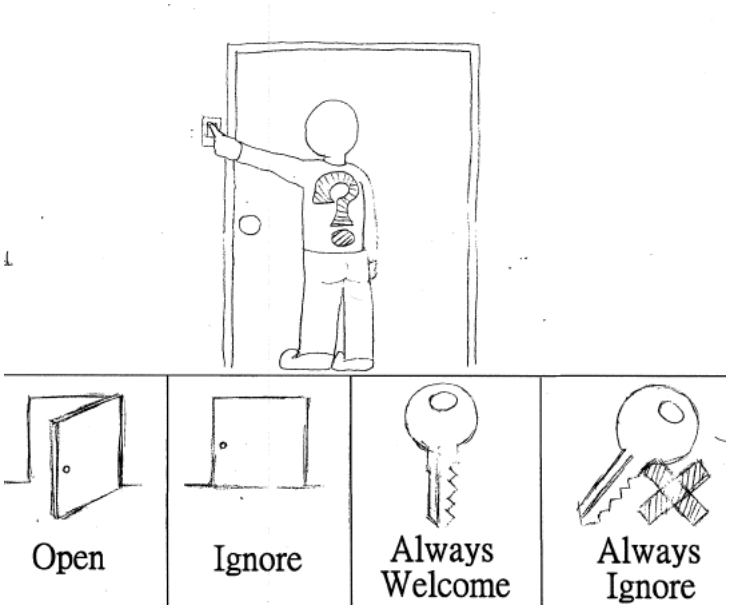

Figure 8: Firewall warning based on physical security by a door (for unrecognized applications).

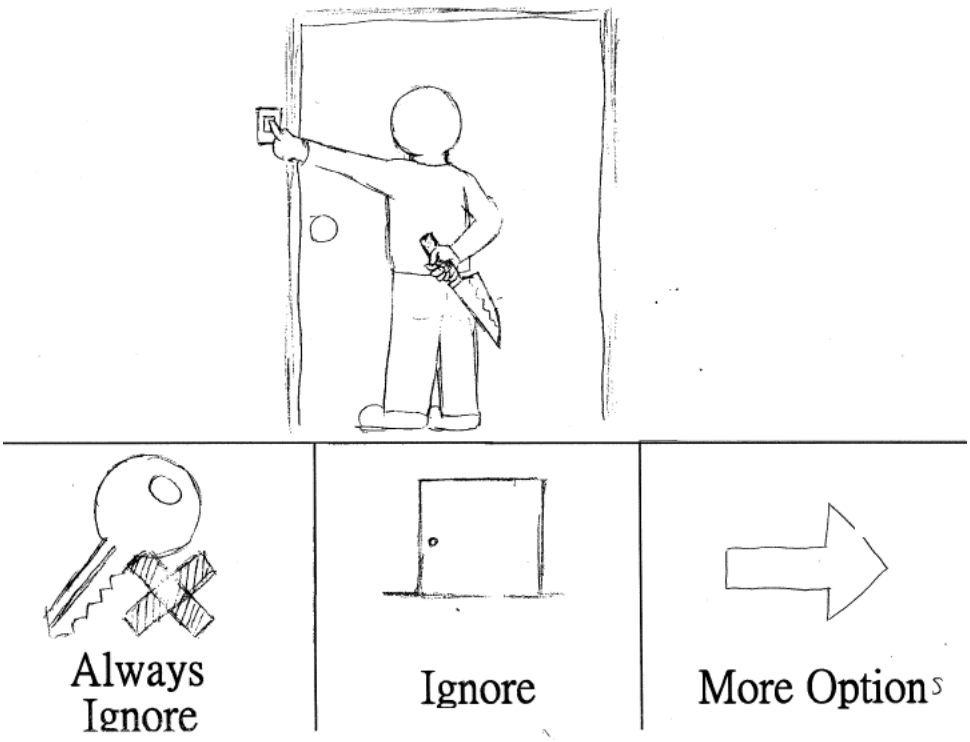

Figure 9: Firewall warning based on physical security by a door (for malicious applications).

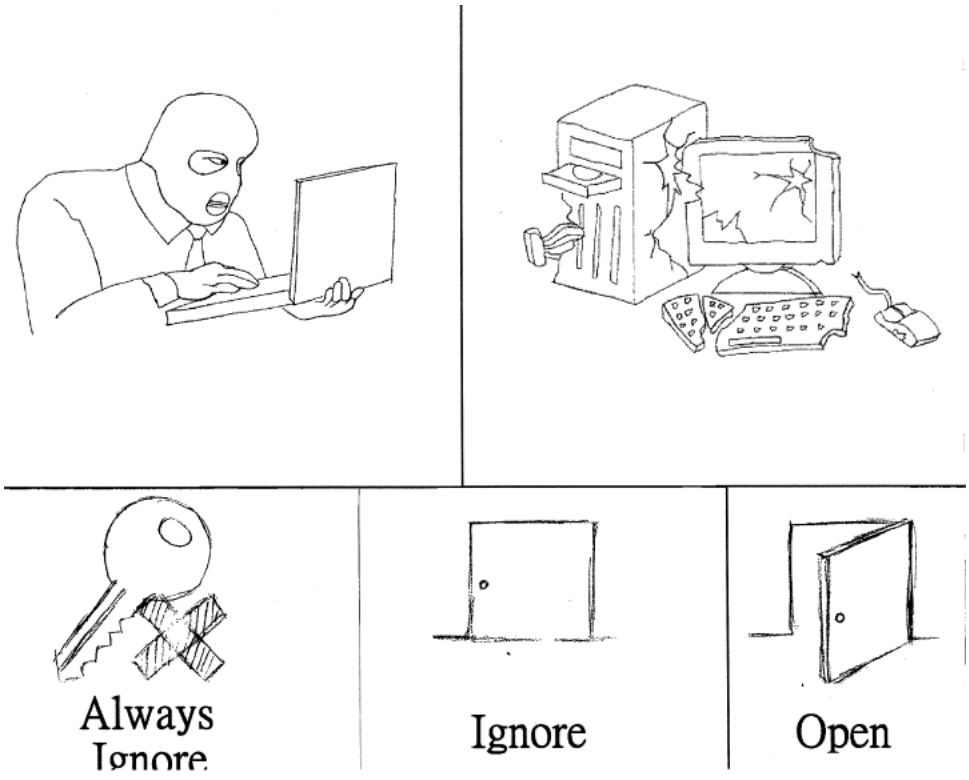

Figure 10: An interface for showing the consequences of allowing a malicious application.

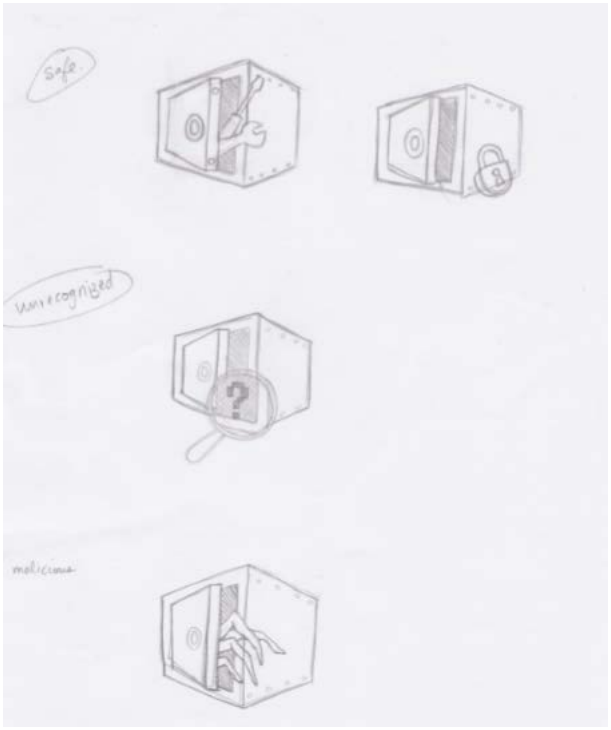

Figure 11: Firewall warning based on physical security by a safe.

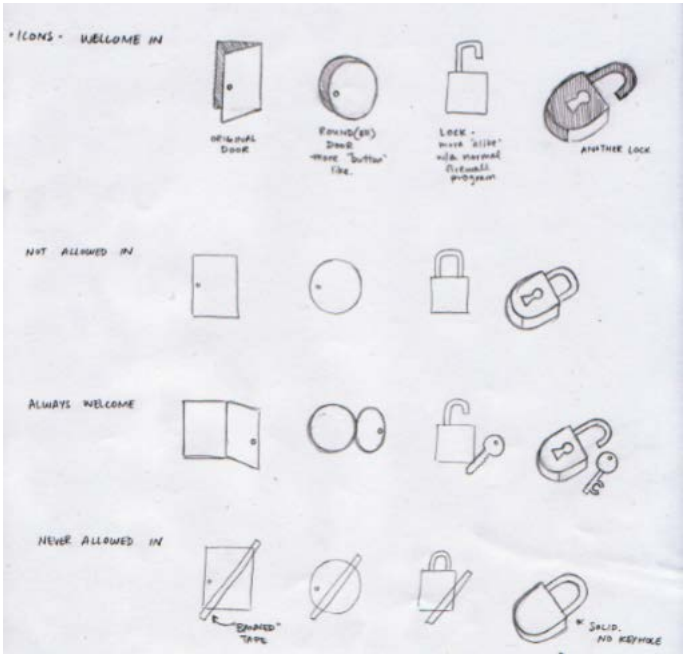

Figure 12: A sample of our initial designs for different actions.

B. SECOND FORMATIVE STUDY WARNINGS

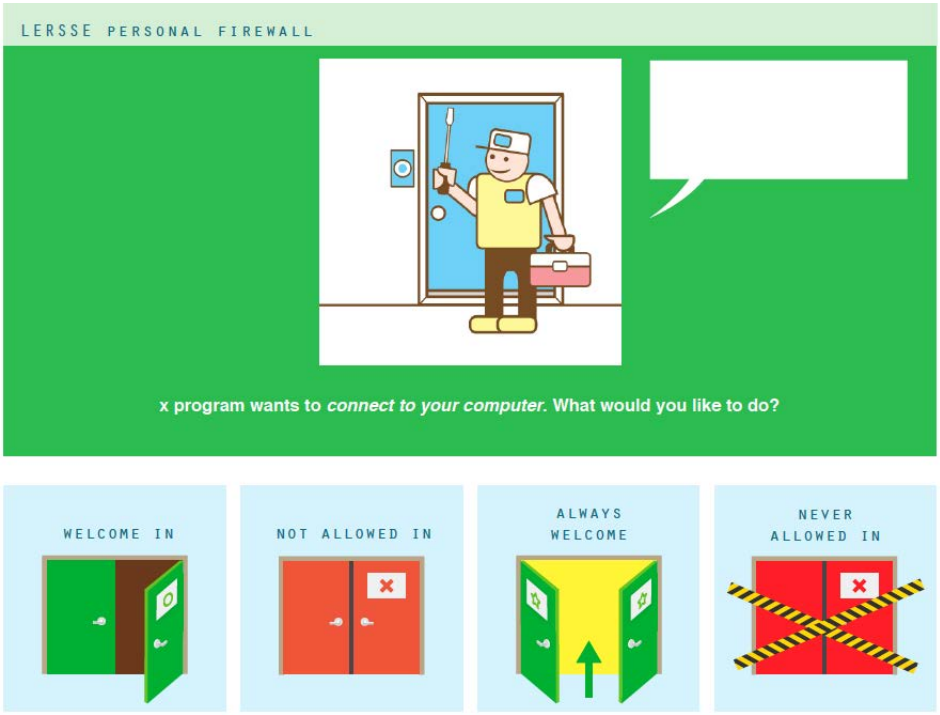

Figure 13: Firewall warning based on physical security by a door (for safe applications).

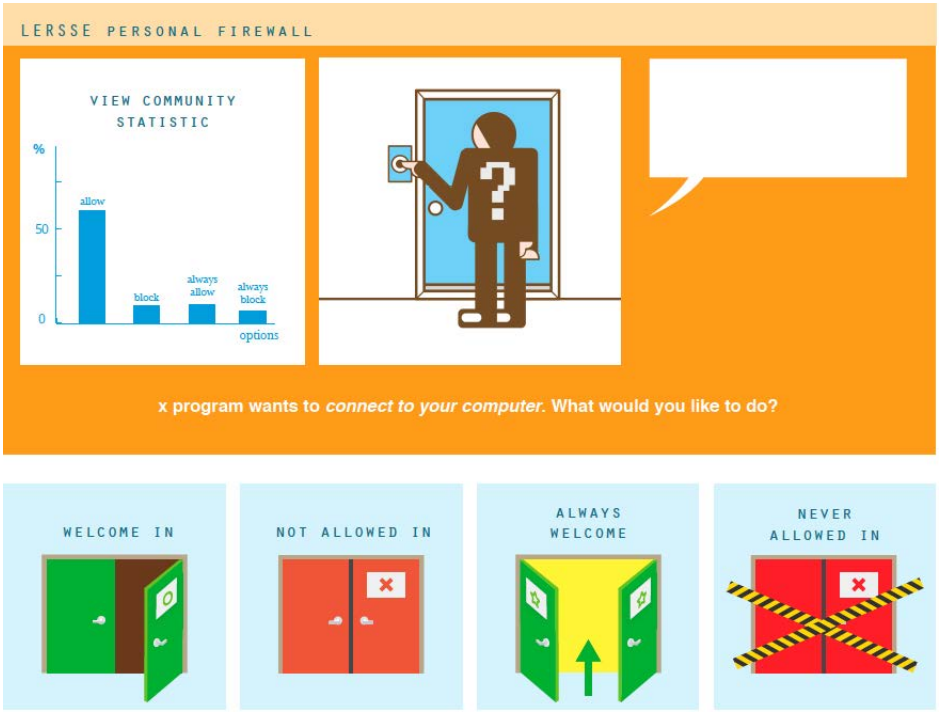

Figure 14: Firewall warning based on physical security by a door (for unrecognized applications).

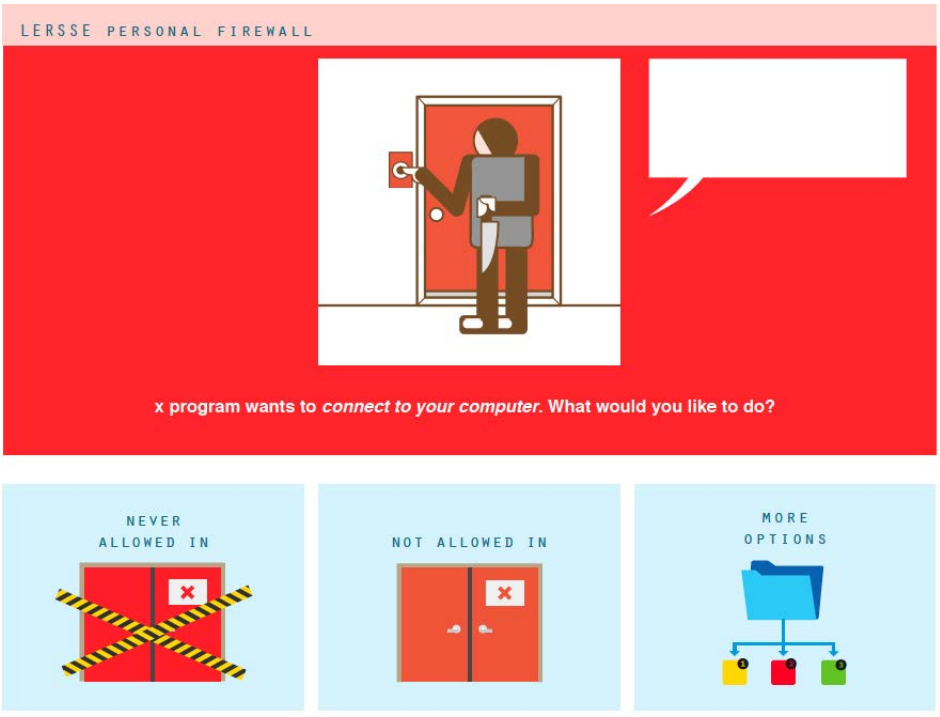

Figure 15: Firewall warning based on physical security by a door (for malicious applications).

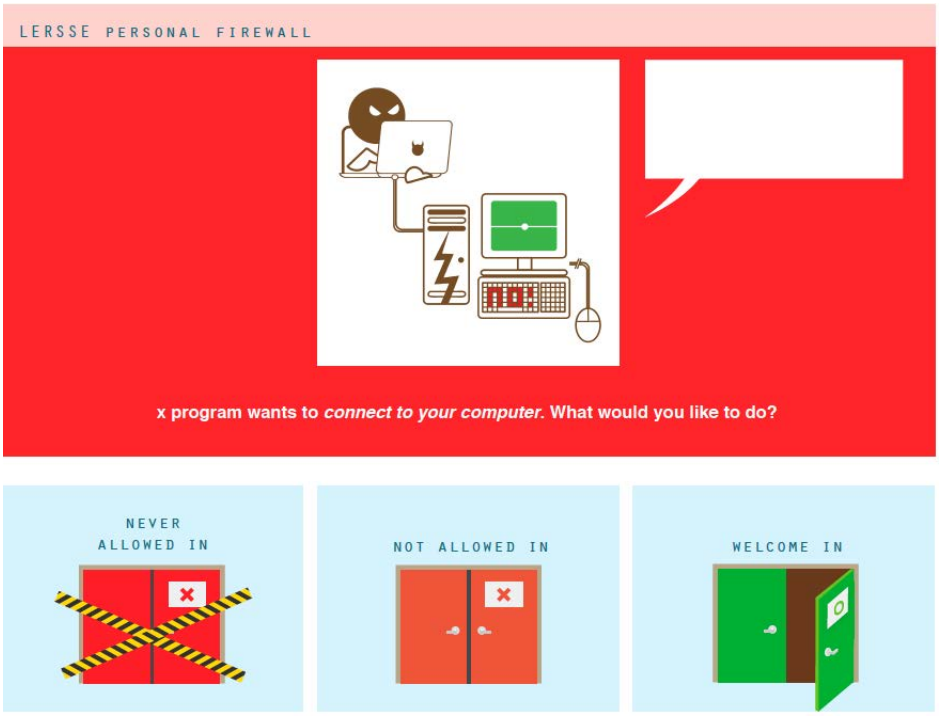

Figure 16: An interface for showing the consequences of allowing a malicious application.