DEFENDING AGAINST SPEAR PHISHING: MOTIVATING USERS THROUGH FEAR APPEAL MANIPULATIONS

Sebastian Walter Schuetz
City University of Hong Kong, sebastian.w.schuetz@outlook.com

Paul Benjamin Lowry
City University of Hong Kong, paul.lowry.phd@gmail.com

Jason Bennett Thatcher
Clemson University, jthatch@clemson.edu

Follow this and additional works at: http://aisel.aisnet.org/pacis2016

Recommended Citation
Schuetz, Sebastian Walter; Lowry, Paul Benjamin; and Thatcher, Jason Bennett, "DEFENDING AGAINST SPEAR PHISHING: MOTIVATING USERS THROUGH FEAR APPEAL MANIPULATIONS" (2016). PACIS 2016 Proceedings. 74.
http://aisel.aisnet.org/pacis2016/74
DEFENDING AGAINST SPEAR PHISHING: MOTIVATING USERS THROUGH FEAR APPEAL MANIPULATIONS

Sebastian Walter Schuetz, Department of Information Systems, City University of Hong Kong, Hong Kong, sebastian.w.schuetz@outlook.com

Paul Benjamin Lowry, Department of Information Systems, City University of Hong Kong, Hong Kong, paul.lowry.phd@gmail.com

Jason Bennett Thatcher, Social Analytics Institute, Department of Management, Clemson University, Clemson, South Carolina, USA, jthatch@clemson.edu

Abstract

Phishing is a pervasive form of online fraud that causes billions in losses annually. Spear phishing is a highly targeted and successful type of phishing that uses socially engineered emails to defraud most of its recipients. Unfortunately, anti-phishing training campaigns struggle with effectively fighting this threat—partially because users see security as a secondary priority, and partially because users are rarely motivated to undergo lengthy training. An effective training approach thus needs to be non-disruptive and brief as to avoid being onerous, and yet, needs to inspire dramatic behavioral change. This is a tremendous, unsolved challenge that we believe can be solved through a novel application of theory: Using fear appeals and protection-motivation theory (PMT), we outline how brief training can educate users and evoke protection motivation. We further invoke construal-level theory (CLT) to explain how fear appeals can stimulate threat perceptions more quickly and more powerfully. This research-in-progress study further proposes a field experiment to verify the effectiveness of our proposed training approach in an ecologically valid environment. Overall, we (1) improve training based on PMT and CLT, (2) expand PMT for guiding fear appeal design; and (3) demonstrate a full application of CLT.

Keywords: Security, Spear phishing, Training, Fear Appeals, Protection Motivation Theory, Construal Level Theory
1 INTRODUCTION

Phishing is a serious threat that caused more than US$5.9 billion in damages in 2013 (RSA 2014); and that number promises to rise: The recent Cyber Threat Report of cyber intelligence provider CYREN reports a steep 51% increase alone in the first three months of 2015. Indeed, the amount of phishing attacks has been rapidly growing, exceeding 7.5 trillion attacks in 2014, which is double the number from 2012 (McAfee Labs 2014). In every quarter of 2014, more than 197,252 new, unique attacks were executed (APWG 2014), and this number rises continuously.

Across these rapidly growing and numerous attacks, one technique—spear phishing—stands out. Spear phishing is a highly targeted, context-specific attack that is directed at specific groups of individuals or organizations and aims to appear genuine to their members and employees (Wang et al. 2012). Compared to phishing, spear phishing is more sophisticated (Hong 2012), and inherently more dangerous—by magnitudes (Jagatic et al. 2007); more effective (its return rates are up to 40 times greater than those of ordinary phishing); and increasingly more frequent (FireEye 2012). Some particularly popular spear-phishing attacks impersonate banks, which often lead to worrisome rates of success (APWG 2014).

This persistent increase in the number of successful attacks suggests spear-phishing is far from being under control. Existing approaches aim to prevent victimization with the use of enhanced automated detection mechanisms and behavioral training (Hong 2012). However, automated solutions struggle to reliably detect customized and fast-moving spear-phishing attacks—rendering users vulnerable when they solely rely upon automated solutions (Abbasi et al. 2012). Moreover, even when these tools correctly raise the alarm, users often disregard such warning messages and continue with their imprudent behavior, leaving themselves vulnerable to spear-phishing attacks (Abbasi et al. 2015). In light of the deficiencies of automated solutions and users’ deviant behavior in using them, we turn to educational approaches to further learn how to increase organizational defenses against spear-phishing victimization.

Although evidence suggests that anti-phishing training generally improves detection rates (Kumaraguru et al. 2010), the results do not evoke confidence in such training’s efficacy. Particularly for spear phishing, more than 70 percent of participants consistently fell victim to these highly tailored and targeted attacks—sometimes even despite previous training (see Dodge et al. 2007; Ferguson 2005; Jagatic et al. 2007). Unfortunately, the existing approaches to behavioral training—which often focus solely on educating users about phishing—are deficient. Some researchers argue that one of the key challenges in delivering security education is motivating users to engage in protective behavior (Kumaraguru et al. 2010; Puhakainen and Siponen 2010). If so, then effective anti-phishing training needs not only to inform users, but also needs to arouse a sense of fear that will elicit protection motivations, resulting in meaningful behavioral change (i.e., not fall victim to the phishing attack). Consequently, the current conundrum in research and practice is how anti-phishing training be conducted effectively without being onerous and disruptive of end-users’ time. This enigma forms the guiding research question of our research.

To identify the attributes of effective anti-phishing training that hold potential to inform users and elicit protective motivations, we turn to protection-motivation theory (PMT). PMT explains the effect of fear appeals by identifying the cognitive processes that engender protective intentions. Fear appeals are “persuasive messages designed to scare people by describing the terrible things that will happen to them if they do not do what the message recommends” (Witte 1992, p. 329). Because fear appeals are brief, fear-inducing communications, they hold potential to overcome the tedious nature of traditional training approaches to effectively raise protection motivation and subsequent behaviors. Recent IS research has adapted PMT from its traditional healthcare context and has argued it is well-suited for privacy (Milne et al. 2009) and security contexts (Boss et al. 2015). Likewise, IS research on security education, training and awareness (SETA) programs has recently explored the tenets of PMT to foster protective intentions (Posey et al. 2015). Whereas PMT has been widely applied across various areas, a theoretical account guiding the
design of fear appeals for effective, brief training of end users is still missing. To address this gap, we further draw from construal-level theory (CLT) to understand the mental representations that underlie the threat and coping response appraisal processes. These mental representations (i.e., construals) form the interpretive basis on which PMT’s cognitive processes engender protection motivation. Our purpose in invoking CLT is thus to explore whether and how the manipulation of these construals can aid evoking protection motivation, and thereby inform the design of anti-phishing training materials.

Within the realm of phishing and PMT, our study offers three contributions. First, we address the necessity for brief and effective anti-phishing training by building upon PMT to understand how to deliver fear appeals that both inform and motivate. Second, we extend PMT by examining the design of fear appeals, using CLT as our theoretical lens to explain how the design of a fear appeal guides the mental representation that informs protective intention. This allows us to offer a unique theoretical account for how to design effective fear appeals that can be used in spear-phishing training. Third, we demonstrate a full utilization of CLT in an applied area, whereas applications of CLT are still rare outside of psychology, and whereas existing publications in IS examine only parts of CLT. These contributions are unique not only to the field of IS but also to all other disciplines that leverage fear appeals, PMT, and CLT.

2 LITERATURE REVIEW

2.1 Phishing and Spear Phishing

Phishing is a method of social engineering that aims to induce recipients to expose personal information, such as contact details, user/network credentials, or credit card details. Typically, phishing targets a mass audience (Wright and Marett 2010) and succeeds in deceiving only a small percentage of the audience (Wright et al. 2014). For phishing to be effective, successful messages must be both believable and persuasive (Hong 2012). Believability is often achieved by mimicking communications from reputable persons or organizations (Wright and Marett 2010). A variety of techniques exist to boost persuasiveness (cf. Wright et al. 2014).

However, spear phishing is dramatically more successful than traditional phishing, as it targets a specific group of recipients and leverages information that is specific to that group so as to increase its believability (Hong 2012). Spear-phishing attacks often appear to be genuine (Wang et al. 2012). Hence, detecting spear phishing requires that users pay close attention not only to obvious visual cues, but especially to the plausibility of the message. For example, Jagatic et al. (2007) mined social media to impersonate real friends of targets, and achieved a deception success rate over 70 percent. Similarly, Dodge et al. (2007) phished students exploiting their sensitivity to grade-related matters, and succeeded with eight out of ten students. Although the effort of tailoring phishing attacks to a specific audience makes this type of attack more costly, such attacks are far more successful and dangerous (FireEye 2012). Users often underestimate the ease of spoofing emails, and thus overrate the security and privacy of mail communications (Jagatic et al. 2007).

2.2 Defense against Phishing

Approaches to counter phishing attacks can be categorized into two broad classes: technical and educational. Technical approaches, such as anti-phishing tools, attempt to automatically identify fraudulent messages and either discard them immediately or warn the users of their potentially harmful nature. Research on technical solutions has led to improvements in detection accuracy (e.g., Abbasi et al. 2015), which have been found to be—along with speed and cost of error—crucial factors driving user’s reliance on such tools (Zahedi et al. 2015). However, technical approaches suffer from severe weaknesses. First, the fast-moving nature of phishing attacks stall detection rates at about 90 percent accuracy (Hong 2012); however, this is a woefully inadequate rate given the sheer volume of phishing attacks. Furthermore, automated warnings are
often overlooked or purposely ignored (Egelman et al. 2008; Wu et al. 2006). Consequently, technical means are incapable of offering sufficient protection against phishing attacks (Abbasi et al. 2012). In particular, spear phishing has proven apt at defeating existing security mechanisms (FireEye 2012). Thus, educational anti-phishing approaches are necessary to complement their technical counterparts, especially training that enables users to draw the appropriate conclusions from automated warnings (Lai et al. 2012).

Educational approaches can significantly reduce the likelihood of falling for phishing attacks (Kumaraguru et al. 2010). Training enhances user’s self-efficacy, security knowledge, and suspicion—all factors that have been found significant in reducing deception success (Wright and Marett 2010). However, making anti-phishing training salient among users has proven to be a difficult task (e.g., Ferguson 2005 who successfully phished 90 percent users who received a four hour training session on the very same day), because many users regard security as a secondary activity that is bothering at best and interfering at worst (Puhakainen and Siponen 2010). As a training technique, researchers have used feigned phishing attacks to probe users as a means to increase user’s motivation to protect themselves (Kumaraguru et al. 2010). However, feigned phishing attacks also often cause negative pushbacks among users, evident in their expression of anger and denial (Jagatic et al. 2007), and are thus—despite being effective in arousing awareness—ethically controversial (Jakobsson et al. 2008). Ideally, an educational approach could overcome motivational issues without evoking negative emotions in users. The motivation to actively engage in protective behaviors becomes especially salient for spear phishing, as this type of attack is well disguised and hard to detect based solely upon visual cues. Ferguson (2005) underlines the important role of motivation by reporting a startling 90 percent success rate when phishing freshmen shortly after preparing them with a four-hour long training session. Although the freshmen reported to be suspicious, the subject matter of the phishing attack (in this study: “a problem with your last grade”) easily overruled the arguably insufficient motivation to be cautious and engage in protective action. These results confirm the essential role of motivation as a salient trigger to an effortful protective process that aims to distinguish phishing messages from innocuous messages.

2.3 Protection-Motivation Theory

PMT is a theoretical model that explains and predicts the processes that elicit protection motivation in recipients of fear appeal communications. PMT’s core assumption is that fear appeals sequentially trigger cognitive appraisal processes, which jointly inform protection intentions. A successful fear appeal thereby raises protection motivation, which captures one’s intent to engage in protective behaviors in an effort to avoid the depicted dangers (Floyd et al. 2000). According to Rogers (1975), fear appeals must contain three essential stimulus variables: “(a) the magnitude of noxiousness of a depicted event; (b) the conditional probability that the event will occur provided that no adaptive behavior is performed […] and (c) the availability and effectiveness of a coping response that might reduce or eliminate the noxious stimulus” (p. 97). A fear appeals thus stimulates protection motivation by firstly communicating a threat, and then subsequently outlining an adequate coping response.

The initial threat appraisal process balances perceived threat severity, referring to one’s belief that a threat is serious and potentially serve (Rohm and Milne 2004), and perceived threat susceptibility, as one’s expectation of being susceptible to that particular threat (Rohm and Milne 2004), against the maladaptive rewards one gains from not engaging in protective action that would alleviate that threat (Floyd et al. 2000). When faced with a comprehensible and personally relevant threat, recipients further experience fear (Leventhal 1970). Fear has been conceptualized as a core mediator of protection motivation (Boss et al. 2015). In contrast, the succeeding coping appraisal process evaluates the recommended coping strategy by comparing response efficacy, which refers to one’s belief in the efficacy of the protective action (Floyd et al. 2000), in conjunction with self-efficacy, relating to one’s perceived ability to actually perform the protective action (Floyd et al. 2000; Maddux and Rogers 1983), versus the response cost, which refers to
all costs that arise from carrying out the protective response (Floyd et al. 2000). Accordingly, PMT proposes that if the noxiousness of the threat outweighs the associated maladaptive rewards—and if the recommended coping response is perceived to be an executable, efficient and affordable resort—only then will a recipient be motivated to engage in the recommended protective response (Boss et al. 2015; Floyd et al. 2000; Maddux and Rogers 1983).

For the design of effective fear appeals, PMT requires threat and coping appraisal perceptions (Milne et al. 2000). Furthermore, Boss et al. (2015) show that threats must be personally relevant in order to engender fear and engender protective motivation. Beyond this basic guidance on how to design effective fear appeal manipulations, it is surprising how few studies have investigated the actual effects of specific manipulations. Instead of focusing on which information needs to be conveyed to effectively stimulate appraisal processes, existing studies have restrained to confirm that comics and pictures are superior to textual fear appeals (Dijkstra and Bos 2015; Kumaraguru et al. 2010); pleasure-evoking colors (blue) are more effective than less pleasure-evoking colors (yellow) (Wauters et al. 2008); or that positive consequences are more persuasive than negative consequences (Anderson and Agarwal 2010), which already deviates from the core purpose of a fear appeal: Thus, across the literature, general account or explanation of how to conduct effective fear-appeal design is missing. However, designing effective fear-appeals is elementary to the study of fear appraisal, and hence presents a unique opportunity to inform and contribute to PMT literature.

2.4 Construal-Level Theory

Construal-level theory holds potential to explain how fear appeals are represented mentally, and can thus inform our understanding of how fear-appeal design shapes cognitive appraisal. CLT proposes that people, to generate predictions about the future, form “abstract mental construals of distal objects” (Trope & Liberman, 2010, p. 440, emphasize added). The construal thereby serves a goal, and manifests itself as either a concrete or abstract mental representation, the latter thereby being more schematic, prototypical and hence simpler and less ambiguous (Fiske and Taylor 1991; Smith 1998). CLT coins concrete construals as being on the low-level end of the abstraction continuum, whereas abstract construals are on the high-level end of the continuum.

The notion of CLT is strictly hierarchical: whereas abstract construals represent the conceptual idea of an object, concrete construals represent a more specific instantiation. Concrete and abstract construals are thus concrete and abstract mental representations of the same object, respectively, and are connected through hierarchical meaning: the high-level (abstract) construal determines the meaning of the low-level (concrete) construal more than the low-level construal the meaning of the high-level construal (Trope and Liberman 2010). As illustration, an abstract spear-phishing construal would be perceived to be dangerous, whereas the same concrete construal would describe the dangers more specifically, such as “steals your credit card information.” When the high-level attribute “dangerous” changes to “safe,” the hierarchically connected low-level attribute “steals credit card information” vanishes, because it is inconsistent with the superordinate attribute of being “safe.” The superordinate attribute is hence more central to the meaning of the construal.

2.4.1 Psychological distance

Because construals serve as the basis of judgement and prediction, CLT proposes that construals mediate the consequences of psychological distance, which refers to the perception of something being either close or far away from the self (Trope and Liberman 2010). According to CLT, psychological distance is directly associated to construal level (Bar-Anan et al. 2006), and therefore describes the perceived distance between the present self and the construal. Feelings of distance can be evoked by moving the construal along a temporal (now vs. in the far future), spatial (here vs. far away), social (me vs. others) or hypothetical (realistic vs. hypothetical) dimension. Accordingly, one perceives a construal to be psychologically
proximate if it is construed to be present in the here and now, involving oneself, realistically. By contrast, a
construal is perceived to be distant if it is located in the future, somewhere else, and involves others, and
takes place hypothetically.

2.4.2 Purpose and Features

CLT proposes that with increasing psychological distance, the gestalt of the construal changes. According
to Liberman and Förster (2009) and Steidle et al. (2011), there is a bilateral association between
psychological distance and construal level, such that a greater psychological distance is accompanied by a
higher construal level (i.e., thinking that is more abstract), whereas a more proximate distance is
accompanied by a lower construal level (i.e., thinking that is more concrete) (Bar-Anan et al. 2006). Trope
and Liberman (2010) propose that this is due to a shift in purpose: Because proximate (i.e., imminent) threats
demand an immediate response, an understanding of “how am I threatened” is needed. To answer such
demand, concrete construals are formed that entail viable details of the specific threat, and thus serve the
purpose of explaining its feasibility—how it exactly works. By contrast, distant threats do not require such
immediate response, but demand a judgement of “whether I will be threatened.” As a result, abstract
construals are formed that focus on the general desirability of the threat (in abstract terms, not one particular
instance), and thus serve the purpose to evaluate its desirability—does it need to be avoided?

Following the purpose, CLT proposes that abstract construals exhibit central features that relate to
desirability and omit peripheral features that relate to feasibility (Trope and Liberman 2010). For example,
if the purpose is to understand how to detect a spear-phishing attack, a low-level construal might feature
specific attributes of the message, such as whether it will carry a personal greeting, impersonate a specific
institution, or requests credit card details, as such information is crucial for detecting the attack. However,
when moving to a higher construal, for which the purpose is to understand the consequences of falling victim
to a spear-phishing attack, these features become irrelevant. Instead, features relating to the concept of spear
phishing are salient characteristics of such an abstract construal.

The application of CLT in research is still rather limited. In IS, two studies published in premier outlets
apply CLT to recommendation systems (cf. Köhler et al. 2011) and systems adoption (cf. Ho et al. 2014).
Despite the pioneering role of these CLT-based studies, these studies did not yet have the opportunity to
fully leverage the full notion of CLT, nor explore construal features and psychological distance
manipulations to stimulate construal formation. However, CLT yields a unique opportunity to understand
the interpretative basis of cognitive processes, and we explore its notion to guide the design of more
persuasive fear appeals by modelling the attributes of low- and high-level construals.

3 RESEARCH MODEL

Based on CLT’s understanding of how concrete and abstract construals emphasize different purposes and
features, and ultimately form the basis for prediction and judgement, we combine CLT and PMT’s appraisal
processes to form our research model (Figure 1). Particularly, we argue that threat and coping appraisal
process each draws upon a construal to enable its cognitive judgement. Thus, we model that a spear-phishing threat construal and an anti-phishing coping response construal informs each respective appraisal process.

**Figure 1. Research model**

### 3.1 Psychological Distance and Self-Perceptions

The first difference between concrete and abstract construals is their relative position to the self (psychological distance). Thus, PMT constructs that evaluate self-perceptions are likely stimulated by traversing construal’s relative position to the self (psychological distance). Such constructs are threat vulnerability, maladaptive rewards, self-efficacy, and response cost. The remaining constructs of PMT’s appraisal processes are independent of the self and relate to the construal’s gestalt; and thus remain unaffected by changes in psychological distance.

Intuitively, the construal of a spear-phishing attack that takes place proximately provokes a more immediate perception of risk than does an attack that takes place distally. Similarly, maladaptive rewards that are construed to be attainable proximately receive higher appraisal than rewards that are attainable distally. Indeed, research on risk taking suggests this temporal connection between reward appraisal (Smithson 2008). As a farther psychological distance is associated with a higher-level construal, we argue:

- **H1a**: A high-level spear-phishing threat construal leads to lower threat vulnerability perceptions.
- **H1b**: A high-level spear-phishing threat construal leads to lower maladaptive rewards perceptions.

Similarly, when evaluating one’s efficacy in executing the coping response, psychological distance dictates the perception of whether the coping response is in need to be executed proximately, that is soon, realistically, by oneself; or distally, that is in the far future, by someone else, hypothetically. Accordingly, a more abstract coping response is likely to elicit lower levels of self-efficacy. Furthermore, because concrete construals convey more specific procedural details of what needs to be done, this argument is further supported by Arachchilage and Love (2014) who find that users require procedural (concrete) and
conceptual (abstract) knowledge to build up self-efficacy perceptions, while in contrast, conceptual knowledge alone is insufficient. Thus, we argue:

\[ H2a: \text{A high-level response construal leads to lower self-efficacy perceptions.} \]

Finally, likewise to maladaptive rewards, response costs are similarly discounted if they are distal. Hence, costs that are construed to incur proximately receive higher appraisal than costs that are construed to occur distally.

\[ H2b: \text{A high-level response threat construal leads to lower response cost perceptions.} \]

3.2 Features of Threat and Coping Response

Differences in psychological distance command differences in construal level: Abstract construals carry high-level features that relate to desirability, whereas concrete construals exhibit low-level features that relate to feasibility. These differences affect two PMT variables that are concerned with evaluating the nature of threat and coping response: For the first, threat severity, CLT argues that a concrete construal of a spear-phishing attack with low-level features (e.g., sender address, firm logo, link to website) enables an understanding of what constitutes an attack; whereas an abstract construal with high-level features (e.g., deceptive, dangerous) enables an understanding of the impact of an attack—which we predict to evoke higher threat perceptions. Indeed, previous research has shown that the abstract handling of information systems (i.e., the way managers do—in contrast to technical staff) increases security risk perceptions (Goodhue and Straub 1991). Thus, we propose that an abstract spear-phishing attack construal’s focus on the (un)desirability of victimization increases the threat severity perceptions.

\[ H1c: \text{A high-level spear-phishing threat construal leads to higher threat severity perceptions.} \]

Similarly, when evaluating the efficacy of the coping response, the question becomes whether it works. Addressing this question requires conceptual knowledge—knowledge about the outcome of the coping response, rather than knowledge about its workings (McCormick 1997). Again, because abstract construals emphasize features that relate to desirability, we predict that such construals increase response efficacy perceptions—rather than concrete construals that focus on feasibility; on how it works.

\[ H2c: \text{A high-level response construal leads to higher response efficacy perceptions.} \]

4 RESEARCH PLAN

We plan to conduct a field experiment in which we will train a large body of individuals about the ins and outs of spear phishing through fear appeal messages; and then will probe the effectiveness of the trainings through a feigned spear-phishing attack targeting the participants, for which we follow the guidelines for ethical fraud experiments outlined by Jakobsson et al. (2008). Because we objectively observe subject’s behavior (i.e., whether they fall for the feigned spear-phishing attack) within the context of their day-to-day lives, we ecologically validate the applicability of our theory, and avoid the problematic discrepancy between self-reported intentions and behaviors which is often found in security contexts (Workman 2008). Our research design aims at treating subjects with concrete or abstract construals of the two core fear appeal elements: threat and response. For the other PMT-related measures, we mainly rely upon validated instruments of Boss et al. (2015). We expect to conduct the experiment later this year.

References

FireEye. (2012). Spear phishing attacks — Why they are successful and how to stop.


RSA. (2014). 2013 A year in review. RSA Online Fraud Report.,


promote individuals’ use and enhance their performance. Journal of the Association for Information Systems, 16 (6), 448–484.