5-5-2022

Security Warning Messages Research: Past and Future

Dustin Ormond
Creighton University, dustinormond@creighton.edu

Jordan Barlow
St. Thomas University, jordan.barlow@stthomas.edu

Security Warning Messages Research: Past and Future

Dustin Ormond
Creighton University
dustinormond@creighton.edu

Jordan B. Barlow
University of St. Thomas
jordan.barlow@stthomas.edu

ABSTRACT

Research on the effects of IT security warning messages has increased in the last several years. Most studies empirically examining such warning messages chiefly focus on warning content and/or aesthetics and their effects on attention and/or behavior. Many of these studies cite the Communication-Human Information Processing (C-HIP) model as a foundation, yet this model includes other important and under-researched constructs, including perceptions of the source of a message, comprehension of a message, attitudes and beliefs, and fear. In this study, we performed a comprehensive literature review of empirically published studies on IT security warning messages. We propose a comprehensive theoretical model that entails both C-HIP and Protection Motivation Theory. We then categorize our catalog of IT security warning message research papers according to which propositions in our model have been previously studied. We focus specifically in this paper on those under-researched areas that provide opportunities for future research.

Keywords

Warning messages, IT security, literature review, C-HIP model, protection motivation.

INTRODUCTION

Despite existing IT security countermeasures aimed at protecting technology users (e.g., antivirus software, firewalls), user behavior represents the last line of defense against malicious actions. Warnings represent communication designed to prevent users from hurting themselves or others (Wogalter 2006b). While effective in many cases, warnings are often ignored by users, either intentionally or due to habituation (Akhawe and Felt 2013; Vance et al. 2018).

Research on computer warning messages indicates that HCI elements are integral parts of these messages (e.g., Bravo-Lillo et al. 2011b), yet many of these studies do not account for content of the message or whether users ignore the message for other reasons. Even research focusing on design of warnings to avoid habituation (Anderson et al. 2014a; Anderson et al. 2014b) do not fully account for the fact that people may reject them based on their content. Taken together, existing studies largely only examine warning content, aesthetics, or attention in isolation; even those studies accounting for all three do not also consider other relevant factors such as perceptions of the source of the message, attitudes and beliefs about the message, fear generated from the message, or motivation to comply.

We completed a thorough literature review to catalog all existing empirical research on computer security warning messages and view the existing knowledge in one comprehensive whole. As a foundation for this research, we use the Communication-Human Information Processing (C-HIP) Model, but also supplement the C-HIP Model with fear appeals (a key element from Protection Motivation Theory) to understand this research in the digital, IT security context.

THEORETICAL BACKGROUND

Researchers of physical warning messages have used the Communication-Human Information Processing (C-HIP) model frame their research (Conzola and Wogalter 2001; Wogalter 2006a). This framework shows that when receiving a message (such as a warning), we must consider the source, channel, and receiver. *Source* refers to the person or entity delivering the message. In the case of information security warnings, the source is often hidden from the user, who only sees warning messages as appearing on the screen. In this context, perceptions toward the source (e.g., trust) may be more important than the literal source itself. *Channel* is the method of delivering the communication (the warning itself). In addition to the actual text or meaning of the message, a channel also consists of visual aspects (e.g., color, size).

The C-HIP model indicates that the *receiver* (i.e., the person toward whom the warning message is addressed) must go through several stages before the information can have a full effect. The first is *attention*. If receivers are not paying attention to the message, it cannot have any further impact on their behavior. The next step is *comprehension*, which refers to an individual's level of understanding of the message itself and the consequences associated with disregarding the message. If a user pays attention to and comprehends a warning, their *attitudes and beliefs* can be changed, which is essential to affect their *motivation* and *behavior*. To apply the C-HIP model to digital security messages, we supplement it with another theory that explains attitudes, beliefs, motivations, and behaviors particular to a security context: Protection Motivation Theory (PMT).

Several IT security studies have used PMT, yet key constructs, such as fear, have often been omitted (Boss et al. 2015). Further, while PMT includes attitudes, beliefs, motivation, and behavior, it does not examine the earlier stages of the C-HIP model such as perceptions of the message source, aspects of the message channel, and the attention and comprehension of users. Thus, while C-HIP is the most appropriate model to understand warning messages, we extend and supplement C-HIP with key elements of PMT to better contextualize C-HIP to the information security context.

COMPREHENSIVE THEORETICAL MODEL

To gain a more holistic understanding of the effect of digital warning messages, we propose an expanded C-HIP model that incorporates the concept of fear appeals from PMT. Attitudes and beliefs in our model include, but are not limited to, the attitudes and beliefs traditionally theorized in PMT (i.e., perceived threat severity, perceived threat susceptibility, perceived response cost, perceived self-efficacy, and response efficacy). See Figure 1. In the following sections, we present the associated propositions, summarize which have been well studied in the literature, and comment on areas where future research is needed, all based on a comprehensive literature review of current digital IT security warning message literature.

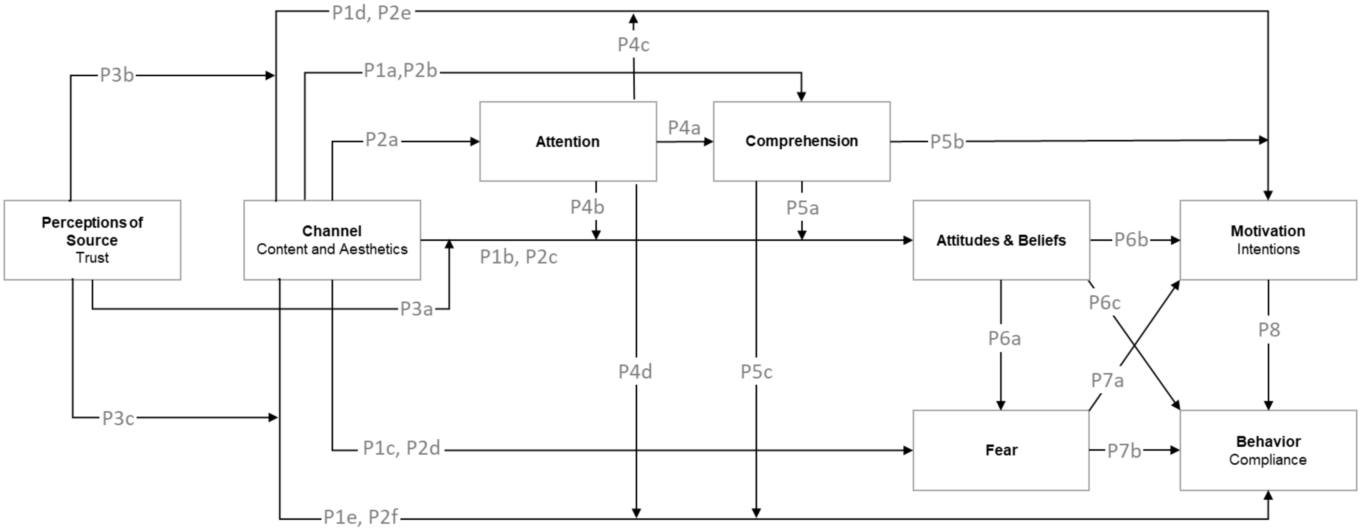

Figure 1. Comprehensive Digital IT Security Warning Message Model (expanded from C-HIP)

COMPREHENSIVE LITERATURE REVIEW

We completed a comprehensive literature review of research on IT security warning messages including all works published through January 2022. Using Google Scholar, we searched for the following key words: "digital warning," "security warning," "online warning," "pop-up warning," "warning banner," "security dialog." From this result set, we searched for works meeting the following criteria: (1) Works should mention security warnings in the abstract, indicating a primary focus on this phenomenon. (2) Works should focus only on digital messages in an IT security context; therefore, works that focused on pop-up messages in other contexts were excluded. (3) Works should theorize or measure human behavior or human perceptions; therefore, commentaries, literature reviews, and design and technical papers were excluded. (4) Works should be published in peer-reviewed journals, conference proceedings, or published books; therefore, posters, forums, essays, editorials, theses, reviews, summaries, short abstracts, and any works 3 pages or less were excluded. (5) Works should be complete; therefore, anything self-classified as work in progress was excluded. Our final set included 76 papers. Full references for these 76 are not included here due to conference paper space constraints but are available upon request.

For each of the 76 papers, we classified them as whether or not they empirically examined each of our propositions. Table 1 summarizes the counts for each proposition. As shown in the table, the most commonly studied relationships in the current literature are (1) the effect of warning message content directly on whether the user behaves in a manner that adheres to or ignores the warning (P1e; 33 studies; 43.4%); (2) the effect of warning message aesthetics directly on user behavior (P2f; 24 studies; 31.6%); (3) how the aesthetics of warnings affect whether individuals pay attention to the warning message (P2a; 19 studies; 25%); and (4) the effect of paying attention on a warning message's effectiveness (P4d; 16 studies; 21.1%).

Proposition	**Studies**
P1a. Variations in the content of a warning message will have a direct effect on comprehension.	5
P1b. Variations in the content of a warning message will have a direct effect on attitudes and beliefs.	2
P1c. Variations in the content of a warning message will have a direct effect on fear.	2
P1d. Variations in the content of a warning message will have a direct effect on motivation.	6
P1e. Variations in the content of a warning message will have a direct effect on behavior.	33
P2a. Variations in the aesthetics of a warning message will be related to the attention paid by an individual.	19
P2b. Variations in the aesthetics of a warning message will be related to comprehension.	5
P2c. Variations in the aesthetics of a warning message will be related to attitudes and beliefs.	2
P2d. Variations in the aesthetics of a warning message will be related to fear.	2
P2e. Variations in the aesthetics of a warning message will be related to motivation.	3
P2f. Variations in the aesthetics of a warning message will be related to behavior.	24
P3a. The effect of warning message channel on user attitudes and beliefs will be stronger when users trust the source of the warning message.	1
P3b. The effect of warning message channel on motivation will be stronger when users trust the source of the warning message.	0
P3c. The effect of warning message channel on behavior will be stronger when users trust the source of the warning message.	4
P4a. Individuals who pay attention to warning messages are more likely to comprehend them.	4
P4b. The effect of warning message channel on attitudes and beliefs is stronger when users pay attention.	1
P4c. The effect of warning message channel on motivation will be stronger when users pay attention.	0
P4d. The effect of warning message channel on behavior will be stronger when users pay attention.	16
P5a. The effect of warning message channel on user attitudes and beliefs will be stronger when users fully comprehend the meaning of the warning.	1
P5b. The effect of warning message channel on motivation will be stronger when users fully comprehend the meaning of the warning.	4
P5c. The effect of warning message channel on behavior will be stronger when users fully comprehend the meaning of the warning.	8
P6a. Attitudes and beliefs, after experiencing a warning message, are associated with fear.	4
P6b. Attitudes and beliefs, after experiencing a warning message, are associated with motivation.	6
P6c. Attitudes and beliefs, after experiencing a warning message, are associated with compliance behavior.	7
P7a. Individual fear, as a result of experiencing a warning message, is associated with motivation.	2
P7b. Individual fear, as a result of experiencing a warning message, is associated with compliance behavior.	2
P8. Individual motivation, as measured by a person's behavioral intentions, is associated with compliance to warning messages.	5

Table 1. Propositions and counts of published works. Propositions with 3 or fewer studies are highlighted.

PROPOSITIONS NEEDING FUTURE RESEARCH

For this conference paper, we focused our limited space to writing on propositions with three or fewer articles. Propositions that have been studied more extensively will be reviewed in an expanded journal manuscript. Here, we describe each of the propositions where more research is needed.

First, only two studies examine the direct effect of warning message content on attitude and beliefs. Bravo-Lillo et al. (2011b) describe a mental model where users carefully think through how the content of a warning message affects a variety of beliefs. Haddad et al. (2020) found that warning messages that explicitly explain reasons and consequences increased perceptions of believability and perceived severity. Although many studies suggest a link between warning message content and actual behavior, we are not aware of any other studies that directly examine the connection between content and attitudes and beliefs. According to the C-HIP model, such a link should exist, but it is not yet fully understood. Therefore, we propose:

P1b. Variations in the content of a warning message will have a direct effect on attitudes and beliefs.

Few studies have measured fear, which is alarming given that the content of a message likely affects a user's fear. Mat Razali et al. (2021) find that specific (non-generic) messages are associated with higher fear. In an fMRI study, Anderson et al. (2014a) found that warning messages increased brain activity associated with fear and anxiety. However, only these two studies examine how warning message content influence the fear a person feels. Accordingly, we propose:

P1c. Variations in the content of a warning message will have a direct effect on fear.

The aesthetics of the warning message are another aspect of channel that should have strong effects. Message aesthetics are non-content-based characteristics of a message that appeal to the senses of the users. Aesthetics are primarily visual, but can also include audio aspects (Datta et al. 2021). Bravo-Lillo et al. (2011b) describe a mental model where users consider the look and feel of warning messages and carefully evaluate their beliefs based on the warning message. Although various studies suggest a link between a warning message aesthetics and behavior, we are not aware of any other studies that directly examine between the aesthetics and the measured attitudes and beliefs. We then posit:

P2c. Variations in the aesthetics of a warning message will be positively related to attitudes and beliefs.

We are aware of only two studies that specifically examine and measure fear as it relates to aesthetic design choices. Mat Razali et al. (2021) found that different levels of fear are associated with aspects such as box color, text color, and icons. Eargle et al. (2016) designed warning messages that integrated pictures of facial expressions and found that they activated parts of the brain related to fear and anxiety. Consequently, we postulate:

P2d. Variations in the aesthetics of a warning message will be positively related to fear.

In terms of how aesthetics affect motivation, Haddad et al. (2020) investigated the effects of using pictures of eyes on warning messages and found that such images actually decreased the motivation to comply with such messages. Bravo-Lillo et al. (2011a) found that warning messages that adhere to a set of best practices for both aesthetics and content led to user motivation because the users found the issues being warned about to be important. Raja et al. (2011) demonstrated that their warning message design led to an increase of user intentions to comply with the warning. Thus, we propose:

P2e. Variations in the aesthetics of a warning message will be positively related to motivation.

Perceptions of the source, the person or entity delivering a message, have been shown to affect adherence to messages in that users discount material from less trustworthy sources (Hovland and Weiss 1951). Kaiser et al. (2021) sought to determine the impact of source when presented with disinformation warning messages. Their study indicated stronger attitudes toward heeding browser warnings when the warning originated from familiar sources (i.e., Google). Hence, we offer the following:

P3a. The effect of warning message channel on user attitudes and beliefs will be stronger when users trust the source of the warning message.

Interestingly, no research to our knowledge has examined the impact of security warning message source on motivation. Outside the context of security warnings, research is full of studies indicating that source has an impact on motivation. For example, one study reported in *Fortune* (Reingold 2016) shows that source trust leads to more employee motivation than pay or other benefits. Therefore, we propose:

P3b. The effect of warning message channel on motivation will be stronger when users trust the source of the warning message.

The main objective for many designers of warning messages is to capture users' attention and convey information about the possible hazard (Bravo-Lillo et al. 2013). If a user's attention is switched to a warning message, users may be more likely understand its meaning (Sunshine et al. 2009) which may lead to increased compliance and better decision making. Jaeger and Eckhardt (2021) conducted a study using eye-tracker software to determine the impact that attention has on situational information security awareness through measuring factors such as phishing experience and the presence of warnings messages. Given the lack of studies evaluating the effect of channel on attitudes and beliefs, no studies to our knowledge show attention moderating these relationships. However, we anticipate that attention may strengthen the relationship between warning message channel and attitudes and beliefs and suggest the following:

P4b. The effect of warning message channel on attitudes and beliefs is stronger when users pay careful attention to the warning.

Research has investigated the impact that attention has on motivation, but this research is outside the context of security warnings. For example, one study (Suri and Gross 2015) indicated that a stimulus must receive attention to be valued and motivated behavior to occur. The study concluded that re-orienting attention could lead to increased motivation. We use this study to inform a similar relationship in the context of security warning messages and propose:

P4c. The effect of warning message channel on motivation will be stronger when users pay careful attention to the warning.

One common mistake of warning designers is to assume that an average user will understand the hazard and its consequences and risks (Wogalter and Laughery 1996). For example, users presented with terms such as “startup disk, encryption, virus, attachment, macro, and certificate” indicated they were familiar with these terms but had difficulties to make sense of them (Bravo-Lillo et al. 2011b). Therefore, the content of warning messages should target the least-skilled users because messages that include complex technical terms are not likely to be comprehended (Wogalter and Laughery 1996). We then posit:

P5a. The effect of warning message channel on user attitudes and beliefs will be stronger when users fully comprehend the meaning of the warning.

Only two IT warning message studies have examined the relationship between fear and measured motivation or intentions. Boss et al. (2015) argued for the inclusion of fear appeals in any study of PMT and examined the effects of fear appeals in the warning message context. They found that fear affected intention to comply with malware warning messages under conditions where the warning message was designed with a fear appeal intent. Jaeger and Eckhardt (2021) found a significant relationship between fear and motivation in a phishing context. Accordingly, we postulate:

P7a. Individual fear, as a result of experiencing a warning message, is positively associated with motivation.

Two recent studies examined the direct effect of fear on warning message behavior. Kaiser et al. (2021) found that fear or risk of harm influenced users to navigate away after seeing a warning message, even when they did not take the time to fully process the written message. Sharma et al. (2021), on the other hand, found no significant relationship between measured fear and the decision of whether to comply with a phishing warning message. Thus, there are conflicting findings in the literature. Further, more research is needed to understand more deeply how fear interacts with attention and comprehension of warning messages. We then propose:

P7b. Individual fear, as a result of experiencing a warning message, is positively associated with compliance behavior.

CONCLUSION

For individuals and organizations to better protect themselves against the adverse consequences of poor choices, a comprehensive understanding is vital to properly safeguard individual and organizational assets. We present the comprehensive digital IT security warning message model that intersects PMT with the C-HIP model.

After undergoing an extensive evaluation of prior security warning message literature, we determined that perceptions of the warning message source, comprehension of the warning message, attitudes and beliefs, and fear have been largely understudied while warning message channel (content/aesthetics), attention, and compliance behavior have been the predominant focus. Diverting attention from these latter topics to focus on the former topics could offer additional light on why individuals continue to not comply with warning messages.

In particular, we conclude that the two biggest opportunities for future research on IT security warning messages are in (1) studying how perceptions of the source affect all aspects of receiving the warning message, and (2) the role that fear plays (or does not play) in processing a warning message. First, regarding perceptions of source, we know very little about how such perceptions change whether the content and aesthetics of a message affect attitudes and beliefs (P3a), motivation (P3b), and ultimate behavior (P3c). Secondly, we call on researchers to further examine the role of fear in processing warning messages. Is fear the most effective mechanism, and if so, how much of a role does it play? More research is needed to understand both what affects fear (P1c, P2d, P6a) and how fear ultimately affects motivation and behavior (P7a, P7b).

REFERENCES

1. Akhawe, D., and Felt, A. P. 2013. "Alice in Warningland: A large-scale field study of browser security warning effectiveness," *Proceedings of the 22nd USENIX Conference on Security*, pp. 257-272.
2. Anderson, B. B., Vance, A., Kirwan, B., Eargle, D., and Howard, S. 2014a. "Users aren't (necessarily) lazy: Using neuroIS to explain habituation to security warnings," *2014 International Conference on Information Systems (ICIS)*.

3. Anderson, B. B., Vance, A., Kirwan, B., Eargle, D., and Howard, S. 2014b. "Why users habituate to security warnings: Insights from fMRI," *2014 IFIP 8.11 Dewald Roode Security Workshop*.
4. Boss, S. R., Galletta, D. F., Lowry, P. B., Moody, G. D., and Polak, P. 2015. "What do systems users have to fear? Using fear appeals to engender threats and fear that motivate security behaviors," *MIS Quarterly* (39:4), pp. 837-864.
5. Bravo-Lillo, C., Cranor, L. F., Downs, J., Komanduri, S., Reeder, R. W., Schechter, S., and Sleeper, M. 2013. "Your attention please: Designing security-decision UIs to make genuine risks harder to ignore," *SOUPS '13 Proceedings of the Ninth Symposium on Usable Privacy and Security*, pp. 1-18.
6. Bravo-Lillo, C., Cranor, L. F., Downs, J., Komanduri, S., and Sleeper, M. 2011a. "Improving computer security dialogs," *IFIP Conference on Human-Computer Interaction*, pp. 18-35.
7. Bravo-Lillo, C., Cranor, L. F., Downs, J. S., and Komanduri, S. 2011b. "Bridging the gap in computer security warnings: A mental model approach," *IEEE Security & Privacy* (9:2), pp. 18-26.
8. Conzola, V. C., and Wogalter, M. S. 2001. "A Communication-Human Information Processing (C-HIP) approach to warning effectiveness in the workplace," *Journal of Risk Research* (4:4), pp. 309-322.
9. Datta, P., Namin, A. S., Jones, K. S., and Hewett, R. 2021. "Warning users about cyber threats through sounds," *SN Applied Sciences* (3:7), pp. 1-21.
10. Eargle, D., Galletta, D., Kirwan, B., Vance, A., and Jenkins, J. 2016. "Integrating Facial Cues of Threat into Security Warnings–An fMRI and Field Study,").
11. Haddad, A., Sauer, J., Prichard, J., Spiranovic, C., and Gelb, K. 2020. "Gaming tasks as a method for studying the impact of warning messages on information behavior," *Library Trends* (68:4), pp. 576-598.
12. Hovland, C. I., and Weiss, W. 1951. "The influence of source credibility on communication effectiveness," *Public opinion quarterly* (15:4), pp. 635-650.
13. Jaeger, L., and Eckhardt, A. 2021. "Eyes wide open: The role of situational information security awareness for security-related behaviour," *Information Systems Journal* (31:3), pp. 429-472.
14. Kaiser, B., Wei, J., Lucherini, E., Lee, K., Matias, J. N., and Mayer, J. 2021. "Adapting Security Warnings to Counter Online Disinformation," *30th {USENIX} Security Symposium ({USENIX} Security 21)*.
15. Mat Razali, N. A., Md Saad, N. J. A., Wook, M., Hasbullah, N. A., Mohd Noor, N., and Ishak, K. K. 2021. "Fear Assessment in Information Security Dialog Box based on Hybrid Kansei Engineering and KJ Method," *International Journal of Affective Engineering*), pp. IJAE-D-20-00021.
16. Raja, F., Hawkey, K., Hsu, S., Wang, K.-L. C., and Beznosov, K. 2011. "A brick wall, a locked door, and a bandit: a physical security metaphor for firewall warnings," *Proceedings of the seventh symposium on usable privacy and security*, pp. 1-20.
17. Reingold, J. 2016. "Why trust motivates employees more than pay" Retrieved from https://fortune.com/2016/04/27/why-trust-motivates-employees-more-than-pay/.
18. Sharma, K., Zhan, X., Nah, F. F.-H., Siau, K., and Cheng, M. X. 2021. "Impact of digital nudging on information security behavior: an experimental study on framing and priming in cybersecurity," *Organizational Cybersecurity Journal: Practice, Process and People*).
19. Sunshine, J., Egelman, S., Almuhimedi, H., Atri, N., and Cranor, L. F. 2009. "Crying wolf: An empirical study of SSL warning effectiveness," *18th USENIX Security Symposium*, pp. 399-432.
20. Suri, G., and Gross, J. J. 2015. "The role of attention in motivated behavior," *Journal of Experimental Psychology: General* (144:4), p. 864.
21. Vance, A., Jenkins, J. L., Anderson, B. B., Bjornn, D. K., and Kirwan, C. B. 2018. "Tuning out security warnings: A longitudinal examination of habituation through fMRI, eye tracking, and field experiments," *MIS Quarterly* (42:2), pp. 355-380.
22. Wogalter, M. S. 2006a. "Communication-Human Information Processing (C-HIP) Model," in: *Handbook of Warnings*, M.S. Wogalter (ed.), Lawrence Erlbaum Associates, Mahwah, NJ, pp. 51-61.
23. Wogalter, M. S. 2006b. "Purposes and scope of warnings," *Handbook of Warnings. Lawrence Erlbaum Associates, Mahwah, NJ*), pp. 3-9.
24. Wogalter, M. S., and Laughery, K. R. 1996. "Warning! Sign and label effectiveness," *Current Directions in Psychological Science*), pp. 33-37.