A. Weisbecker, M. Burmester & A. Schmidt (Hrsg.): Mensch und Computer 2015 Workshopband, Stuttgart: Oldenbourg Wissenschaftsverlag, 2015, S. 621-646.

User Perception and Response to Computer Security Warnings

Wolfgang Börger, Luigi Lo Iacono

Cologne University of Applied Sciences, Germany

Abstract

This paper gives necessary foundations to understand the mechanism of warning processing and summarizes the state of the art in warning development. That includes a description of tools, researchers use to work in this scientific field. In detail these are models that describes the human way of processing warnings and mental models. Both are presented detailed with relevant examples. The paper tells how these tools are connected and how they are used to improve the effectiveness of warnings.

1 Introduction

Together with the growth of the digital technologies, there appear many new security threats to computer users like phishing or non-trusted SSL-Certificates. Handling these threats can be quite difficult. Research pointed out that the main reason for a lack of security is wrong human behavior to an appeared warning (Akhawe & Felt 2013). This can be caused by insufficient knowledge, a problem of motivation or wrong beliefs and a disadvantagous and to highliy technical Design of a warning (Cranor 2008). Because designing out the human in decision-making is not always possible (Cranor 2008, Wogalter 2006a), it is important to understand the perception of computer users towards security warnings with the goal to improve the design of warnings, which leads to a higher chance of safe behavior.

This paper aims to create foundations to understand problems around computer warnings, to briefly review already developed methods to solve these problems and to summarize the current state of the art.

2 Foundations

To understand the ongoing chapters, it is important to know about used mechanisms. In detail these are warnings together with the way of humans processing to them and the definition of mental models. This Chapter describes these foundations.

2.1 Physical Warnings

Warnings are created as a communication to users to inform them about hazards and guide them to a safe behavior. As it was mentioned in Chapter 0, human behavior in case of hazardous situations is one of the main reasons for security systems to fail. Because of that, warnings should not be the first step in dealing with hazards. A model for the handling of hazards has been described in the Hazard Control Hierarchy, as illustrated in Figure 1 (Wogalter 2006b).

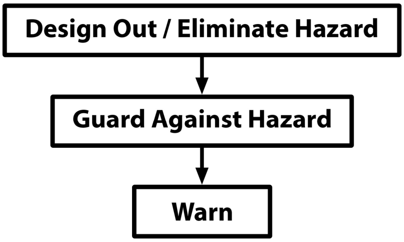

Figure 1: Hazard Control Hierarchy (Wogalter 2006b).)

When a hazard is detected, designing out the hazard should be the first step. For example selling glass bottles during a big event like a football-match will cause a hazard by broken bottles. Instead of placing a warning, a better way would be to replace glass bottles with bottles made out of plastic.

If designing out is not possible, the next step should be to analyze whether the hazard can be guarded or blocked. For example a needed hole in a construction (e.g. for trash) should be surrounded by a fence, giving access only to people, who have to work with it.

In some cases, designing and guarding is not possible. E.g. one cannot remove every harm from a saw without making it useless. In this case, it is necessary to inform the user about the hazard through a warning.

If analyzing the use of the warning proofs that people got injured anyway, a last step can be to completely remove the harming object. E.g. if a company finds out, that their product is not safe, they should withdraw sold products from the customers (Wogalter 2006b).

The main-functions of a warning can be summarized in four mechanisms (Wogalter 2006b).

1. Inform about hazards to make safe behavior possible.
2. Influence the user’s behavior to a safer one.

3. As a result of the second point, reduce or prevent damage.
4. When people work a long time in an environment with constant hazard exposure, a warning can be used to remind people on the presence of the hazard.

The implementation of these mechanisms should be realized by a good design of warnings. It is necessary to get the users' attention. To achieve this, one can use alarming colors, like red, or symbols that almost everyone would intuitively refer to as dangerous. Research found out many rules to improve warnings. An example Warning Design Guideline can be found in the paper from M. S. Wogalter (Wogalter 2006b). The design-recommendations for digital warnings will be discussed in Chapter 4.

One cause for ineffective warnings is the Cry Wolf-Syndrome (also referred to as Warning Fatigue (Akhawe & Felt 2013) or Habituation Effect (Thorley et al. 2001). Sometimes a warning appears without an actual hazardous situation. E.g. a housekeeper cleans the floor of a public building with water and places a warning sign of slippery floor. After the floor has been drying, he forgets to take the sign away. People will see the warning but will not find a hazard. The more often this false warning is perceived, the less attention it will get. Even when this warning later on indicates a real hazard, it is very likely be ignored. Therefore user attention should be defined as a finite resource which is lowered through false warnings. This leads to the result, that the sensitivity of an alarm system is important. Too much sensitivity will produce many wrong warnings and change the human behavior to ignore them (even in hazardous situations). Too little sensitivity can cause damage because of missing warnings (Breznitz 1984).

2.2 Warning Processing

There are many studies on warning improvements in diverse cases. In 2006, M. S. Wogalter developed the Communication-Human Information Processing (C-HIP) Model to organize and structure research literature of warnings (Wogalter 2006a). The model describes the human warning-handling as a result of different stages. The model is illustrated in Figure 2.

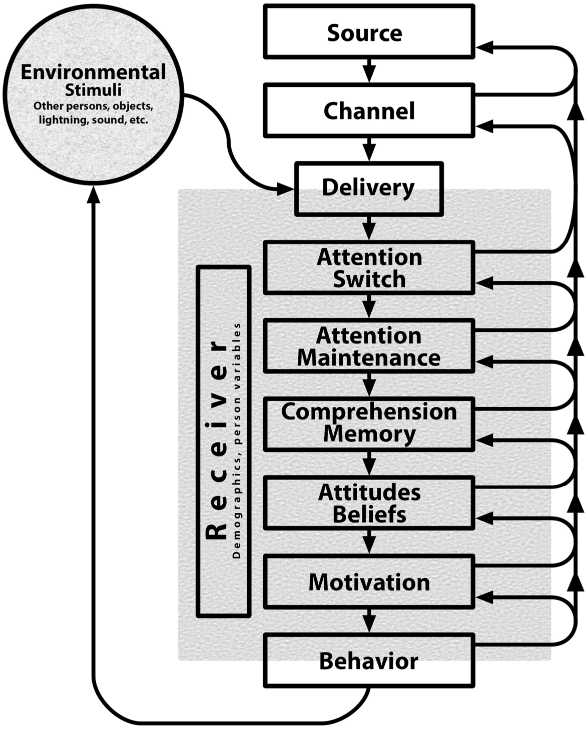

Figure 2: Communication-Human Information Processing (C-HIP) Model (Wogalter 2006a).

Each process is part of a different research-field (which can overlap). The process-chain follows a top-down path. Feedback loops account for later processing stages affecting earlier stages. Furthermore, it is possible, that the processing ends at any stage. Only if the processing reaches the end of the model, a change to safe behavior is possible. An abnormal termination does not necessarily has to be a total failure as it can still result in positive effects like learning.

The model can help researchers to detect bottlenecks inside the processing and to focus limited resources on fixing these bottlenecks rather than working on wrong aspects. In every process, specific techniques are available to examine whether the failure of the warning is due to false accomplishment of the process (Wogalter 2006a). In the following I will explain the steps of this model in detail.

Source: The source has to decide, if a hazard is present and to analyze, if a warning is required based on the following guidelines (Laughery & Wogalter 1997):

1. The hazard cannot be designed out or guarded.
2. The hazard, consequences, and appropriate safe modes of behavior are not known to persons at risk.

3. The hazard is not open and obvious, which means that the appearance of the product or environment does not clearly expose hazards.
4. A reminder is needed to promote awareness of the hazard at proper time.

When the analysis of the described guidelines is positive, the source has to decide, which channel to use. This can have effect on people’s beliefs. For example, an expert explaining the hazard and consequences strengthen the belief in this hazard more than an untrusted channel like an Internet forum would (Wogalter, Kalsher & Rashid 1999).

To exclude the source of having problems, a hazard analysis is necessary (Young 2006). It is important to know, if the system can detect the above mentioned guidelines and if there is information missing about the hazard and the risks (Wogalter 2006a).

Channel: The channel represents the way the warning is send to the end-user. In the physical world, this can be a label printed on the product, a person giving the warning orally to the user or a warning light. It is also possible to use more then one channel simultaneously (Cohen et al. 2006).

To prove the channel, a good way is to ask end-users if they received the warnings. If not, this can be an indicator that it is necessary to think about a better channel (Wogalter 2006a).

Delivery: Even if multiple channels are used, a warning may not reach the end-user. The delivery process describes the condition that the warning has successfully arrived at the target at risk (Wogalter 2006a).

Environment: After arriving at the end-user, the warning has to compete with environmental stimuli, called noise, for the attention of the user. This noise can have many forms. Examples are a ringing telephone, weak light, fog or smoke. This noise can lead to the result that the user classifies the warning non-important and ignores it (Wogalter 2006a).

Receiver and their Variables: The following stages are linked to the targeted person(s) or audience, called the receiver (Wogalter 2006a). The receiver has different variables (Rogers et al. 2000), that form the character of the person. This includes self-efficacy (Lust et al. 1993), locus of control (Donner 1991, Laux & Brelsford 1989), mental workload (Wogalter & Usher 1999), processing strategy (deTurck & Goldhaber 1988) and time-stress (Wogalter, Magurno, Rashid & Klein 1998). Furthermore there are demographic differences, like gender (LaRue & Cohen 1987, Smith-Jackson 2006, Young et al. n.d.) or age (Mayhorn & Podany 2006). These factors will influence all following steps and therefore also the effectiveness of the warning (Young et al. 1999).

It is not completely cleared whether and which variables have a small or big influence on the whole processing (Smith-Jackson 2006, Wogalter 2006a).

Attention Switch: After the delivery the warning has to compete with other stimuli or tasks the user is performing to get his attention (Wogalter 2006a). To reach this goal, warnings should be designed highly salient (Wogalter & Leonard 1999, Wogalter & Vigilante Jr. 2006). A positive result is an attention switch to the warning. Many results from different studies can be used to make warnings as salient as possible. Examples are using high contrast colors compared to the background (Bzostek & Wogalter 1999, Laughery et al. 1993), big letters (Wogalter 2006a) and common symbols (Bzostek & Wogalter 1999). Another significant element is the Cry Wolf-Syndrome, which was mentioned in Chapter 2.1 (Breznitz 1984, Thorley et al. 2001). To prevent this syndrome, different designs for the same warning can be helpful (Wogalter 2006a).

To analyze this process, important questions are if, how and how fast the user perceives the warning. Common techniques include eye-tracking and interviews with users (Wogalter 2006a).

Attention Maintenance: After the user's attention is focused on the warning, this attention needs to stay focused until the user has encoded all information of the warning. The needed time for encoding should be minimized (Wogalter 2006a). Many positive influencing factors can be the same as the ones used for the attention switch. Examples are using short and precise text in simple language and big and clear fonts with high contrast and resolution (ANSI 2002, Frascara 2006, Wogalter et al. 2005). Studies show that the used format has a big influence on the attention maintenance (Desaulniers 1987, Hartley 1994, Wogalter & Vigilante Jr. 2006). Environmental conditions like fog or smoke can weaken the legibility of text and symbols (Collins & Lerner 1982).

Comprehension: The next step is to understand the encoded information of the warning. This means users need the knowledge and ability to connect the encoded information to the situation (Wogalter 2006a). Important components are the usage of understandable symbols and signal words, maybe linked to a warning color. For example, the ANSI (2002) Z535 standard recommends three signal words depending on the situation: DANGER, WARNING and CAUTION, linked to the colors red, orange and yellow respectively (ANSI 2002, Wogalter, Kalsher, Frederick, Magurno & Brewster 1998, Wogalter & Silver 1990). Furthermore the level of knowledge of the user is important. This includes the language skill, the reading ability and technical knowledge. It is important to ensure that the target group is able to understand the whole warning based on the knowledge of the user. Many warnings fail because their content is too technical for the users to understand (Wogalter 2006a).

Testing methods include methodologies like memory tests, open-ended response tests, structured interviews and so on. The results can be used to determine which parts are not understood and further how and where to improve the warning for better understanding (Wogalter 2006a).

Beliefs and Attitudes: The user may have beliefs and attitudes about the warning that are not true. These wrong thoughts can for example be created by experiences they made with similar situations or from hearing from other people in similar situations. These beliefs can lead to ignoring the warning (Goldhaber & DeTurck 1988, Wogalter et al. 1991). To overcome this problem, the warning has to compete successfully against these wrong beliefs and attitudes (Wogalter 2006a). The warning may need to change beliefs a user has. Some mechanisms can be used to increase the success rate (Riley 2006). The greater the consequences and damage in ignoring a warning seem to be, the higher the chance that the user will change his beliefs and attitudes (Wogalter, Young, Brelsford & Barlow 1999). Furthermore, if communicated consequence affects the user in a personal way like injury or even death rather than impersonal like a broken device, the perception of the hazardous situation is much higher (Wogalter et al. 1991). The warning should be personalized (Wogalter et al. 1994). A colorful description of the hazard can be supporting (Wogalter 2006a).

For testing this process, questionnaires are used to analyze pre-existing beliefs and the grade of hazard perception. If the perceived hazard is too low, a greater persuasiveness may be needed (Wogalter 2006a).

Motivation: The motivation determines, whether the will for reacting to the warning is existent or not (Wogalter 2006a). This decision is influenced among other things by the cost of compliance. Time and effort to carry out the behavior is included here (Wogalter et al. 1989, Wogalter et al. 1987). To minimize these effects, warnings should give instructions that are easy to perform (Dingus et al. 1991). Linked to this are the costs of noncompliance. This means the bigger the injury or damage is by ignoring the warning, the higher the motivation to perform the recommended instructions (Wogalter et al. 1991). Furthermore the social influence should be considered (Edworthy & Dale 2000, Wogalter et al. 1989). If the user knows or sees that other people comply with the warning, the likelihood of the user also complying is higher. At last, stress (Wogalter, Magurno, Rashid & Klein 1998) and mental workload (Wogalter & Usher 1999) can prevent the user from reacting to the warning.

Research tries to measure the behavioral intentions. When the complying intentions are small, the warning maybe needs to stronger communicate the consequences while the cost of performing tasks should be reduced (Wogalter 2006a).

Behavior: If the warning was successful in the previous processes then this step describes the user performing the suggested task from the warning that is supposed to prevent him from the hazard.

This process is one of the most important measurement for the warning effectiveness (Kalsher & Williams 2006, Silver & Braun 1999). Researchers mostly measure the compliance likelihood. The measurement is difficult, because researchers cannot put the participants into real danger while the scenario has to let the participants think of an available real risk. Furthermore to create believable scenarios is difficult, because situations that lead to real injury are rare. These tests also cost much time and effort (Wogalter 2006a).

2.3 Mental Models

Kenneth Craik formed the term Mental Model in his work The Nature of Explanation (Craik 1967). A mental model is described as an internal scale-model representation of an extern reality (Craik 1967). Humans understand the world in serial stimuli. These stimuli have to be filtered and then put together to one understandable construct. This created mental model is a reduced view of reality and is then able to be processed by the brain with its limited resources. It is important to notice, that every individual has different mental models based on their experiences with the world. In the same situation, this lets people react in different ways. As a result, changing a mental model can help to change behavior (Johnson-Laird 1983).

In todays psychology, mental models are one way to describe the human reasoning. The model shows one possibility and all different ways, that include this possibility. Mental models are based on the following properties (Davidson et al. 1999):

- They are based on the *Principle of Truth*: Only possible ways are shown.
- They are built *on-the-fly* from knowledge based on prior experiences.
- They are unstable and can change at any times.
- They are used to make decisions in new but similar situations.

Cognitive scientists often use studies about mental models, to get information about processes of the mind. Because the main content is about how humans react in different situations and also to predict this behavior, the results are also used in the scientific field of Artificial Intelligence. Because of the explosive growth of computer usage, mental models are used in Computer-Human Interaction to improve the users understanding and to facilitate the usage of this new complex world (Davidson et al. 1999).

3 Warnings in the Digital World

Almost all features of warnings in the physical world described in Chapter 2.1 can also be applied to digital computer warnings. As mentioned before, warnings are communications to users. In the digital world, apart from the warnings there exist four more types of communications (Bauer et al. 2013, Cranor 2008). They will be described in Chapter 3.1.

Studies have shown that the Cry Wolf-Syndrome has a strong influence on the effectiveness of computer warnings (Akhawe & Felt 2013, Bauer et al. 2013, Cranor 2008). Because the intention of warnings will fail when they are not read by users, it is necessary to determine precisely whether a warning must be shown or not. First step to avoid habituation effects is to use the Hazard Control Hierarchy model, described in Chapter 2.1, to make sure that warnings are only shown as a last resort (Bravo-Lillo et al. 2010, Cranor 2008).

In case designing out or blocking the hazard is not possible, the requirement and the type of the warning can be determined based on two variables. The first one is the probability of occurrence and the second one is the impact of a risk. A perfectly automated warning system should be able to calculate both variables and make a decision on the proportion of the variables as shown in Figure 3 (Bauer et al. 2013).

In zone 1, no warning has to be send due to the low impact of a risk. The impact in the second zone is immense. In this case, the intended action should not be allowed and blocked against the user; there is also no warning needed. In the third zone, the impact is neither high or low, therefore the user might be asked what to do depending on the probability of occurrence (Bauer et al. 2013).

In reality the probability of occurrence is a more objective variable, which is mostly based on external factors. Therefore in most cases, a system should be able to calculate the value of this variable. The impact of a risk is always a subjective variable relative to the user and therefore very hard to estimate by a system. Because of that, L. Bauer *et al.* suggested to always interpret a warning as an implicit question to the user: “How bad it would be if... ?” (Bauer et al. 2013).

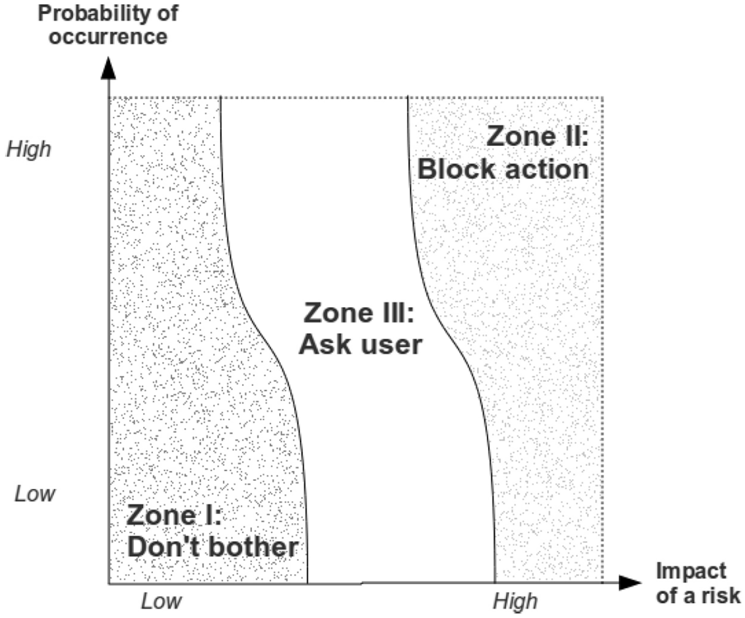

Figure 3: Ideal automatic risk-assessment graph (Bauer et al. 2013).

The next chapters will expand the C-HIP model from Chapter 2 to the needs of the digital world. Mental models are important to be considered for creating effective warnings (Bauer et al. 2013). One model will be explained in detail as well as other work with mental models will briefly be summarized. Next, available studies about warnings will be presented.

3.1 The Human in the Loop-Model

In Chapter 2.2 the C-HIP model was described. L. F. Cranor adapted this model to the needs of digital communications (Cranor 2008). She named her model the human-in-the-loop security framework. Like the C-HIP model, the main usage of the human-in-the-loop model is to find bottlenecks in processing of communications. It is included in the human threat identification and mitigation process pictured in Figure 4. It is used as a four-step iterative process which helps identifying and mitigating human threats to system security. In the *task identification* step the designer identifies all parts of the system, where humans are involved in security questions. In the step *task automation* the designer will try to automate the identified parts completely or at least partially (Cranor 2008). This contains thinking about default settings and whether a human will be able to make better decisions than the default settings or not (Cranor & Garfinkel 2005). Depending on the situation this will not always be possible (Edwards et al. 2008, Flechais et al. 2005). The designer then identifies potential failure modes for remaining security-critical human tasks in the *failure identification* step. In this step, the human-in-the-loop security framework offers a systematic approach to identifying the failure modes (Cranor 2008). In the *failure mitigation* step the designer tries to improve the warnings and to find out how to support the user in reaching a safe behavior (Gross & Rosson 2007). After a first iteration, the designer may find out, that the human failure rates still are very high. In this case, he might want to repeat one or more previous steps to improve the warnings until the human failure rates are acceptable or to proof, whether an automated process, even if not perfect, will get better results (Cranor 2008).

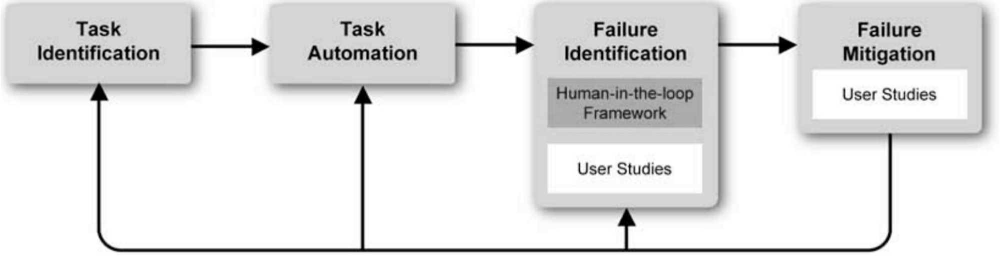

Figure 4: Human thread identification and mitigation process (Cranor 2008).

The *failure identification* step is important to identify failure spots in the warning processing. The human-in-the-loop security framework, pictured in Figure 5, is a good choice to test every step of warning processing. As mentioned before it is based on the *C-HIP* model. It is not a strictly linear process, which means, that steps can be omitted or repeated. The communications are now defined in five different kinds and the environmental stimuli now come along with interferences. The structure of the human receiver differs. Attitudes and beliefs and motivation are not longer inside the process-chain and capabilities are new. The process chain now is divided into three stages each including two steps. In the following part a brief description of the HITL model is given and differences between the C-HIP model will be examined (Cranor 2008, Wogalter 2006a).

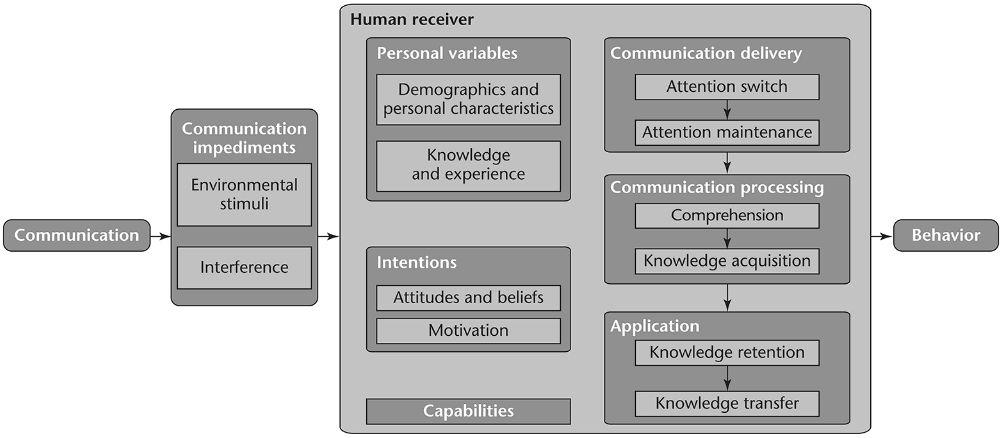

Figure 5: The human-in-the-loop security framework (Cranor 2008)

Communication: The communication is now one of five different types:

- *Warnings* alert users about detected hazards and demand the user to immediately take action to avoid the hazard. They should have a successful design to get through the whole proccessing and in the end lead to a safe behavior.
- *Notices* are used to provide an user with information. This information can be used to make decisions whether there is a hazardous situation or not. Notices are not meant to interrupt the user. Therefore notices should not be used, if a real hazard is detected.
- *Status indicators* are passive messages, that inform users about system status information (E.g. symbols in the taskbar, that inform users about plugged devices or enabled bluetooth). Status indicators do not necessarily ask users to take action.
- *Training material* teaches users about security threats and how to react in given cases. The material can include games, videos, instruction manuals and more. A training is effective, if users not only remember the training, but also apply it in usable situations.
- *Policy communications* define to users, how they have to act in different environments. For example this can include the definition of the structure and symbols a password must have or what types of documents must be encrypted.

A communication can be active, which means it is designed to interrupt the user's primary task and force the user to pay attention, passive, which means they are there but can easily be ignored or something in between active and passive. Secure system designers should consider the frequency the hazard is encountered, the severity of it and the cost the user has to take to perform secure tasks. With this information they should decide, which type of communication should be used (Cranor 2008).

Communication Impediments: Environmental stimuli are the same as in the C-HIP-model like weak light or the user's primary task. Passive designed communications will get more influenced by environmental stimuli then active designed communications. The interference is everything that prevents the user from receiving and noticing the original communication. This contains active attacks or technological failures. In the worst case the user does not even know that a communication was sent (Cranor 2008).

Human Receiver: The human receiver is the user who obtains the communication and might influence the security. He has personal variables, intentions and capabilities which all will influence the progressing-chain containing communication delivery, communication processing and application (Cranor 2008).

Personal Variables: Personal Variables are similar to the ones in the C-HIP-model. Demographic and personal characteristics include age, gender, culture, education, occupation and disabilities. Knowledge and experience also contains education and occupation as well as prior experiences. These variables make it important to design the warning from the sight of the target audience (Cranor 2008).

Intentions: Intentions will influence whether the user will rate the communications as important or not. Attitudes and beliefs as well as motivation exist also in the C-HIP-model. In the human-in-the-loop-model, these variables are not part of the process-chain, but apart from that similar to the one from the C-HIP-model. Attitudes and beliefs contain the belief on accuracy of the communication, the ability to perform the recommended tasks connected to self-efficiency, the belief, whether performing the task will be effective or not connected to response-efficiency, the time the user has to spend and the general attitude to the warning (Cameron & DeJoy 2006). Motivation is the incentive users have to perform the tasks and how proper they will do it. Important factors are risk-perception and conflicting goals (with the primary task of the user) (Kalsher & Williams 2006).

Capabilities: Capabilities are the foundation needed to be able to perform the task. Even if everything is understood, the user will not be able to perform the security task without it. Depending on the situation, this can include physical or cognitive skills as well as needed devices or software. E.g. remembering random strings for passwords can be one ability the user must have to complete tasks (Cranor 2008).

Communication Delivery: Communication delivery contains attention switch and attention maintenance. Both were already described in the C-HIP model in Chapter 2.2. Environmental stimuli, interference, characteristics of the communication (format, font size, length and delivery channel) as well as the Cry Wolf-Syndrome will have an impact on this step (Wogalter & Vigilante Jr. 2006).

Communication Processing: Communication processing includes comprehension and knowledge acquisition. The former was also mentioned in the C-HIP model and describes the ability to understand the warning. It is impacted by the structure and design of the content of the communication, conceptual complexity as well as personal variables (Hancock et al. 2006). Every user must understand the content, which can be difficult to implement because in many cases it is challenging to write about computer security concepts without using technical jargon. Computer warnings often fail on this task. Knowledge acquisition means the ability to perform the understood task practically. Impact properties are exposure or training time, involvement during training and personal variables (Cranor 2008).

Application: This step contains knowledge retention and knowledge transfer. The first term describes the ability to memorize symbols and instructions and to remember the communication if a situation arises in which they need to apply it. Frequency and familiarity of the communication, long term memory abilities and personal characteristics of the user and the level of interactivity of training activities will have an impact to this step. Knowledge transfer is the ability to recognize situations where the communication is applicable and to apply to this situation. It is also impacted by the level of interactivity of training activities and personal variables as well as the degree of similarity between training examples and situations where knowledge should be applied (Cranor 2008).

Behavior: The previous stages will lead to a behavior. The best case is, that the user understands the communication perfectly and performs the security tasks. Failures can occur. Norman defined the Gulf of Execution as the gap between the intention of a user to carry out an action and mechanisms provided by a system to facilitate that action. E.g. the user gets informed that he has to update the anti-virus software but did not get information where and how to do it. To close this gap it is necessary to give clear instructions to users and also to examine the hardware with whom users has to interact with. Norman also defined the Gulf of Evaluation. This means the user has performed the recommended tasks but is not sure, if he did it the right way and if the system is now secure. To avoid the Gulf of Evaluation it is important to give a feedback to the user about the current state (Norman 2002).

3.2 A Mental Model in Detail

To determine the weak spot in warning processing, mental models can help to learn about the perception users have towards warnings and to determine the exact spot of warning failure inside the human in the loop model (Bravo-Lillo et al. 2010). By knowing the weak step in the human in the loop model one also knows the specific variables that can have an impact on the warning processing (as described in Chapter 3.1). After determining the exact variables, the next step of the human threat identification and mitigation process is the failure mitigation, in which the warning may be improved to higher the success rate of the warning (Cranor 2008).

C. Bravo-Lillo *et al.* published a mental model, which was created by analyzing the results of a previous performed user study (Bravo-Lillo et al. 2010). In this study, they split participants in advanced and novice users. They showed the participants warning dialogs and gave a brief scenario in which case this dialog can appear. Then they ask the participants, if they know about the content of the message, if they know, what would happen by ignoring it and which recommendation of processing they would give a friend who has to handle this dialog. The results were then analyzed, similar behaviors were summarized and then the mental model was created. It is illustrated in Figure 6 (Bravo-Lillo et al. 2010).

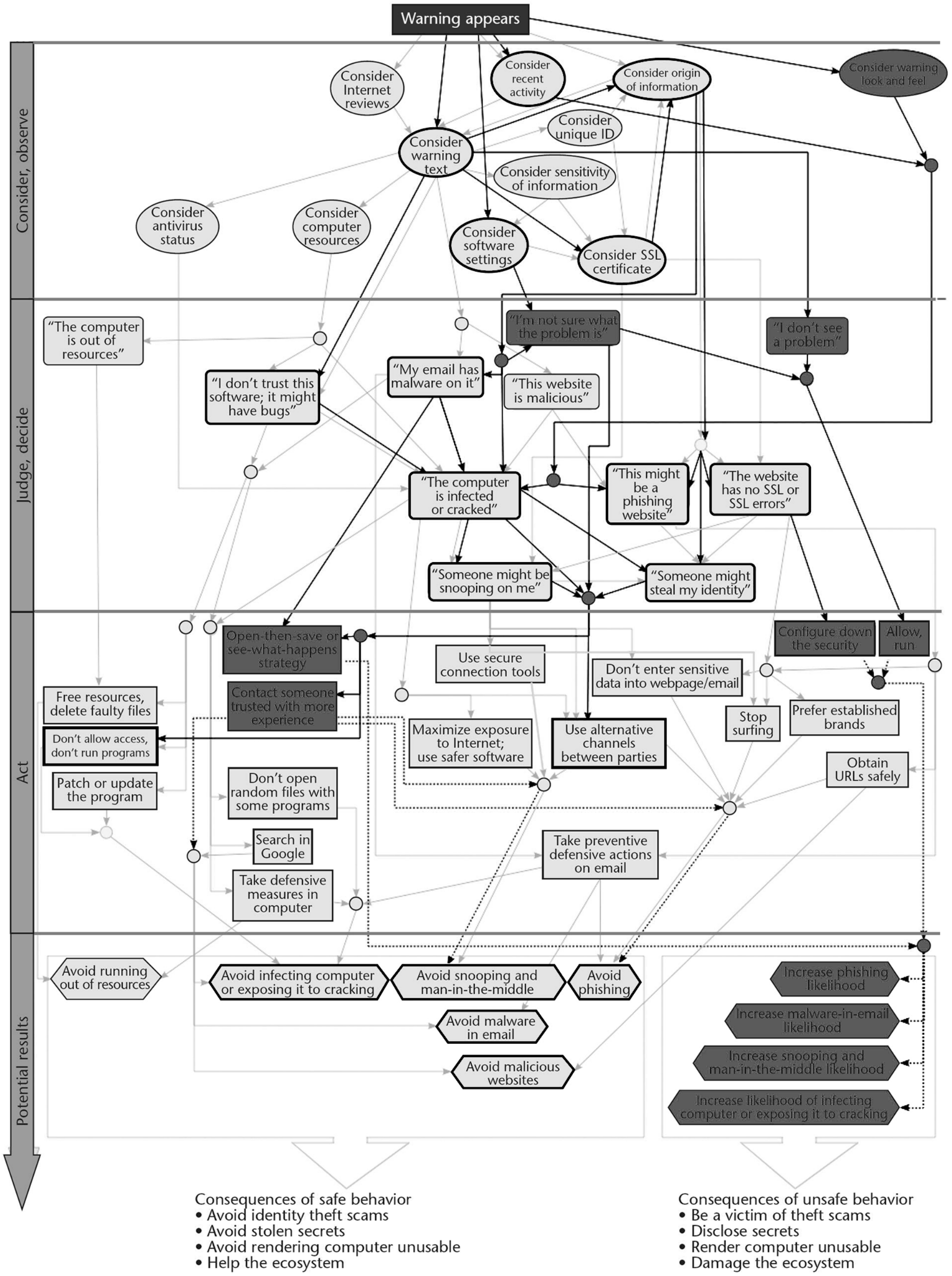
Warning appears
Consider, observe
Judge, decide
Act
Potential results
Consider Internet reviews
Consider recent activity
Consider origin of information
Consider warning look and feel
Consider unique ID
Consider warning text
Consider sensitivity of information
Consider antivirus status
Consider computer resources
Consider software settings
Consider SSL certificate
"The computer is out of resources"
"I'm not sure what the problem is"
"I don't see a problem"
"My email has malware on it"
"I don't trust this software; it might have bugs"
"This website is malicious"
"The computer is infected or cracked"
"This might be a phishing website"
"The website has no SSL or SSL errors"
"Someone might be snooping on me"
"Someone might steal my identity"
Open-then-save or see-what-happens strategy
Contact someone trusted with more experience
Use secure connection tools
Don't enter sensitive data into webpage/email
Configure down the security
Allow, run
Free resources, delete faulty files
Don't allow access, don't run programs
Maximize exposure to Internet; use safer software
Use alternative channels between parties
Stop surfing
Prefer established brands
Patch or update the program
Don't open random files with some programs
Obtain URLs safely
Search in Google
Take preventive defensive actions on email
Take defensive measures in computer
Avoid running out of resources
Avoid infecting computer or exposing it to cracking
Avoid snooping and man-in-the-middle
Avoid phishing
Avoid malware in email
Avoid malicious websites
Increase phishing likelihood
Increase malware-in-email likelihood
Increase snooping and man-in-the-middle likelihood
Increase likelihood of infecting computer or exposing it to cracking
Consequences of safe behavior
• Avoid identity theft scams
• Avoid stolen secrets
• Avoid rendering computer unusable
• Help the ecosystem
Consequences of unsafe behavior
• Be a victim of theft scams
• Disclose secrets
• Render computer unusable
• Damage the ecosystem

Figure 6: The detailed mental model of warning response behaviors. Bright items indicate advanced users' responses and dark items represent novice users' responses. Bright items with a dark outline were mentioned by both (Bravo-Lillo et al. 2010).

The model is split into three main stages of processing which are shown in the red arrow on the left side. First, users observe and consider the warning and its related factors and events to try to understand, what the warning is communicating. Second they try to find out the cause of the warning and decide or judge, which problem, out of severals potential options, they are dealing with. Last they perform one or more actions they believe will solve the problem. If the diagnosis was correct and the behavior was appropriate, the problem will be solved (safe behavior). Otherwise it might persist and create more problems. L. Bauer summarized and simplified the processing of the model as pictured in Figure 7 (Bauer et al. 2013, Bravo-Lillo et al. 2010).

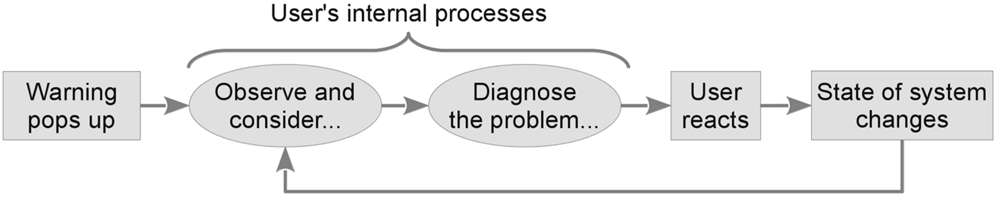

Figure 7: Simplified mental model, showing the sequence of activities performed by users (adapted from Figure 6) (Bauer et al. 2013).

The connecting attribute is the *warning appears* block at the top of the model in Figure 6. From there all found possible ways of processing the warning are shown separated in advanced (bright items) and novice (dark items) users. The model also separates ways that both users mentioned (bright items with a dark outline). Depending on the chosen way it results in a safe or unsafe behavior at the bottom of the model. There is also a summary of the consequences of safe and unsafe behavior. For example, many novice users consider the warning text, but do not see the problem to which this warning is pointing. As a result, they allow the program that caused the warning to run anyway. This increases the likelihood to infect the computer and therefore leads to an unsafe behavior. In the same case after considering the warning text advanced users often did not trust the software and think about bugs the program may have. As a result, they patch or update the program to minimize the likelihood for getting the computer infected or cracked which is a safe behavior. Using this model, one can separate all ways that can lead to an insecure behavior and in a next step the warning can be improved to eliminate the insecure ways (Bravo-Lillo et al. 2010).

It will follow an overview about the results of the mental model, which were also summarized by L. Bauer *et al.* Novice users often get influenced by the look and feel of a warning. They connect almost every warning with viruses. They often consider wrong variables and factors or handle them in wrong order. They also do not have a dynamic image of the state of their computers and do not look in public expert forums that could help dealing with common problems. Often they are not aware of the consequences of their actions. They do not understand most technical terms and get frustrated when being confronted with them instead of gathering more information. Long messages are often ignored. While advanced users often estimate the safety of an option before engaging on it, novice users do that afterwards. Advanced users use forums to inform about states, they update and patch their programs, use safe URLs and scan for viruses. In general, the behavior of advanced users differs strongly from novice users' behavior. C. Bravo-Lillo *et al.* pointed out three main actions novice users take that leads to the consequences of unsafe behavior. First they configure down the computer's global security level. Second they let unknown programs run and third they perform a set of simple strategies. This for example can be to first save files and then look what is inside (can contain viruses) instead of checking them before (Bauer et al. 2013, Bravo-Lillo et al. 2010).

Possible solutions to solve the above named problems can be to write into the warning, that files are already checked for viruses and to present only relevant variables and factors in a meaningful order as well as point out variables and solutions users should not take. Shading offending applications can also be useful. To improve the dynamic image of the state of their computers, it can be useful to log all security related tasks and present parts of it in the warning. One can also present a link to expert forums. To strengthen the perception of consequences one should give an explanation of them for every possible option.Technical terms should be avoided or explained and the text should be short, but useful and understandable. In general it is important, to always develop warnings with the question in mind, if it is possible to misunderstand the warning from the sight of a novice user (Bauer et al. 2013, Bravo-Lillo et al. 2010).

3.3 More Work about Mental Models

R. Wash informed about folk models of home computer security (Wash 2010). He defines this term as mental models of people which are shared among similar members of a culture. This means, they are mostly formed by stories passed on by their friends and colleagues. They are not necessarily correct and right in the real world and can therefore lead to erroneous decision making. He pointed out, that especially in technological context many incorrect folk models exists. Through an user study he classified folk models related to internet security from novice users into two big groups, each with four common mental models inside. The first group are the virus related models and the second the hacker related models. He found out, that his participants do not distinguish between different malicious software and instead calling them all virus. Some participants do not have any particular model of a virus, some interpret them as a buggy software causing crashes, some as programs written by mischievous individuals to cause harm by infecting the computers and some as written programs from criminals who want to get sensitive financial information. He also found out, that his participants use the term hackers for everything, where human agents perform an active attack against a system. Some participants interpret hackers as young students who want to show their skills and causing damage without further reason, some see them as criminals who want to sneak in and get sensitive financial data, some think about them also as criminals, but only aiming at rich people and companies and some mixed up the last two models. Depending on which model the participants believe, they choose different ways to perform security tasks (e.g. the ones who think that hackers only go for rich people are not likely to install a firewall, because they do not see themselves as valuable targets). They also think, that they take a safe behavior based on their beliefs (Wash 2010).

Building on those results R. Walsh *et al.* published another study in which they investigate if and how it is possible to change the mental models of computer users (Wash & Rader 2011).

They pointed out three approaches, that are used to improve computer security. The *stupid user* approach tries to place the human out of the decision-making (like described in the Hazard Control Hierarchy in Chapter 2.1). The *education* approach tries to teach users about security. It can work in companies, but home users will in most cases not have the motivation to perform training tasks. The *understand how users think* approach tries to find out mental models and how to change them in order to improve the security. To change mental models, first it is important to find out how people build them up and how to influence them. The second step is to identify which models are associated with which security behavior to know which models are necessary to change. The conclusion of their work is to accept that mental models mostly are simplified and incorrect. The aim should not be to try to train people, because like mentioned before, in most cases they will not have the motivation to do so or even not the ability to understand all the information. Instead of this the results of the models are more important. This means, as long as the models lead to valuable security behaviors, even incorrect models are worth and one should try to change the folk models into these models (Wash & Rader 2011).

L. J. Camp created mental models in order to use them as a framework for communicating complex security risks to the general populace (Camp 2006). She did not invent the models to get insights in their thinking but constructed five different models, that can be useful to explain complex topics to standard users. These models take the form of analogies or metaphors with other similar situations. The physical security model tries to explain the internet security with physical metaphors like a neighborhood watch which is difficult due to the big differences between the physical and the digital world. The medical model can be used to explain the necessity of keeping security systems up to date and that the first infected computer has not to be the victim but can be used to attack other computers. Also the responsibility can be taught. The criminal model explains that users are responsible for their own security (maybe supported by experts). The problem is, that users are not always able to protect themselves. The market model explains the costs of performing as well as not performing security tasks to higher the perception and motivation for security (Camp 2006).

Based on the results from L. J. Camp F. Asgharpour et al. built up a user study to determine the difference of behavior between novice and advanced users (Asgharpour et al. 2007). For their study they used the card sort technique. This is a structured elicitation technique done by requiring a subject to sort a pile of cards with words written on them into different piles. The results mirror the ones from the study from C. Bravo-Lillo *et al.* presented in Chapter 3.2 (Bravo-Lillo et al. 2010). They pointed out, that most of the novice users have different mental models than advanced users about the same tasks (Asgharpour et al. 2007).

J. Blythe *et al.* designed a network security test bed, where they implement previously identified models in agents to show that the implementations produce behaviors similar to those of users who hold these models (Blythe & Camp 2012). For testing they used models from R. Walsh’s work first mentioned in this chapter (Wash 2010). The test proved that the agents are able to predict human behavior. They want to make this model simulator available for researchers to test the impact of human behavior on security tools, to share the results and to compare the outcomes of the mental models (Blythe & Camp 2012).

3.4 Available Studies about Warnings

In their work from 2005 T. Whalen *et al.* used eye-tracking software to test the effectiveness of passive indicators used for SSL-Warnings (Whalen & Inkpen 2005). These are used for example as small lock or key icons in the URL bar. Often browsers change the color of these icons to show that websites have an Extended Validation (EV) SSL Certificate. None of the participants did look at the icons in the URL bar (Whalen & Inkpen 2005). Other studies came to the same result. The participants of C. Jackson's *et al.* study did not find the EV indicators helpful in order to identify hazardous websites (Jackson et al. 2007). J. Sobey *et al.* came to the result, that the EV indicators do not influence the decision-making of the participants (Sobey et al. 2008).

Many studies determine click-through rates[1] to test the effectiveness of warnings. Often laboratory studies are used where participants get a task and during completing it getting confronted with warnings. In 2009 J. Sunshine *et al.* performed a laboratory study where participants were confronted with SSL warnings during performing information lookup tasks. Depending on the design of the warning (amongst other they used Mozilla Firefox 2 with a modal[2] warning and Firefox 3 with a separate warning site) the click-through rates differ. By Firefox 2, the rate was 90% and by Mozilla Firefox 3 it was 50% when trying to access their bank websites. When visiting the library site the rates increased to 95% and 60% (Sunshine et al. 2009). Similar high rates were determined by A. Sotirakopoulos *et al.* in a similar study in 2011. While using Mozilla Firefox 3.5 the rate was at 80% and using Internet Explorer 7 it was 72%. These high click-through rates lead to the conclusion, that many people do not deal with warnings. In the second study, many participants said afterwards that they felt safe due to the laboratory environment. This could have an impact on the results (Sotirakopoulos et al. 2011).

In 2013 D. Akhawe *et al.* found contrastive results (Akhawe & Felt 2013). They performed a large scale field study in which 25 million samples of warning handling were collected through telemetry data[3] from Google Chrome and Mozilla Firefox. They determined click-through rates from malware warnings which were at 7,2% and 23,2% for Mozilla Firefox and Google Chrome respectively, phishing warnings which were at 9,2% and 18,0% and SSL warnings which were at 33% and 70,2%. Except for the SSL warnings at the Google Chrome browser all rates were low enough to enhance the security of users. The high rate for SSL warnings at the Google Chrome browser can be explained by a different appearance, by using certificate pinning[4] and by the missing ability to save chosen settings. They pointed out different reasons for the differences of the click-through rates from previous studies. First like mentioned before most studies were performed in a laboratory environment which

[1] The rate in percent in which users click through the warning without giving attention.

[2] no other interaction with any other application is possible until the dialog has been dismissed

[3] Anonymous data that can optionally be send by users by activating them in the options from the browsers.

[4] warnings triggered by sites on an internal maintained list cannot be clicked through and were not considered in the study

produces a feeling of safety to the participants and thus leads to a different behavior. Second most studies were made in the past. The design of warnings got improved over time. Some changes were directly motivated by published user studies. The improved warnings are more efficient in getting users attention. To lower the impact of the first reason they conclude to use add-ons for browsers to collect data instead of using an laboratory environment. They found out, that warnings are able to get users attention and that they deal with warnings. Therefore warnings are an effective instrument for computer security (Akhawe & Felt 2013).

4 Designing Computer Warnings

The last chapters offered insight to how knowledge about human reasoning and perception about warnings is received. This knowledge can be considered by warning designers when designing new warnings as well as improving existing ones. L. Bauer *et al.* established six warning design guidelines based on the results of studies and the discovered mental models elucidated in the previous chapters as well as some others (Bauer et al. 2013).

Describe the risk comprehensively: Any warning must include information about the risk, consequences of ignoring and instructions to make safe behavior possible. If one of this information is missing, the user will underestimate the risk or will not know, how to comply. This will lead in a high chance to ignore the warning (Egelman 2009, Wogalter 2006b).

Be concise and accurate: Text presented in warnings must be as short as possible yet informative enough to comply with it. It should be written from the perspective of the user without using technical terms. Ambiguous terms should be avoided and the warning should be polite, supportive and encouraging instead of being offensive (Apple human interface guidelines n.d., Egelman 2009, Gnome human interface guidelines n.d., Nodder 2005**,** Windows User Experience Interaction Guidelines n.d., Wogalter 2006b).

Offer meanigful options: Warnings always must ask for a decision about at least two options (otherwise it should be labeled as a notification or a status indicator; see Chapter 3.1). After reading the warning the user must be able to make a safe decision and also understand all available options. There should always be a safe choice which is marked as a default value. The safest option should always be above all other options or in the lower right corner (when the options are presented in a row). Furthermore the close action should not be labeled as *Ok*, *Close* or *Cancel*. Better choices are *Ignore this warning* or *Cancel the update*. The user will then be more aware of the closing (Apple human interface guidelines n.d., Cranor 2008, Egelman 2009, Egelman et al. 2008, Gnome human interface guidelines n.d., Nodder 2005, Windows User Experience Interaction Guidelines n.d., Wogalter 2006b).

Present relevant contextual information: When a decision has to be made whether to run an unknown or unverified application, it is important to give detailed information about this application. This should include the name, execution path, if it was checked by an anti-virus program and the current state of the anti-virus program. When handling with an agent, presented information should contain if and which data will be transferred (especially if there is any sensitive data), the identification of the agent, in which term the data will be used and if the data will be stored. If the hazard includes acting with an insecure channel, the system should scan all information, that will be send for personally identifiable information. If nothing is found, the user should be asked to rate the sensitivity of the data (e.g. in presenting a scale). Only if personally identifiable information is found or the rate of the sensitivity is high a warning should be presented (Bauer et al. 2013, Bravo-Lillo et al. 2010).

Present relevant auditing information: A warning system should record the user's interactions within the environment which the warning system is a part of. This includes who, what, when and what for accesses are granted to information. When showing a warning, parts of the records, that are relevant to the hazard, should be presented. If the information is too big, a link to a log-file should be placed within the warning (Bauer et al. 2013, Bravo-Lillo et al. 2010).

Follow a consistent layout: Based on the upcoming information about designing the look and feel of warnings, L. Bauer *et al.* have designed a structure for warnings. It is shown in Figure 8. When presenting critical warnings, there should not be a *close* button on the upper right corner, the screen should be shaded and the warning should be modal. Like shown in mark one of the picture, the warning should include only one icon conveying the level of urgency. Mark two is about the primary text, which should be presented in big letters and should not be longer then one sentence while the secondary text, shown at mark three, when needed, should be in smaller letters and give further explanations to the content of the primary text. The secondary text may be initially hidden and only presented, when the user presses a *more information* button. Directly above the options should be an explicit question to the user like presented in mark four. The options should all be possible answers to this question. Instead of using buttons, command links are a better choice[5]. Each command link should contain a brief description of its usage in big font and if necessary a brief explanatory text in small font below the description. The warning may include primary options, which are important for safe behavior, and secondary options, which can contain a *help* or an *ignore this warning* button. An example is shown by mark five and six. Technical terms, when not avoidable, should be marked as links and open small pop-ups with explanations or definitions of the terms (Apple human interface guidelines n.d., Bauer et al. 2013, Bravo-Lillo et al. 2010, Egelman 2009, Egelman et al. 2008, Gnome human interface guidelines n.d., Nodder 2005, Windows User Experience Interaction Guidelines n.d., Wogalter 2006b).

[5] Command links are part of the Windows User Experience Interaction Guidelines [71]

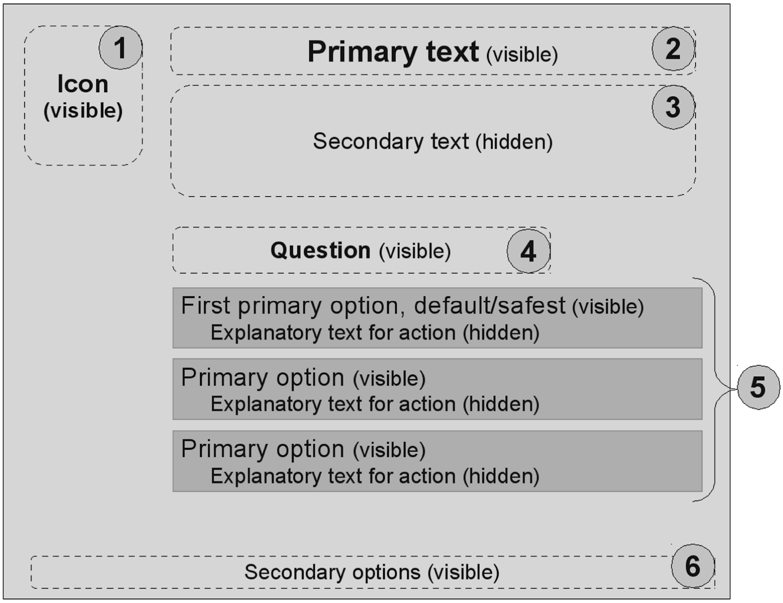
1
Icon (visible)
Primary text (visible)
2
3
Secondary text (hidden)
Question (visible)
4
First primary option, default/safest (visible)
Explanatory text for action (hidden)
Primary option (visible)
Explanatory text for action (hidden)
Primary option (visible)
Explanatory text for action (hidden)
5
Secondary options (visible)
6

Figure 8: Suggested warning dialog layout (Bauer et al. 2013).

5 Summary and Conclusion

This paper summarized the state of the art in warning design and user perception. This includes necessary foundations to understand the mechanism of warnings and introducing a model which can be used to understand how humans process warnings. Different steps of the human threat identification and mitigation process in which the warning processing is a part of were described. There exist many variables that have to be considered when designing warnings. Mental models can effectively be used to detect the step where the warning processing fails and also which variables have to be changed. Depending on this, built up warning guidelines can be used when designing warnings. Studies have been summarized that pointed out how important the design as well as the content of warnings are. With all this knowledge, warnings got improved heavily the last years. The click-through rates of current studies show, that warnings are an effective way to improve computer security, when designing out or guarding against the hazard is not possible. Many studies were done with only a small group of participants. Therefore further studies are needed to prove the results and also to improve warning effectiveness even more.

References

Akhawe, D. & Felt, A. P. (2013), Alice in warningland: A large-scale field study of browser security warning effectiveness., *in* ‘Usenix Security’, pp. 257–272.

ANSI (2002), ‘Accredited standards committee on safety signs and colors - z535.1-5’, *National Electrical Manufacturers* .

Apple human interface guidelines (n.d.). https://developer.apple.com/library/mac/documentation/-UserExperience/Conceptual/OSXHIGuidelines/ [assessed 21-Dec-2014]

Asgharpour, F., Liu, D. & Camp, L. J. (2007), Mental models of computer security risks., *in* 'WEIS', Citeseer.

Bauer, L., Bravo-Lillo, C., Cranor, L. F. & Fragkaki, E. (2013), 'Warning design guidelines', *CMU-CyLab-13-002* .

Blythe, J. & Camp, L. J. (2012), Implementing mental models, *in* 'Security and Privacy Workshops (SPW), 2012 IEEE Symposium on', IEEE, pp. 86–90.

Bravo-Lillo, C., Cranor, L. F., Downs, J. S. & Komanduri, S. (2010), 'Bridging the gap in computer security warnings: A mental model approach', *Security & Privacy, IEEE* .

Breznitz, S. (1984), *Cry wolf: The psychology of false alarms*, Psychology Press.

Bzostek, J. A. & Wogalter, M. S. (1999), Measuring visual search time for a product warning label as a function of icon, color, column and vertical placement, *in* 'Proceedings of the Human Factors and Ergonomics Society Annual Meeting', Vol. 43, SAGE Publications, pp. 888–892.

Cameron, K. & DeJoy, D. (2006), 'The persuasive functions of warnings: Theory and models', *Handbook of Warnings. Lawrence Erlbaum Associates, Mahwah, NJ* pp. 301–312.

Camp, L. J. (2006), 'Mental models of privacy and security', *Available at SSRN 922735* .

Cohen, H. H., Cohen, J., Mendat, C. C. & Wogalter, M. S. (2006), 'Warning channel: Modality and media', *Handbook of warnings* pp. 123–134.

Collins, B. L. & Lerner, N. D. (1982), 'Assessment of fire-safety symbols', *Human Factors: The Journal of the Human Factors and Ergonomics Society* 24(1), 75–84.

Craik, K. (1967), *The nature of explanation*, CUP Archive.

Cranor, L. F. (2008), 'A framework for reasoning about the human in the loop.', *UPSEC* 8, 1–15.

Cranor, L. F. & Garfinkel, S. (2005), *Security and usability: designing secure systems that people can use*, "O'Reilly Media, Inc.".

Davidson, M. J., Dove, L. & Weltz, J. (1999), 'Mental models and usability', *Depaul University, Cognitive Psychology* 404.

Desaulniers, D. R. (1987), Layout, organization, and the effectiveness of consumer product warnings, *in* 'Proceedings of the Human Factors and Ergonomics Society Annual Meeting', Vol. 31, SAGE Publications, pp. 56–60.

deTurck, M. A. & Goldhaber, G. M. (1988), 'Consumers' information processing objects and effects of product warning', *Proceedings of the Human Factors Society* (32), 445–449.

Dingus, T. A., Hathaway, J. A. & Hunn, B. P. (1991), A most critical warning variable: Two demonstrations of the powerful effects of cost on warning compliance, *in* 'Proceedings of the Human Factors and Ergonomics Society Annual Meeting', Vol. 35, SAGE Publications, pp. 1034–1038.

Donner, K. A. (1991), 'Prediction of safety behaviors from locus of control statements', *Proceedings of Interface* 91, 94–98.

Edwards, W. K., Poole, E. S. & Stoll, J. (2008), Security automation considered harmful?, *in* 'Proceedings of the 2007 Workshop on New Security Paradigms', ACM, pp. 33–42.

Edworthy, J. & Dale, S. (2000), Extending knowledge of the effects of social influence in warning compliance, *in* 'Proceedings of the Human Factors and Ergonomics Society Annual Meeting', Vol. 44, SAGE Publications, pp. 107–110.

Egelman, S. (2009), *Trust me: Design patterns for constructing trustworthy trust indicators*, ProQuest.

Egelman, S., Cranor, L. F. & Hong, J. (2008), You've been warned: an empirical study of the effectiveness of web browser phishing warnings, *in* 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', ACM, pp. 1065–1074.

Flechais, I., Riegelsberger, J. & Sasse, M. A. (2005), Divide and conquer: the role of trust and assurance in the design of secure socio-technical systems, *in* 'Proceedings of the 2005 workshop on New security paradigms', ACM, pp. 33–41.

Frascara, J. (2006), 'Typography and the visual design of warnings', *Handbook of Warnings* pp. 385–406.

Gnome human interface guidelines (n.d.). https://developer.gnome.org/hig/stable/index.html.en [assessed 21-Dec-2014]

Goldhaber, G. M. & DeTurck, M. A. (1988), 'Effects of consumers' familiarity with a product on attention to and compliance with warnings', *Journal of Products Liability* 11(1), 29–37.

Gross, J. B. & Rosson, M. B. (2007), Looking for trouble: understanding end-user security management, *in* 'Proceedings of the 2007 Symposium on Computer Human interaction For the Management of information Technology', ACM, p. 10.

Hancock, H. E., Bowles, C. T., Rogers, W. A. & Fisk, A. D. (2006), 'Comprehension and retention of warning information', *Handbook of Warnings. Lawrence Erlbaum Associates, Mahwah, NJ* pp. 267–277.

Hartley, J. (1994), *Designing instructional text*, 3rd edn, Routledge.

Jackson, C., Simon, D. R., Tan, D. S. & Barth, A. (2007), An evaluation of extended validation and picture-in-picture phishing attacks, *in* 'Financial Cryptography and Data Security', Springer, pp. 281–293.

Johnson-Laird, P. N. (1983), *Mental models: Towards a cognitive science of language, inference, and consciousness*, number 6, Harvard University Press.

Kalsher, M. J. & Williams, K. J. (2006), 'Behavioral compliance: Theory, methodology, and results', *Handbook of warnings* pp. 313–331.

LaRue, C. & Cohen, H. H. (1987), Factors affecting consumers' perceptions of product warnings: An examination of the differences between male and female consumers, *in* 'Proceedings of the Human Factors and Ergonomics Society Annual Meeting', Vol. 31, SAGE Publications, pp. 610–614.

Laughery, K. R. & Wogalter, M. S. (1997), 'Warnings and risk perception', *Handbook of human factors and ergonomics* 2, 1174–1197.

Laughery, K. R.and Vaubel, K. P., Young, S. L., Brelsford Jr., J. W. & Rowe, A. L. (1993), 'Explicitness of consequence information in warnings', *Safety science* 16(5), 597–613.

Laux, L. & Brelsford, J. W. (1989), 'Locus of control, risk perception, and precautionary behavior', *Proceedings of Interface* 89, 121–124.

Lust, J. A., Celuch, K. G. & Showers, L. S. (1993), 'A note on issues concerning the measurement of self-efficacy1', *Journal of Applied Social Psychology* 23(17), 1426–1434.

Mayhorn, C. B. & Podany, K. I. (2006), 'Warnings and aging: Describing the receiver characteristics of older adults', *Handbook of Warnings* pp. 355–361.

Nodder, C. (2005), 'Users and trust: A microsoft case study', *Security and Usability* pp. 589–606.

Norman, D. A. (2002), *The design of everyday things*, Basic books.

Riley, D. M. (2006), 'Beliefs, attitudes, and motivation', *Handbook of warnings* pp. 289–300.

Rogers, W. A., Lamson, N. & Rousseau, G. K. (2000), 'Warning research: An integrative perspective', *Human Factors: The Journal of the Human Factors and Ergonomics Society* 42(1), 102–139.

Silver, N. C. & Braun, C. C. (1999), 'Behavior', *Warnings and risk communication* pp. 245–262.

Smith-Jackson, T. L. (2006), 'Receiver characteristics', *Handbook of warnings* pp. 335–344.

Sobey, J., Biddle, R., van Oorschot, P. & Patrick, A. (2008), 'Exploring user reactions to new browser cues for extended validation certificates'.

Sotirakopoulos, A., Hawkey, K. & Beznosov, K. (2011), On the challenges in usable security lab studies: Lessons learned from replicating a study on ssl warnings, *in* 'Proceedings of the Seventh Symposium on Usable Privacy and Security', ACM, p. 3.

Sunshine, J., Egelman, S., Almuhimedi, H., Atri, N. & Cranor, L. F. (2009), Crying wolf: An empirical study of ssl warning effectiveness., *in* 'USENIX Security Symposium', pp. 399–416.

Thorley, P., Hellier, E. & Edworthy, J. (2001), 'Habituation effects in visual warnings', *Contemporary ergonomics* pp. 223–230.

Wash, R. (2010), Folk models of home computer security, *in* 'Proceedings of the Sixth Symposium on Usable Privacy and Security', ACM, p. 11.

Wash, R. & Rader, E. (2011), Influencing mental models of security: a research agenda, *in* 'Proceedings of the 2011 workshop on New security paradigms workshop', ACM, pp. 57–66.

Whalen, T. & Inkpen, K. M. (2005), Gathering evidence: use of visual security cues in web browsers, *in* 'Proceedings of Graphics Interface 2005', Canadian Human-Computer Communications Society, pp. 137–144.

Windows User Experience Interaction Guidelines (n.d.). http://msdn.microsoft.com/en-us/library/-Aa511258.aspx [assessed 21-Dec-2014]

Wogalter, M. S. (2006*a*), 'Communication-human information processing (c-hip) model', *Handbook of warnings* pp. 51–61.

Wogalter, M. S. (2006*b*), 'Purposes and scope of warnings', *Handbook of Warnings* pp. 3–9.

Wogalter, M. S., Allison, S. T. & McKenna, N. A. (1989), 'Effects of cost and social influence on warning compliance', *Human Factors: The Journal of the Human Factors and Ergonomics Society* 31(2), 133–140.

Wogalter, M. S., Brelsford, J. W., Desaulniers, D. R. & Laughery, K. R. (1991), ‘Consumer product warnings: The role of hazard perception’, *Journal of Safety Research* 22(2), 71–82.

Wogalter, M. S., DeJoy, D. & Laughery, K. R. (2005), *Warnings and risk communication*, CRC Press.

Wogalter, M. S., Godfrey, S. S., Fontenelle, G. A., Desaulniers, D. R., Rothstein, P. R. & Laughery, K. R. (1987), ‘Effectiveness of warnings’, *Human Factors: The Journal of the Human Factors and Ergonomics Society* 29(5), 599–612.

Wogalter, M. S., Kalsher, M. J., Frederick, L. J., Magurno, A. B. & Brewster, B. M. (1998), ‘Hazard level perceptions of warning’, *International Journal of Cognitive Ergonomics* 2(1-2), 123–143.

Wogalter, M. S., Kalsher, M. J. & Rashid, R. (1999), ‘Effect of signal word and source attribution on judgments of warning credibility and compliance likelihood’, *International Journal of Industrial Ergonomics* 24(2), 185–192.

Wogalter, M. S. & Leonard, S. D. (1999), ‘Attention capture and maintenance’, *Warnings and risk communication* pp. 123–148.

Wogalter, M. S., Magurno, A. B., Rashid, R. & Klein, K. W. (1998), ‘The influence of time stress and location on behavioral warning compliance’, *Safety Science* 29(2), 143–158.

Wogalter, M. S., Racicot, B. M., Kalsher, M. J. & Simpson, M. J. (1994), ‘The role of perceived relevance in behavioral compliance in personalized warning signs’, *International Journal of Industrial Ergonomics* 14(3), 233–242.

Wogalter, M. S. & Silver, N. C. (1990), ‘Arousal strength of signal words’, *Forensic Reports* 3(407-420).

Wogalter, M. S. & Usher, M. O. (1999), Effects of concurrent cognitive task loading on warning compliance behavior, *in* ‘Proceedings of the Human Factors and Ergonomics Society Annual Meeting’, Vol. 43, SAGE Publications, pp. 525–529.

Wogalter, M. S. & Vigilante Jr., W. J. (2006), ‘Attention switch and maintenance’, *Handbook of warnings* pp. 245–266.

Wogalter, M. S., Young, S. L., Brelsford, J. W. & Barlow, T. (1999), ‘The relative contributions of injury severity and likelihood information on hazard-risk judgments and warning compliance’, *Journal of Safety Research* 30(3), 151–162.

Young, S. L. (2006), ‘Hazard analysis as part of the safety information development process’, *Handbook of Warnings* pp. 431–436.

Young, S. L., Laughery, K. R., Wogalter, M. S., Lovvoll, D. R., Karwowski, W. & Marras, W. S. (1999), ‘Receiver characteristics in safety communications’, *The occupational ergonomics handbook* pp. 693–706.

Young, S. L., Martin, E. G. & Wogalter, M. S. (n.d.), ‘Gender differences in consumer product hazard perceptions’.

Contact Information

Wolfgang Börger (wolfgang.boerger@fh-koeln.de)
Luigi Lo Iacono (luigi.lo_iacono@fh-koeln.de)