Abstract

[1] There is a critical need in computer security to communicate risks and thereby enable informed decisions by naive users. Yet computer security has not been engaged with the scholarship of risk communication. While the existence of malicious actors may appear at first to distinguish computer risk from environmental or medical risk, the impersonal un-targeted nature of the exploitation of computing resources and the technical complexity of the risks are similarities. This work is a first experimental step in evaluating the informal, implicit, and unexamined use of mental models in computer security. The experiments described in this paper have three results. First, the experiments show that for a wide range of security risks self-identified security experts and non-experts have quite distinct mental models. Second, a stronger definition of expertise increases the distance between the mental models of non-experts and experts. Finally, the implicit and informal use of models through metaphors in the computer security community has not resulted in metaphors that match the mental models of naive users, and more surprisingly , of self-identified experts. We close with a description of our research agenda targeted at developing a better understanding of the mental models of naive users.

Keywords Security, Privacy, Mental Models, Card Sorting, Risk Communication, Computer Risks

[1]This work was produced in part with support from the Institute for Information Infrastructure Protection research program. The I3P is managed by Dartmouth College, and supported under Award number 2003-TK-TX-0003 from the U.S. DHS, Science and Technology Directorate. This material is based upon work supported by the National Science Foundation under award number 0705676. Opinions, findings, conclusions, recommendations or points of view in this document are those of the author(s) and do not necessarily represent the official position of the U.S. Department of Homeland Security, National Science Foundation, the Science and Technology Directorate, the I3P, the NSF, Indiana University, or Dartmouth College.

1 Introduction

Solutions to the problem of lack of security adoption have included insurance that correlates with security practices [2], changing defaults so that security is difficult to avoid [28], and improved system usability [42], [16] [43]. It is the core contention of this paper and our team's associated research agenda that there is a clear and profound need for effective risk communication. While there have been studies of user conceptions of privacy [20] and usable security; these have focused on system design rather than content communication. In this paper we offer our own step forward in examining risk communication in the context of computer security.

Effective security risk communication requires both communicating risk information and motivating the appropriate risk behaviors. One may think that since experts best understand the risks to be described, that the expert's mental model is the most reliable for designing risk communication instruments. The essential point is that the purpose of risk communication is not conveying the perfect truth to the users, but rather prompting them 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 that risk, efficacy of the risk communication requires communication that is aligned with the mental model of the target group. Effective risk communication often requires more of an understanding of the risk perception of the communication target, as opposed to a communication optimized for technical accuracy. [41, 32]

The mental models approach is a risk communication method grounded in the conceptual models of the recipients of risk communication. A mental model is a simplified internal concept of a process. This concept is case specific and may depend on life experience, stigmatization of the risk, perception of the risk, and individual information processing strategies [37]. The mental models approach has been successfully applied in

environmental as well as medical risk communication [25] [38]. There has been one initial exploration of mental models in privacy research [15] but none in security research. (Obviously excluding the work of described here.)

The first step in out application of mental models to computer security research was to determine the scope of mental models used in the computer security profession. We began with an examination of the security literature and found five widely used conceptual models implicit in language or explicit in metaphors. A more complete description of the use of these metaphors and more examples of their use in security can be found in [7]. These conceptual models form the basis of this experimental exploration of mental models.

Physical Safety: The physical concept of security is implicit in descriptions of 'locks', 'keys', 'safe computing' and other mechanisms for physical risk mitigation. This concept implies individual and localized control.

Medical Infections: The model of security incidents as medical infections is grounded in the patterns of diffusion of 'computer infections'. 'Worms', and 'swarms' are terms of art in computer security. A research domain in computer security is the importance of heterogeneity in the larger network, or the network as ecosystem. [27].

Criminal Behavior: Computer security violations can be criminal with the launching of 'malicious' code, but these virtual acts cannot include a physical 'break-ins' or 'intrusion'. The concept of computer risks as risks of being a victim of crime acknowledges the existence of a malicious and motivated attacker.

Warfare: The warfare concept of computer security implies the existence of a determined implacable enemy, with 'demilitarized zones,' 'firewalls', and 'offensives'. This

model has the potential to leverage horror by leveraging the horrors of war [14], yet this model also leaves the individual home computer owner out of the equation.

Economic Failure: Security and network or software vulnerabilities can be seen as market failures [3, 35, 45]. Vulnerabilities, in particular, are externalities [8]. Computer security failures cause downtime and costs [10, 21].

These metaphors arose as ad-hoc labeling with some being evocative (e.g., an 'intrusion' into a computer) and some approaching non-sensical (e.g., 'phishing'). Previously there had not been explicit mental models evaluation of these metaphors with respect to either expert or non-expert conceptions of computer security risk. Therefore we designed this experiment to answer a series of preliminary questions. The goals is not only to obtain first cut answers at these questions, but also to inform the design of future qualitative work in evaluating computer security risk communication.

The first question is if the above metaphors and models implicit in the security literature correlate with the mental models of experts or non-experts. Second, do the mental models of experts correlate with the mental models of lay users? Third, how sensitive is the correlation between experts' and non-experts' mental models to the definition of expertise? Fourth, to what extent can we characterize these differences?

In order to answer the above questions we implemented two card sorting [24] experiments. The two experiments differ in definition of expertise.

Our results argue that the concepts of security as embedded in literature are not well matched to the mental models of non-experts, and more surprisingly, of experts. Further, we find experts and non-expert users have significantly different mental models. These results proved sensitive to the definition of expert. The more stringent the definition of the expert, the greater the distance between the mental models of self-defined expert and non-experts.

Section 2 identifies some the related work in risk communication. Section 3 explains the details of our experimental setup. Section 4 covers the data analysis and findings. Section 5 concludes the paper with a review of the findings and our future research direction.

2 Related Work

Better risk communication with respect to computer privacy and security risks is needed to change the risk behaviors of individual naive users. Risk communication in the computer security context has typically consisted of messages designed by security experts to inform a community of non-experts. These communications attempt to communicate complex risk in an environment where there is competition for the individual's attention, and the range of expertise is significant. This communication is made more complex by the lack of validated security metrics in which the communication could be grounded. In contrast to health and environmental research, there is no widely used or supported security risk assessment mechanism. Two security experts completing an assessment with the same data may come to very different conclusions. Best practices and rules of thumb abound. For example, a leading textbook is named *Computer Security: Art and Science.* [4] As a result, uninformed risk behaviors by naive users remain a challenge in computer security.

A common user experience is that of a text box popping up during the context of an online choice. The current, pop-up box method of risk communication embeds many elements that have been rejected in risk communication literature. First, the information lacks context. [40] The findings are not put into perspective. There is no indication of a standard that you can use to judge the risk, or instructions on enabling the action while mitigating the risk. [23].

Figure 1: Extremely High Level Information

The security warning provides either an option of overly technical information, as shown in Figure 1 or extremely high level information, as shown in Figure 2. When the more detailed information is provided, it is not effectively communicated.

Rarely are the risks corresponding to the actions in the risk communication clearly identified. In no case is there an indicator of risk-reducing action that might be taken in order to reduce the risks should the user choose the particular action. Examples of actions that individuals could take to mitigate risks include assuring that the latest version of the browser is installed or that there are functioning firewall and virus programs. Another example is that the browser application can confirm that the person is not browsing in super-user mode. Additional functionality could determine if the latest patches are installed. In the communication to the user, the risks are not associated with the actions. The risk of having a subverted machine and thus being both a victim and an unknowing participant in computer crime is in no way visible in the communication about the action. [17]

The information in Figure 2 is presented in its most technical form, bereft of even minimal narrative. While the extent of narrative may be debated, the complete absence of narrative and context is rarely optimal. [22]

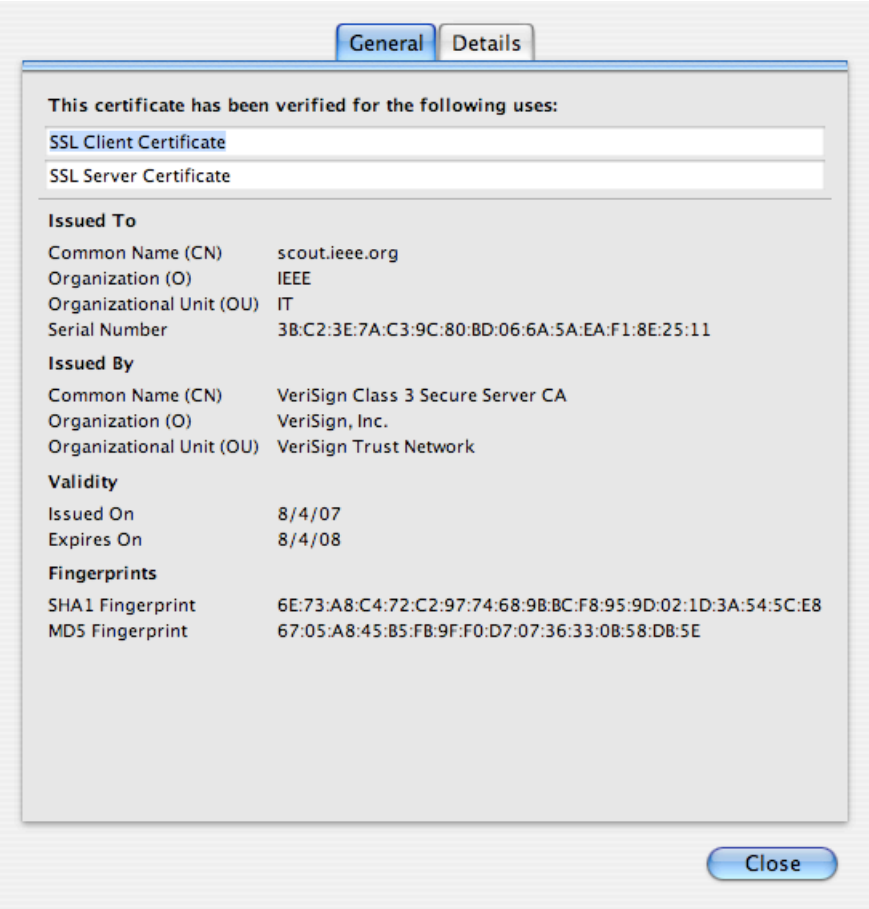

Figure 2: Technical Details with Minutia

In modern interaction information systems, the risk communication mechanisms themselves are not built to allow individuals to annotate, share or easily examine their own histories of risk-taking to evaluate their decisions. The result is that individuals take social and individual risks that, on their face, would seem unacceptable. Individuals with computers that are not secure support socially unacceptable behaviors including spamming others, hosting phishing sites, participating in extortion, and other malicious behaviors. [19] The lack of computer security can result in socially malicious behavior as well as personal loss, yet there is no information on these risks, just warning about specific behaviors. [44]

Risk communication to end users has consistently failed in the computer security realm. As a result individual home users of machines are not only at risk from losing all their personal information, but in fact are unknowing and unwilling participants in electronic crime. While there are a range of technologies to mitigate this state of affairs, adoption of even free technologies has proven inadequate to mitigate these threats. Examples of free technologies include the stand alone (e.g., Zone Alarm is a free and open source firewall) and the bundled (e.g. WPA is available on all wireless routers sold for home markets). The problem of individuals who do not address known security vulnerabilities is so great that the most damaging worms have been the least technically sophisticated. [29]

There is a critical need for informed decision-making by naive computer owners. Zombies, botnets and DDoS are examples of attacks that are enabled by the large number of home-based machines which are not secure. The growing reliance on online services combined with the the exponential growth of security breaches [9], zombies, and botnets [31] suggests a need for better security practice by average users.

Zombies are machines that have been subverted by malicious code, so that they can

be controlled by a remote malicious user. Zombies may appear to function normally to a home user, but will generate traffic and respond to commands by the controlling remote malicious user. Zombies are used to host phishing sites, store illegal data, or collect information on the users via keystroke logging or password harvesting. Botnets are large numbers of zombies, from tens to tens of thousands, that are controlled by a single entity, usually through commands issued over Internet relay chat. These are used in more sophisticated (i.e., distributed) phishing attacks. Zombies and botnets harvest the passwords, credit card numbers and other information entered by the machine users. Botnets are also used for distributed denial of service attacks (DDoS). A denial of service attack (DoS) overwhelms the target server by generating an order of magnitude more requests, including many carefully malformed requests, for service. These malicious attacks prevent legitimate users from access the server. A DDoS attack is one in which many computers, usually comprising a botnet, participate.

A primary goal of effective risk communication is to assist individuals in making informed decisions about which risks are acceptable and which are not. Clarity in risk communication can assist individuals with improved decision-making. How can computer security move from its current state of ineffective practice to clear communication to individuals that enables informed interaction? A natural but naive first step would be to use the metaphors already in computer security (e.g., keys, worms, firewalls) to design risk communication. This has been widely done in user education programs with limited success. In fact, after year-long investment in computer security education at Indiana University, the percentage of students who did not use encryption on their home routers increased. [13] In contrast, we propose that the first step is understanding if the current implicit metaphors are aligned with the mental models of users. Our experimental results suggest that this is not the case.

Mental models have been used to examine user attitude towards privacy and purposeful information disclosure. Diesner et al. [15] have studied the mental models of Indian information professionals by conducting in-depth interviews and found them similar to Americans with the same level of expertise. Acquisti and Gross [1] have shown that individuals have unrealistic risk assumptions about privacy risks in online social networks. Their study shows that individuals expressed tolerance for privacy risks are at best a weak predictor of their information provision to the network.

Mental models have been widely used in human-computer interaction and usability [34]. However, the concepts of mental models in HCI and risk communication are distinct. In HCI, a mental model defines how people interact with systems. Norman [37] suggests that the usability, functionality and learnability of the conceptualized model of the designer depend on the alignment between the conceptualized model of the design and the mental models of the end users. From these three factors, the functionality, and learnability of the risk communication refer to its potential to prompt the target group to take the desired action to mitigate the addressed risk. Cosantine and Lockwood define four criteria for usability of a product: learnability, retainability, efficiency of use and, user satisfaction. Learnability and retainability are the two criteria pointing to the role of mental models in usability. Risk communication and usability have in common that to to the extent that a correct mental model can be learned and retained by a user, the more effective the user will be in managing a system or avoiding risk.

In risk communication the concept of mental models is subtly distinct from the concept of mental models in usability, e.g., [36, 37, 11] . This work is grounded in mental models as it has been developed in risk communication [12].

Mental models in risk communication suggests that one predictor of the efficacy of risk communication is the alignment between the recipient's internal conception of a risk-

generating process and the conception of the process embedded in the communication. Morgan has applied mental models to a wide range of environmental applications. [33] Bostrom has applied mental models in home hazards. [5] Fischoff advocates using mental models to minimize over-confidence in individual perceptions of personal safety. [18] Acquisti illustrated over-confidence in the case of on-line privacy. [1] Keller proposes using mental models to mitigate the availability heuristic which produces irrational risk behavior. [26]

One us of mental models in environmental research is to enhance risk awareness about a particular risk, e.g. household toxins, and alter consumer behavior [33]. Like computer security risks, environmental risks are much more complex to manage in household than in industrial or corporate conditions. Paint stripper and other chemical hazards are, like computers, more easily regulated and controlled in the work place than home.

3 Experimental Design and Implementation

3.1 Card Sorting

Due to the complexity of human knowledge acquisition and psychology, the discovery of implicit mental models is a difficult task. This task could be done using various elicitation techniques such as Teachback Interviews, Repertory Grid, Goal-Oriented Approach, Grounded Theory or, our choice here, Card Sorting [6]. Card sorting [9, 39] is a structured elicitation technique based on requiring participants to to sort a pile of words written onto cards into different piles.

There are two kinds of card sorting: closed and open. In a closed card sort a subject must sort each card into one of some number of predefined groups. In an open card sort, no predetermined groups are given. Thus in an open card sorting experiment, the subject

can sort the words into arbitrary groups according to that subject's perception. A benefit of the card sorting technique is that it is easy for the subjects to perform. In this case, we began with a set of groups derived from the security literature. (See [7]) We used a closed card sort to evaluate the correspondence of mental models of lay users and experts with each other and with the implicit models in the literature.

To summarize, the goal in the design of the experiment is to answer the following questions –

1. Do the mental models implicit in the security literature correlate with the mental models of experts or non-experts?

2. To what degree do the mental models of experts correlate with the mental models of lay users?

3. How sensitive is the correlation between experts' and non-experts' mental models to the definition of expertise?

To find the correlation between individual mental models and level of expertise in security the same experiment was repeated, with two different definitions of expertise. Throughout the paper we refer to the two implementations of the experiment as Experiment 1 and Experiment 2 with the definition of expertise being weaker in Experiment 1.

We implemented an on-line card sorting experiment. The participants were given six label and color pairs: physical security = green, medical infection = blue, warfare = red, criminal behavior = orange, economics failure = yellow, and "I can't decide" = purple. To avoid effecting the participants' decisions, we did not follow any specific cultural pattern in associating colors with labels. For instance, the color green is associated with peace for some people and with the environment for others. The arbitrary color selection made the

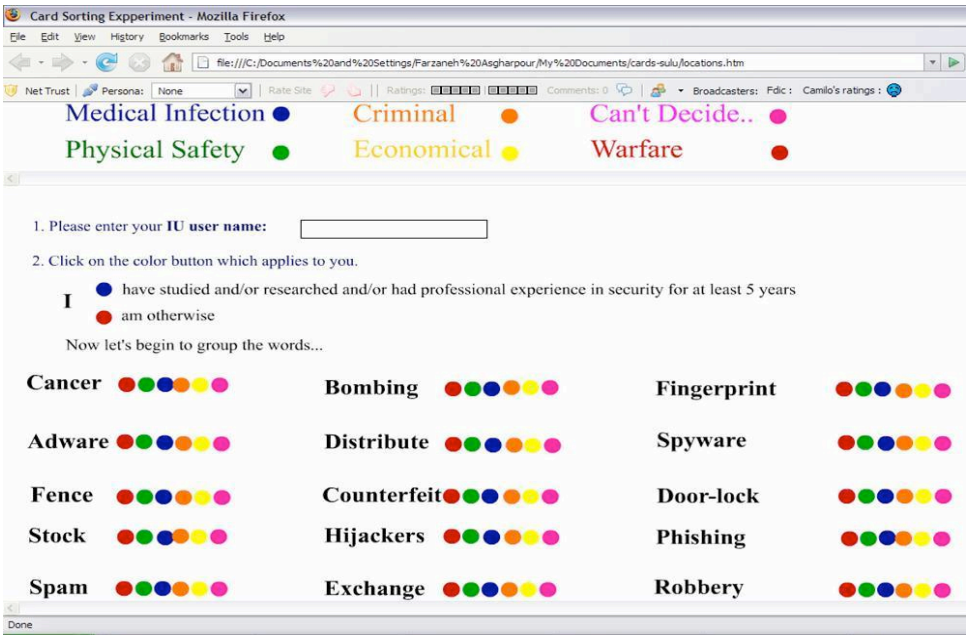

Figure 3: Screenshot of the card sorting experiment

participants refer to the instructions frequently, so the instructions were placed in a frame above the scrolling words in refining the experiment.

The participants labeled a word by changing its color. All words were initially black. Each word was accompanied with six colored buttons. Clicking on each button changed the color of the corresponding word into the color of the selected button. Words could be changed multiple times, until the participant completed the experiment. Figure 3 shows a screenshot of the experiment's interface.

The experiment was developed using Macromedia Flash and PHP.

A closed card sort experiment requires first and foremost a set of words in order to label the cards. Recall that we asked participants to group the given words into these six labels: physical security, medical infections, criminal behavior, economic failure, warfare or indeterminate. Recall, for the last category, participants were instructed to label a word

Crime	**Medical**	**Physical**	**Warfare**	**Economic**
Theft	*Epidemic*	*Fence*	*War*	*Trade*
Housebreaking	*Fever*	*Door-lock*	*Bombing*	*Export*
Kidnapping	*Illness*	*Shield*	*Destroy*	*Stock*
Fingerprint	Cancer	Inviolability	Terror	Distribute
Counterfeit	Detoxification	Invulnerability	Attack	Exchange
Robbery	Nausea		Suicide	Endorse
Mugging	Inflammation			Advertise
Vandalism	Contagious			Monetary Risk
	Sore			

Table 1: Word List

with "I can't decide" if they could not determine a single category, had no impression, or felt the word fit none of the other categories.

The words related to each mental model were selected using Webster's Thesaurus. For instance, the words selected for security as crime are related to "theft" according to Webster's Thesaurus. Table 3.1 provides the word list. The participants were allowed and encouraged to look up the words with which they were not familiar. As the experiment was online, dictionaries were readily available. Using an introductory security textbook, we selected 29 security related words from the index. Using what later were confirmed as obvious words (see Section 4 we used Webster's Thesaurus to select 6 words related to theft for physical security, 9 related to disease for medical, 9 related to assault for criminal behavior, 7 words related to trade for the economic model and 6 words related to "war".

Recall that there were three to four times as many computer security words as words in other categories, and the participants had to sort these into the groups above. The computer security words were: trojan, keystroke logger, junk mail, virus, worm, hacking, binder, exploit, zombie, authentication, click fraud, password, userid, firewall, backdoor, blacklist, spoofing, address book, honeypot, drive-by-download, dos attack, spam, phishing, spyware, adware, cookies and identity theft.

Expertise was asserted by the participants. In the two experiments there were two distinct definitions of expertise. In Experiment 1 participants asserted their expertise according to the following set of questions.

Expert (E_1): An expert is a person who knows the technical definitions of all the security-related words listed below.

Non-Expert (NE_1): One who does not know the technical definition of the security words and at most knows some practical aspects of the risks.

In Experiment 2, participants asserted their expertise according following set of questions.

Expert (E_2): An expert is a person who has had at least five years experience in security as a researcher, student or practitioner.

Non-Expert (NE_2): Otherwise

In both experiments the definitions of expert and non-expert were given in the instruction section. The participants declared their expertise according to the given definitions while viewing the words but before beginning the experiment.

For Experiment 1 (E_1) the results as classified as weak expert (WE) and weak non-experts (WNE). Results from the second experiment (E_2) are similarly referred to the experts as strong experts (SE) and strong non-experts (SNE).

The first experiment included 22 self-defined experts and 49 non-experts. The second experiment included 11 self-defined experts and 27 non-expert participants. In both experiments the participants were 18-50 years old. Participants included faculty, staff, graduate and undergraduate students in informatics or computer science departments.

4 Analysis

The methodology and definitions introduced in this section apply to both experiments unless otherwise noted. During this analysis of results, please recall that the first three words under each mental model are the *obvious words* for that mental model in Table 3.1 . While all words were selected from a thesaurus, the selection of obvious words was verified by participant behavior rather than initial observation of placement. Not surprisingly, these were the same.

There are two steps to the analysis of the results. The first subsection provides is a summary that offers some basic descriptions of the results. The second subsection begins with a description of the multidimensional scaling model and provides more detailed analysis of our findings.

4.1 Analytical Summary

After completing the experiment with both groups of participants, we calculated the matrix of intra-similarity between the words. First the original data were tabulated. Each time a participant marked a pair of words with the same color, that was counted that as an indicator for similarity between the two words.

As an example, if most of the participants mark the words "trade" and "stock" with the same color, then we can say these two words are similar. In contrast, if only a few participants assign the words "war" and "fever" with the same color, we interpret this result as these two words being dissimilar.

Based on participants choice of word correspondence, we constructed two 66×66 matrices, one for weak experts and one for weak non-experts from E_1. We named these two matrices *Weak Expert's Similarity Matrix* (WESM), and *Weak Non-expert's Similarity Matrix* (WNSM). We constructed corresponding matrices (SNSM and SNSM) for E_2.

	Cancer	Bombing	Fingerprint	Adware
Cancer	49	4	1	2
Bombing	4	49	1	5
Fingerprint	1	1	49	12
Adware	2	5	12	49

Table 2: Number of E_1 Weak Non-experts labeling each word pair with the same color

MM	Weak Experts	Weak Non-experts	Strong Experts	Strong Non-experts
Criminal	48$	38$	55$	59$
Physical	14$	45$	11$	28$
Warfare	0$	0$	17$	0$
Economic	34$	17$	3$	10$
Medical	3$	0$	14$	3$

Table 3: Percentage of 29 Risks Corresponding to Each Mental Model

As an example, Table 2 shows part of the Weak Experts' Similarity Matrix matrix from E_1. As shown in this table, 12 weak-non-experts consider "Adware" and "Fingerprint" similar words.

In order to reveal underlying dimensions of the participants' choices we calculated multidimensional scaling maps from the matrices. Table 3 offers a high level summary of the distinctions between security risks using the multidimensional scaling (as described in Subsection 4.2

The leftmost column lists the mental models as derived from the security literature. Each of the other columns corresponds to one experimental group. The first two columns correspond to the non-experts and experts in the first instantiation of the experiment, E_1. The second two correspond to experts and non-experts in the second instantiation of the experiment, E_2. Notice that the distribution of mental models among experts and non-experts in the two experiments are significantly different.

Based on the methodology detailed in Section 4.2, for each risk r and each group of participants, the mental model with minimum distance from r is assigned as the mental

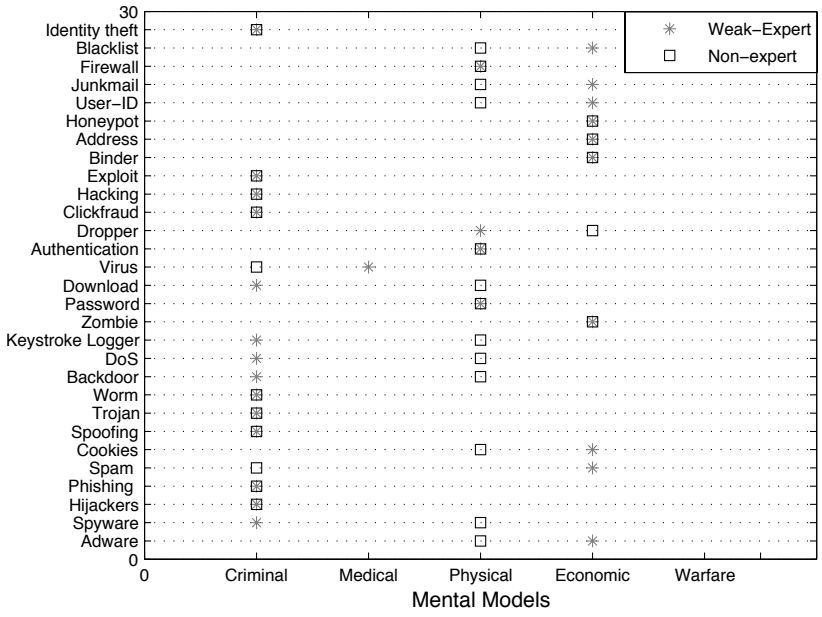

Figure 4: Mental models of Weak Experts and Corresponding Non-experts in E_1

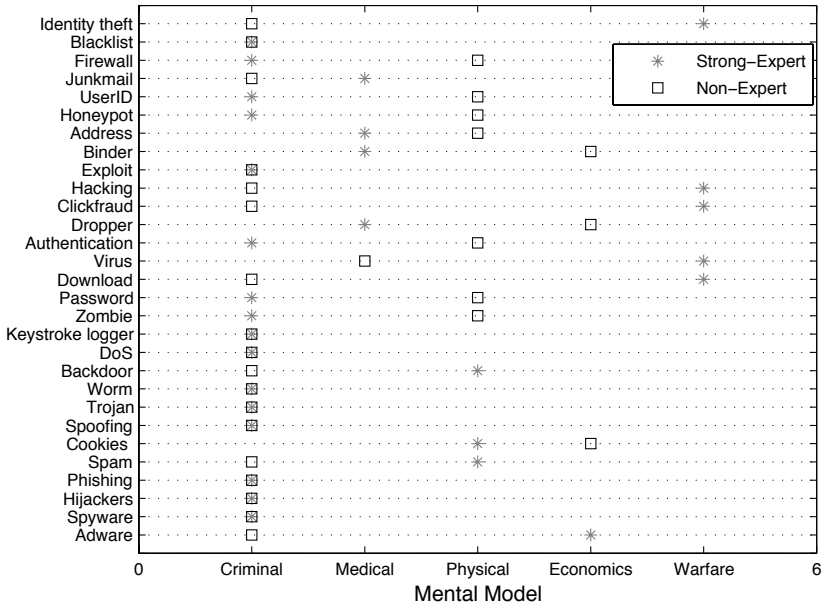

Figure 5: Mental Models of Strong Experts and Corresponding Non-experts in E_2

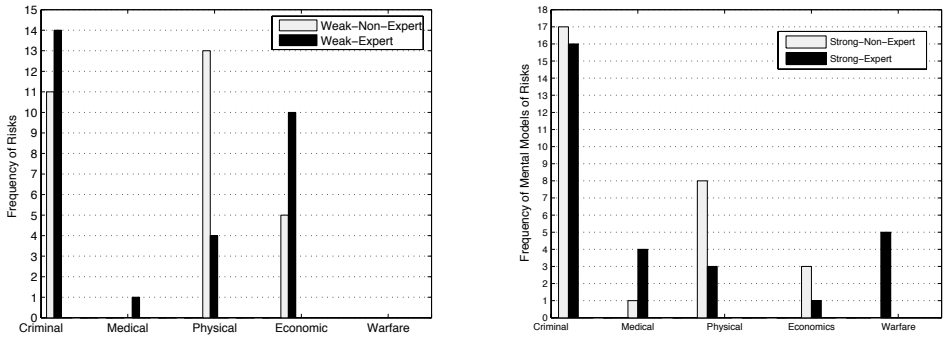

Figure 6: Distribution of Mental models For Security Risks in Experiment 1 (left) and Experiment 2 (right)

model of that risk r. The mental models corresponding to each security risk that corresponds for each set of participants is found by applying the least distance measure to each word. Figures 4 and 5 illustrate the results for E_1 and E_2, respectively.

4.2 Multidimensional Scaling

As noted above, we use the multidimensional scaling (MDS) [12] method to locate the expert's and non-expert's similarity matrices WESM, WNSM, SESM, and SNSM in two dimensional space. The multidimensional scaling (MDS) method is used to evaluate the structure in a dataset by measuring the (dis)similarities in the data as distances. MDS assigns objects to specific points in a conceptual space such that the distances between points in that space correspond to dissimilarities in the data. [30]

Since MDS operates on either relative distance or similarity between observations, one can map the observations using similarity or dissimilarity matrices. These are straightforward matrix transformations, so the choice of similarity versus dissimilarity does not bias of the results. We mapped the similarities, i.e. the matches of the words from the

card sorting experiments, into two dimensional space. We then used relative distances between the words to assign mental models to each security risk with the least distance corresponding to the closest match.

The Statistical Package for the Social Sciences (SPSS) was used to convert the similarity matrices into the distance matrices. Despite the fact that the set of related words for each mental model was chosen from Webster's Thesaurus, the participants sometimes labeled the words differently from our original assumptions. We therefore re-evaluated which of words would be the best *obvious words*. (Recall Table 3.1 shows a list of three obvious words under each mental model). We confirm that these word groups form a set of *obvious Words* based on the observation that for each group of three words a minimum of 85 of the participants have labeled all the words with the same mental model. For instance, 90 of all the participants labeled the words "illness", "epidemic", and "fever" as medical infections. Therefore "illness", "epidemic", "fever" are the three *obvious words* that are used to measure distances to define each mental model.

There are a number of risks r_i, in the set of risks $R = \{r_1, r_2, ..., r_{29}\}$ given in the card sorting experiments. For each mental model there are a set of three obvious words which correspond to that mental model, (w_1, w_2, w_3). For every r_i, the distance D from the mental model j where $M_j = \{w_{j1}, w_{j2}, w_{j3}\}$ is, for the case of experts

$$D_E(M_j, r_i) = \sum_{1 \leqslant n \leqslant 3} d_E(w_{jn}, r_i) \tag{1}$$

Where $D_E(M, r)$ is the distance between w_{jn} and r_i according to the expert distance matrix. We define the non-expert-distance, $D_{NE}(M_j, r_i)$, similarly.

Finally for each risk r_i the expert/non-expert mental models according to the based on the least distance. Suppose that the risk r_i has the following expert or non-expert distances from the obvious mental models:

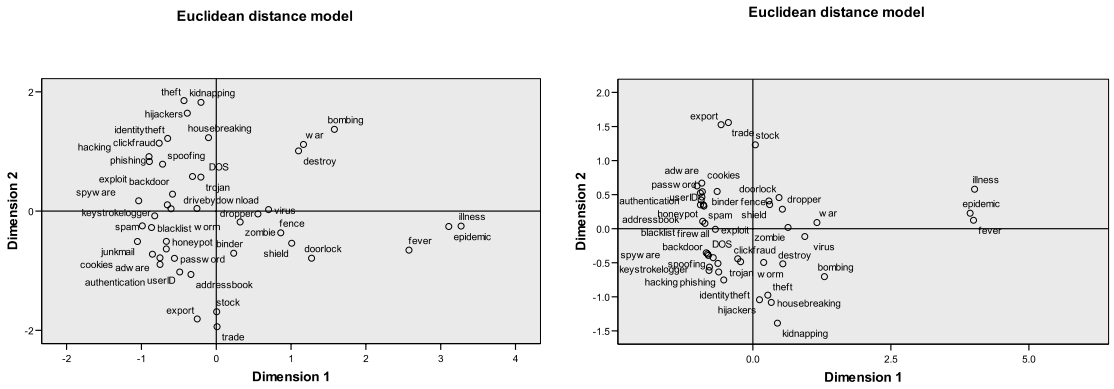

Figure 7: Distance Maps for Weak Experts and for Strong Experts

$$D_{medical} \leqslant D_{criminal} \leqslant D_{physical} \leqslant D_{warfare} \leqslant D_{economic} \tag{2}$$

Then the mental model in the case of r_i would be the medical mental model. This means that the expert/non-expert mental models were assigned to any risk r_i for each data set (WESM, WNSM, SESM, and SNSM) based on the relative distance between the obvious words and the risk r_i. Therefore, in the general case where r_i has distances

$$D_1 \leqslant D_2 \leqslant D_3 \leqslant D_4 \leqslant D_5 \tag{3}$$

the mental model of r_i is M_1 which has the smallest distance, D_1, from r_i.

Figures 7 show the MDS maps of the dissimilarity matrices of all the security and experimentally selected obvious words from E_1(with the weaker definition of expertise) and E_2 (with stronger definition). Notice that those who define themselves as experts isolate warfare and medical models from computer security in the first experiment. Yet in the second case, as the definition of expertise is tightened, warfare is no longer isolated.

The distance maps for all words for non-experts are similarly shown in Figure 8. That is, the left hand graph shows the distances calculated between the obvious words and security words for non-experts when the threshold for self-definition of expert is low. The right hand graph is of the distribution of distances for non-experts when the threshold for self-definition of expert is high. Notice that there is more clear grouping of medical models for the both cases (meaning the medical model is the most distant from security risks).

4.3 Discussion

Recall that the experiments were designed to answer a set of questions. First, do the existing metaphors match mental models of experts or non-expert users? Second, how close are the mental models of experts and non-experts? Third, how sensitive is the difference between expert and non-expert mental models to a definition of expertise? And finally, can we characterize these differences? Our findings made significant progress in answering these questions.

- Neither weak-experts and to a lesser degree experts in E_2 associate the medical mental model with the computer security risks.
- Weak-experts associate neither the warfare nor the economics mental model with computer security risks.
- When the definition of expertise becomes more stringent, the economic and warfare models became associated with the security risks for the experts.
- As the definition of expertise became more stringent the medical model remained the least associated with the security risks, but the distance between the medical model and the computer security risks significantly decreased.

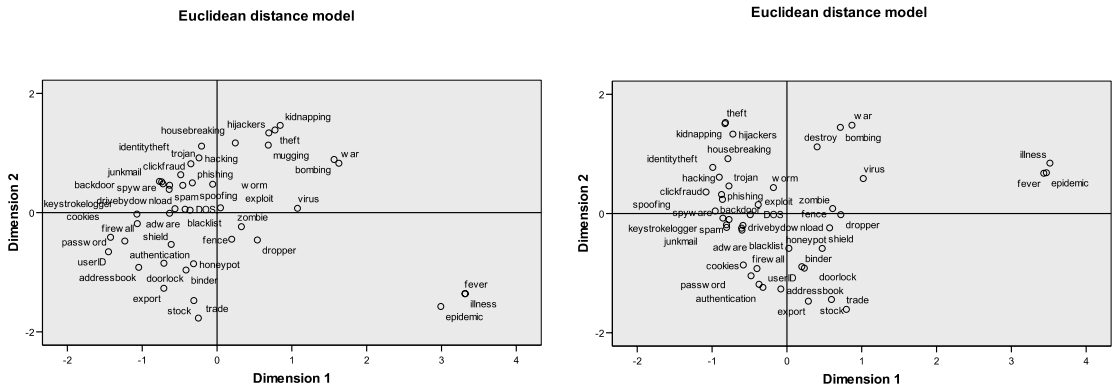

Figure 8: Distance Maps for Non-Experts under the Weak (E_1) and Strong Expertise (E_2) Definitions

Consider the left diagram in Figure 8 of non-experts' association of various words from E_2, and thus putatively mental models. Again, in E_2 (e.g., stronger definition of expert) the medical mental model has the highest distance from the computer security risks and from the other mental models for non-experts. With the exception of the risk "Virus" for strong-non-experts, the medical mental model is not well matched to the mental models of strong-experts and strong-non-experts. We assert that individuals did not associate "Virus" with "computer virus", and believe this was a flaw in the experimental design.

Criminal, physical and economic mental models are closer to computer security risks than other mental models for non-experts. This did not significantly change when the definition of expertise is increased.

The distances between the criminal, physical and warfare mental models and the security risks did not significantly change in the non-expert mapping when the requirement for self-assertion of expertise increases. This implies that non-experts for the purposes of the second experiment that were self-identifying as experts in the first experiment.

As summarized in Figure 6, detailed in Figure 4 and graphically illustrated Figure 7 and Figure 8 weak-experts and non-experts have two different mental models for 13 computer security risks. Similarly Figure 6 illustrates and Figure 5 demonstrates that more stringently defined experts and non-experts have different mental models for 18 computer security risks.

Figure 8 illustrates and Table 3 summarizes that non-experts in both experiments reject warfare as a mental model for computer security risks. Recall that when the definition of computer security expertise was made more stringent, the result was a 17 increase in the selection of the warfare mental model as a correct model for a computer security risk. These facts illustrate that the more stringent definition of expertise resulted in a greater distance between the mental models of expert and non-experts.

Both experiments show a significant difference between experts and non-experts particularly in choosing physical security risks as the appropriate mental model for computer security risks. However, while the difference between the mental models increased with the increased requirements for expertise this was a result of changes in the choices of self-identified experts, not non-experts. In both experiments the non-experts consistently choose either the physical or criminal mental model as related to computer security risks. This suggests that of the mental models embedded in the security literature, the best fits for risk communication are the physical safety and criminal models.

The core argument of this paper is that the communication of security risks to non-expert computer users requires deviating from the mental models of computer security experts. For example, strong-experts mark passwords as corresponding to a criminal model, while non-experts overwhelmingly appear to conceive of passwords as belonging to the realm of physical security. This suggests that non-experts perceive password loss as closer to the risk of an naive loss of a key; while experts perceive passwords loss

as resulting from more malicious activities. The higher standard for self-defined experts resulted in a greater difference overall, not just for passwords. This arguably supports our assertion that the mental models embedded in risk communication be targeted for non-experts rather than based on the models of the communicating experts.

One of the labels given in card sorting experiment was "I can't decide". The participants choose this label for words they didn't know and words that fit no other category. Almost 50 of both weakly defined and strongly defined experts labeled "firewall", "userID" and "cookies" as "I can't decide". This is particularly interesting as the word "firewall" itself indicates conflagration, and warfare was an offered option. The percentage of non-experts who could not categorize "firewall" and "userID' ' in a known category dropped to 40 when the standard for expertise was made more strigent.

The average of "I can't decide" for all the security risks, in the case of experts was 40 and in the case of non-experts was 30.These facts suggest that the five mental models implicit in the security literature do not correlate with the mental models of non-experts *or experts.* Thus moving forward with the mental models from the literature for communication of computer security risks may result in miscommunication to significant elements of the population.

5 Conclusions

This paper reports our initial exploration of the difference between the mental models of security experts and non-experts with regard to computer security risks. To being this work we completed a study of the computer security vocabulary using standard textbooks and our own expertise. We extracted models that had been implicit in security risk communication. Our goals are to identify implicit mental models for computer security, makes these explicit, test these mental models for fit for risk communication, and

use these mental models in a systematic manner to improve risk communication to non-expert computer users. These experiments correspond to the second and third steps in our trajectory towards improving risk communication for computer security.

Our resuls suggest that experts and non-experts have two different mental models for many security risks. With the weak definition of expertise, 45 of the security risks were differently classified by experts and non-experts. With the stronger definition of expertise, 62 of the risks were classified in different mental models by experts and non-experts. The results from the two experiments implies that individual mental models of security is strongly a function of their level of expertise. It is not surprising that this well-noted phenomena occurs in computer security, but the magnitude is notable and the question had not been previously asked in this specific domain. Recall that, due to locality of recruitment, participants included faculty, staff, and students in informatics and computer science. Because of the population it is reasonable to assume that the experiments would tend to underestimate rather than over-estimate the difference between the mental models of experts and lay users in the case of computer security; however, we have no experimental confirmation of this.

Given the results of this research, how can computer security risk communications be aligned with the lay users' mental models? First we determined the implicit mental models in the literature and made them explicit. Second we have used a quantitative approach to evaluate the fit of mental models from the computer security community for experts and non-experts.

Our research agenda is to move forward next with qualitative interviews with experts and non-experts to better understand the mental models associated with computer risk. We have initial designs of risk communication that uses the physical safety and criminal mental models in visual narrative mechanisms. We have completed an initial test with 16

participants that illustrated an expressed desire to change behaviors as a result of these narrative risk communications. We are currently working on devising both larger quantitative experiments and more in-depth qualitative evaluations. We are developing more narratives for risk communication using the physical security and criminal mental models. In parallel, we have proposed a series of interviews of a wide range of users from grade school to those in elder care. (Due to the range of recruitment and the nature of the subjects, the human subjects protection and review is particularly critical for these experiments.)

6 Acknowledgements

We would like to thank professor Youn-Kyung Lim (Indiana University-Bloomington) for her helpful comments on this paper and Christian Briggs (Ph.D. student at Indiana University-Bloomington) for his suggestions on our experiment's interface. We would like to acknowledge to reviewers from WEIS 07 for their suggestions, which significantly improved the paper.

References

[1] A. Acquisti and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. In *Privacy Enhancing Technologies*, pages 36–58, 2006.

[2] R. Adkins. An insurance style model for determining the appropriate investment level. In *Third Workshop on the Economics of Information Security*, Minneapolis, MN, USA, June 2004.

[3] A. Arora, R. Telang, and H. Xu, editors. *Optimal Policy for Software Vulnerability Disclosure*, Minneapolis, MN, 2004. Third Workshop on the Economics of Information Security.

[4] M. Bishop. *Computer Security: Art and Science*. Addison-Wesley Professional, 2003.

[5] A. Bostrom, B. Fischhoff, and M. G. Morgan. Characterizing mental models of hazardous processes: A methodology and an application to radon. *Journal of Social Issues*, 48(4):85–100, 1992.

[6] T. A. Byrd, K. L. Cossick, Cossick, and R. W. Zmud. A synthesis of research on requirements analysis and knowledge acquisition techniques. *MIS Quarterly*, 16(1):117–138, March 1992.

[7] L. J. Camp. Mental models of security. *IEEE Technology and Society*, 2008. accepted.

[8] L. J. Camp and S. Lewis. *The Economics of Information Security*. Kluwer Academic, Boston, MA, 2004.

[9] H. Cavusoglu, B. Mishra, and S. Raghunathan. A model for evaluating it security investments. *Communications of the ACM*, 47(7):87–92, 2004.

[10] J. P. Choi, C. Fershtman, and N. Gandal, editors. *Internet Security, Vulnerability Disclosure, and Software Provision*, Cambridge, MA, 2005. Fourth Workshop on the Economics of Information Security.

[11] L. L. Constantine and L. A. Lockwood. *Software For Use - A Practical Guide to the Models and Methods of Usage Centered Design*. Addison-Wesley, 2000.

[12] T. Cox and M. Cox. *Multidimentional Scaling*. Boca Raton, Florida: Chapman and Hall/CRC, 2001. 2nd edition.

[13] M. Deniszczuk. A temporal investigation into student use of wireless security. In *Informatics Working Paper*. School of Informatics, 2006.

[14] D. Denning. *Information Warfare and Security*. Addison-Wesley Publication, Boston, MA, 1998.

[15] J. Diesner, P. Kumaraguru, and K. M. Carley. Mental models of data privacy and security extracted from interviews with indians. *55th Annual Conference of the International Communication Association*, 2005. New York, NY.

[16] P. Dourish and D. Redmiles. An approach to usable security based on event monitoring and visualization. *Proceedings of the 2002 Workshop on New Security Paradigms*, 2002.

[17] A. B. . R. E.Lofstedt. Communicating risk: Wireless and hardwired. *Risk Analysis*, 23(2):241–248, 2003.

[18] B. Fischhoff, W. B. de Bruin, S. Byram, M. Embrey, and S. Thorne. Mental models of women with breast implants regarding local complications. *Behavioral Medicine*, 27:4–14, 2007.

[19] J. Franklin, V. Paxton, S. Savage, and A. Perrig. An inquiry into the nature and causes of the wealth of internet miscreants. In P. Syverson, editor, *ACM CCS'07*. ACM, 2007.

[20] B. Friedman, D. Hurley, D. C. Howe, H. Nissenbaum, and E. Felten. Users' conceptions of risks and harms on the web: a comparative study. In *CHI '02: CHI '02 extended abstracts on Human factors in computing systems*, pages 614–615, New York, NY, USA, 2002. ACM Press.

[21] D. E. Geer. Security of information when economics matters. *Verdasys, Inc.*, May 2004. available online, at http://www.verdasys.com/resources/resources.html.

[22] D. Golding, S. Krimsky, and A. Plough. Evaluating risk communication: Narrative vs. technical presentations of information about radon. *Risk Analysis*, 12(1):27–35, 1988.

[23] B. J. Hance, C. Chess, and P. M. Sandman. Setting a context for explaining risk. *Risk Analysis*, 9(1):113–117, 1988.

[24] W. Hudson. Playing your cards right: Getting the most from card sorting for navigation design. *HCI & Higher Education Column: People: HCI & the web*, 12(5):56–58, Sep 2005.

[25] H. Jungermann, H. Schutz, and M. Thuring. Mental models in risk assessment: Informing people about drugs. *Risk Analysis*, 8(11):147–155, 1981.

[26] C. Keller, M. Siegrist, and H. Gutscher. The role of the affect and availability heuristics in risk communication. *Risk Analysis*, 26(3):631–639, 2006.

[27] J. Kephart, D. Chess, and S. White. Computers and epidemiology. *IEEE Spectrum*, 30(5):20–26, 1993.

[28] J. Kesan and R. Shah. Establishing software defaults: Perspectives from law, computer science, and behavioral economics. *The Notre Dame Law Review*, 82(2):583–634, 2006.

[29] D. M. Kienzle and M. C. Elder. Recent worms: a survey and trends. In *WORM '03: Proceedings of the 2003 ACM Workshop On Rapid Malcode*, pages 1–10, New York, NY, USA, 2003. ACM Press.

[30] J. Kruskal and M. Wish. *Multidimensional Scaling*. Sage Publication, 1978.

[31] B. Laurie and R. Clayton. Proof-of-work' proves not to work. *Third Workshop on the Economics of Information Security, Minneapolis, MN*, 2004. available online, at http://www.dtc.umn.edu/weis2004/clayton.pdf.

[32] M. G. Morgan, B. Fischhoff, A. Bostrom, and C. J. Atman, editors. *Risk communication*. Cambridge University Press, Cambridge, UK, 2002.

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

[34] J. Nielsen. *Usability Engineering*. Academic Press, San Diego, CA, 1993.

[35] D. Nizovtsev and M. Thursby, editors. *Economic Analysis of Incentives to Disclose Software Vulnerabilities*, Cambridge, MA, 2005. Fourth Workshop on the Economics of Information Security.

[36] D. Norman. *The Design of Everyday Things*. New York: Doubleday/Currency, 1983.

[37] D. Norman. Some observations on mental models. In *Mental Models*, pages 7–14. Lawrence Earlbaum Associates, 1983.

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

[39] S. L. Schensul, J. J. Schensul, and M. D. Lecompte. *Essential Ethnographic Methods*. AltaMira Press, Lanham, MD, 1999.

[40] C. S.Konheim. Risk communication in the real world. *Risk Analysis*, 8(3):367–373, 1988.

[41] P. Slovic. Perception of risk. *Science*, 236(4799):280–285, April 1987.

[42] D. K. Smetters and R. E. Grinter. Moving from the design of usable security technologies to the design of useful secure applications. *Proceedings of the 2002 workshop on New security paradigms*, 2002.

[43] S. W. Smith. Humans in the loop: Human-computer interaction and security. *IEEE Security and Privacy*, 1(3):75–79, 2003.

[44] O. Svenson. Mental models of risk, communications, and action: Reflections on social amplification of risk. *Risk Analysis*, 8(2):199–200, 1988.

[45] R. Telang and S. Wattal, editors. *Impact of Software Vulnerability Announcements on the Market Value of Software Vendors, an Empirical Investigation*, Cambridge, MA, 2005. Fourth Workshop on the Economics of Information Security.