Journal of the
Association for Information Systems

Research Paper ISSN: 1536-9323

# Overconfidence in Phishing Email Detection

**Jingguo Wang**
Information Systems and Operations Management
College of Business, University of Texas at Arlington
jwang@uta.edu

**Yuan Li**
Business, Mathematics and Sciences
Columbia College
yli@columbiasc.edu

**H. Raghav Rao**
Information Systems and Cyber Security
College of Business, University of Texas at San Antonio
hr.rao@utsa.edu

## Abstract:

This study examines overconfidence in phishing email detection. Researchers believe that overconfidence (i.e., where one's judgmental confidence exceeds one's actual performance in decision making) can lead to one's adopting risky behavior in uncertain situations. This study focuses on what leads to overconfidence in phishing detection. We performed a survey experiment with 600 subjects to collect empirical data for the study. In the experiment, each subject judged a set of randomly selected phishing emails and authentic business emails. Specifically, we examined two metrics of overconfidence (i.e., overprecision and overestimation). Results show that cognitive effort decreased overconfidence, while variability in attention allocation, dispositional optimism, and familiarity with the business entities in the emails all increased overconfidence in phishing email detection. The effect of perceived self-efficacy of detecting phishing emails on overconfidence was marginal. In addition, all confidence beliefs poorly predicted detection accuracy and poorly explained its variance, which highlights the issue of relying on them to guide one's behavior in detecting phishing. We discuss mechanisms to reduce overconfidence.

**Keywords:** Phishing Email Detection, Overconfidence, Judgmental Bias, Judgmental Confidence, Judgmental Accuracy, Phishing Detection Self-efficacy, Cognitive Strategies, Motivational Factors.

Mikko Siponen was the accepting senior editor. This article was submitted on November 8, 2013 and went through three revisions.

Volume 17 Issue 11 pp. 759 – 783 November 2016

# 1 Introduction

Phishing causes significant financial loss and erodes trust in online business communications (Symantec, 2014). Understanding why people fall for phishing attacks (e.g., replying to a fraudulent email with personal information such as bank accounts, credit card numbers, and even social security numbers) despite the ubiquity of technology solutions has drawn much research attention (Kumaraguru, Sheng, Acquisti, Cranor, & Hong, 2010; Sheng, Magnien, Kumaraguru, Acquisti, & Cranor, 2007). To develop effective countermeasures and intervening programs to fight phishing, researchers have taken a behavioral approach to understand users' motivations, beliefs, and mental models in phishing detection (Downs, Holbrook, & Cranor, 2006; Hong 2012). As individuals' initial judgment of an email largely determines their subsequent behavioral reaction to the email's requests, this study brings an important aspect of decision making, confidence, in exploring individuals' judgmental processes in detecting phishing emails.

Research in behavioral decision making suggests that confidence plays a critical role in predicting human behavior and that individuals who are more confident with their judgment behave more consistently with it (Fazio & Zanna, 1978). However, if a person's confidence is misplaced[1], where confidence exceeds performance, it may lead the person to adopting risky or erroneous behavior (Moore & Healy, 2008). In the context of phishing detection, for instance, a person who judges a phishing email as a genuine business email with high confidence (i.e., the person is overconfident) may respond to the email and release personal information and, thus, cause identity theft; if the person makes the same wrong judgment with low confidence (i.e., the person is less overconfident), the person may avoid such consequence by taking further action to verify the email's authenticity. Similarly, a person who mistakes an authentic email (such as a bank notice) for a phishing email with high confidence would ignore the email and, thus, take no action; if the person makes the same false judgment but with low confidence, the person may spend more time verifying the email's authenticity. As Tang et al (2014) note, "overconfidence may cause people to take an action that they should not take, possibly leading to suboptimal outcomes or even disasters (p. 26)". This illustration of judgmental confidence and the issue of overconfidence in phishing detection implies that a better correspondence between confidence and accuracy would help prevent one from falling to phishing.

To the best of our knowledge, little research has focused on understanding overconfidence in phishing detection (and in security threat assessment as well). Although prior studies (Hong, Kelley, Tembe, Murphy-Hill, & Mayhorn, 2013; Dhamija, Tygar, & Hearst, 2006; Downs et al., 2006; Kumaraguru, Rhee, Acquisti, Cranor, & Hong, 2007) hint at the possible existence of overconfidence in individuals' phishing detection, the nature of confidence and subsequently overconfidence deserve further investigation. For instance, Hong et al. (2013) studied confidence as a person's self-efficacy beliefs (Stone, 2000): they first measured subjects' prior experience with phishing attacks and found, "approximately 92% of participants misclassified phishing emails even though 89% indicated they were confident of their ability to identify phishing emails (p. 4)". Interestingly, Hayes, Tanner, and Schmidt (2012, p. 109) applied a similar technique to measure confidence (i.e., "confidence in their knowledge of viruses, Trojans, spyware, and phishing attacks") in a study on small business professionals' knowledge of computer threats but found no evidence of overconfidence. Both studies compared prospective confidence (Busey, Tunnicliff, Loftus, & Loftus, 2000) prior to the judgment with task performance in threat detection but not retrospective confidence (or one's judgmental confidence; Tang, Hess, Valacich, & Sweeney, 2014) after a judgment was made, similar to most other behavioral decision making studies (Koriat, Lichtenstein, & Fischhoff, 1980; Lichtenstein & Fischhoff, 1977; Lichtenstein, Fischhoff, & Phillips, 1982). These conflicting but limited results mean we need to better understand confidence and overconfidence in phishing detection.

In this study, we distinguish retrospective confidence (i.e., judgmental confidence) from prospective confidence (e.g., self-efficacy beliefs; Busey et al., 2000) and examine factors that cause one's retrospective overconfidence in phishing detection. We present a starting point to devise intervention programs to address this type of bias and to improve effectiveness in phishing detection. For simplicity, we use the term overconfidence for retrospective overconfidence. Drawing from psychological literature in judgment under uncertainty (Heath & Tversky, 1991; Johnson & Payne, 1985; Keren, 1997; Koriat et al., 1980; Libby & Rennekamp, 2011; Payne, 1982), we analyze cognitive and motivational factors that influence overconfidence, including cognitive effort, variability in attention allocation, familiarity with the business entities in the emails, perceived self-efficacy in detecting phishing emails, and dispositional optimism. We carried out a national survey experiment with 600 subjects to test the research model. The

[1] People are either overconfident or underconfident depending on whether their judgmental confidence is higher or lower than their actual performance.

survey asked each participant to judge a set of emails randomly drawn from a pool of phishing emails and authentic business emails. We measured their judgmental confidence and accuracy, and we used two alternative operationalization metrics (i.e., overprecision and overestimation) to measure the extent of their overconfidence. The results we obtained from regression analysis show that cognitive effort reduced overconfidence while variability in attention allocation, dispositional optimism, and familiarity with the business entities all increased overconfidence in phishing detection. Perceived self-efficacy had a marginal effect on detecting phishing. In addition, all confidence beliefs—prospective confidence such as self-efficacy beliefs, and retrospective confidence (or judgmental confidence)—poorly predicted detection accuracy and poorly explained its variance, which highlights the issue of relying on confidence beliefs to guide one's behavior in phishing detection. We discuss our findings' implications at the end of the paper.

Our sample from a broad demographic basis provides evidence of the prevalence of overconfidence in phishing detection. As such, both research and practice need to pay more attention to this issue. This study makes two contributions. First, to our knowledge, this is the first systematic study on retrospective overconfidence in phishing email detection. We conceptualize and measure retrospective overconfidence as differing from overconfidence in the context of self-efficacy beliefs and, thus, help advance theories in both behavioral decision making and phishing detection. Second, the study identifies several factors that influence overconfidence, which enables researchers and practitioners to devise practical mechanisms to help users reduce bias and improve judgment, such as providing feedback to individuals in confidence training (Moores & Chang, 2009; Sharp, Cutler, & Penrod, 1988).

# 2 Conceptualization and Theoretical Background

In this section, we conceptualize overconfidence and several related concepts (Table 1 summarizes the main concepts). We also introduce corresponding theories of overconfidence's causes.

**Table 1. Explanations of Key Concepts**

| Concepts | Explanations |
|---|---|
| Confidence | There are two types of confidence in judgmental tasks (Busey et al., 2000): prospective confidence, which refers to the confidence in one's ability to make good judgments (e.g., "I can effectively detect phishing emails"); and retrospective confidence (also called judgmental confidence; Berger, 1992), which refers to the degree to which individuals believe that their judgment is accurate (e.g., "I am very much sure that my judgment on this email is correct").<br>In our study, we focus on retrospective confidence in judgments. For simplicity and consistency with literature (e.g., Keren, 1997; Tang et al., 2014), we use the terms retrospective confidence and confidence interchangeably.<br>One's confidence in judgment can be high or low regardless of what judgment one makes or its correctness. For instance, a person may judge an email to be a phishing email (or an authentic email) with high or low confidence (see examples in Section 1). |
| Accuracy | Also called judgmental performance, accuracy refers to the correctness of one's judgments in a task. |
| Overconfidence | Overconfidence refers to the extent to which confidence exceeds performance (or accuracy). Because there are two types of confidence, there are correspondingly two types of overconfidence: prospective overconfidence (e.g., when one compares their performance with their self-efficacy beliefs; Moores & Chang, 2009; Stone, 1994) and retrospective overconfidence (i.e., when one compares their performance with their judgmental confidence; Keren, 1997; Tang et al., 2014). In this study, we examine the latter.<br>As overconfidence is the difference between confidence and performance, it differs from "high" confidence: a person can be overconfident even if that individual has low confidence and if the accuracy is even lower than confidence. |

## 2.1 Confidence, Accuracy, and Overconfidence

While much research in phishing detection has focused on judgmental accuracy (i.e., whether a participant correctly identifies a phishing email or not), one should not ignore the importance of judgmental confidence. Studies in behavioral decision making suggest that judgmental confidence plays a critical role in predicting human behavior because it moderates the judgment-behavior relationship (Berger, 1992; Berger & Mitchell, 1989; Fazio & Zanna, 1978; Ho & Bodoff, 2014). These findings suggest that individuals more confident in judging emails for phishing will take more definite action in response to them (either to follow the instructions in the emails if they believe they are authentic or delete/ignore them if they believe they are phishing emails). They also suggest that individuals less confident in judging emails for phishing may make more effort to verify the authenticity of the email such as contacting the email sender by phone.

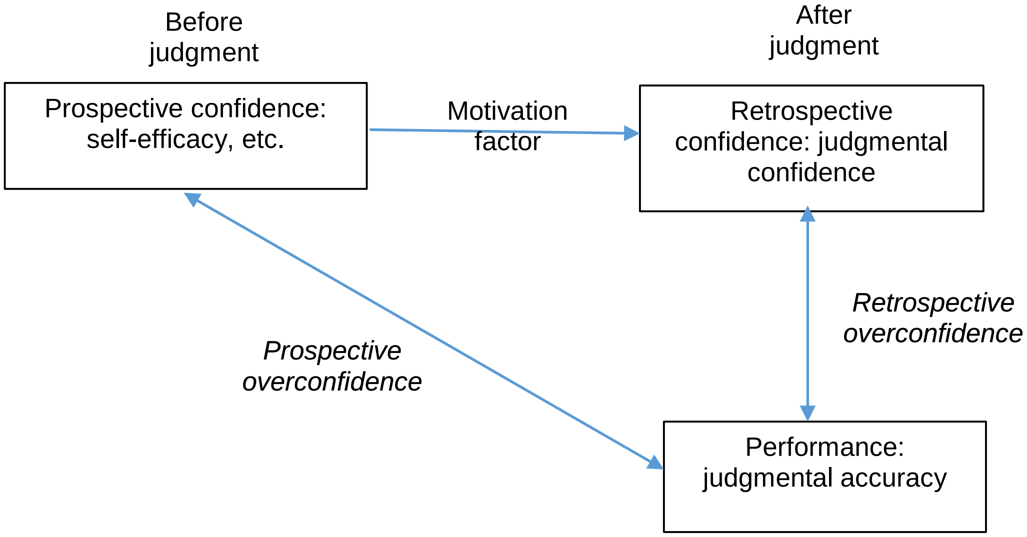

Before judgment
After judgment
Prospective confidence: self-efficacy, etc.
Motivation factor
Retrospective confidence: judgmental confidence
Retrospective overconfidence
Prospective overconfidence
Performance: judgmental accuracy

**Figure 1. Confidence, Accuracy, and Overconfidence**

Judgmental confidence differs from another type of confidence belief: self-efficacy. Self-efficacy reflects one's prospective confidence (Busey et al., 2000); that is, confidence in one's ability to accomplish a task (Bandura, 1997; Cramer, Neal, & Brodsky, 2009; Stone, 2000) rather than confidence in how well one achieves it (Stone, 2000). Social cognitive theory (Bandura, 1997) suggests that self-efficacy beliefs may exceed one's actual ability to maintain motivation and perseverance to accomplish a task. Factors such as past performance and vicarious experience (among others) determine one's self-efficacy beliefs (Bandura, 1997). On the other hand, beyond past performance and vicarious experience, the number of cues in a task also determine judgmental confidence (Bjorkman, 1992, 1994): the more information cues a task contains, the more confident one is with the task outcome. Cramer et al. (2009) comprehensively compare self-efficacy beliefs and judgmental confidence belief and show their distinctions in theoretical support, practical application, and construct composition. While prior studies have suggested the possible existence of overconfidence in the context of self-efficacy beliefs (Hayes et al., 2012; Hong et al., 2013; Stone, 1994), in this study, we focus on retrospective overconfidence. Figure 1 illustrates the relationships between these core constructs.

## 2.2 Theoretical Causes to Overconfidence

Studies in behavioral decision making and judgment under uncertainty have shown that two types of factors determine overconfidence: cognitive factors and motivational factors (Alba & Hutchinson, 2000; Keren, 1997). Cognitive factors explain how individuals process information to make judgments and attribute overconfidence to cognitive strategies that misinterpret and/or selectively process information (Koriat et al., 1980). That is, due to limited mental resource and capacity, individuals often adapt their cognitive strategies to balance the goal of being accurate and the goal of conserving limited cognitive resources (Johnson & Payne, 1985; Payne, 1976, 1982). Research has found individuals' cognitive strategies to significantly affect decision quality and overconfidence in decision making (Brucks, 1985; Kassin, Rigby, & Castillo, 1991; Payne, 1976; Siegel-Jacobs & Yates, 1996; Stone, 1994; Tetlock & Kim, 1987).

Many studies have investigated the effects of cognitive effort and attention allocation, which reflect one's cognitive process in decision making (Alba & Hutchinson, 2000; Palmer, Brewer, & Horry, 2013; Johnson & Payne, 1985; Payne, 1982). Cognitive effort in judgment plays an important role in decision quality: for example, the more time an individual expends, the better the decision the individual reaches (Johnson & Payne, 1985; Payne, 1982). Individuals generally avoid effortful reasoning and thought (Payne, 1982), and the lack of decision effort (such as time) in judgment results in overconfidence bias in decision making

(Kassin et al., 1991; Siegel-Jacobs & Yates, 1996; Tetlock & Kim, 1987). Variability in attention allocation is another important factor that influences judgmental bias because people biasedly treat incoming information as valid rather than of indeterminate validity since they need to expend additional cognitive activity to properly encode and identify information as invalid (Gilber, 1991; Alba & Hutchinson, 2000). Therefore, when individuals have constrained resources (such as decision time) or something diverts their attention, they may not properly identify invalid information and, thus, form false beliefs (Alba & Hutchinson, 2000). Variability in what one pays attention to may influence one's judgment outcomes (Keren, 1997).

The motivational factors, on the other hand, suggest that individuals make biased judgments as a self-motivating or self-enhancement (Alba & Hutchinson, 2000) mechanism. An example is the overconfidence exhibited by amateur bridge players who develop unrealistic and insensible perspectives, which leads to less accurate assessments than experts (Keren, 1997). Behavioral theories suggest that individuals are motivated to view themselves in a positive light and think that they are intelligent and knowledgeable (Taylor & Brown, 1988; Klayman, Soll, & Gonzalez-Vallejo, 1999). In the scenario of probability estimation, individuals may engage in wishful thinking and self-enhancement and, thus, estimate something (Griffin & Brenner, 2004) with excessive confidence.

Such a distorted estimation may originate not only from individuals' dispositional traits such as dispositional optimism (Scheier, Carver, & Bridges, 1994; Libby & Rennekamp, 2011) but also from their perception about their knowledge or competence of the judgment task. According to Heath and Tversky's (1991) competence hypothesis, individuals feel more confident where they consider themselves knowledgeable or competent than in a context where they feel ignorant or uninformed. What an individual knows relative to what the individual can know determines whether the individual feels competent in a given context. General knowledge, familiarity, and experience all enhance one's feeling competent possibly because individuals may have learned from experience that they generally do better in situations they understand than in situations where they have less knowledge.

Research on consumer behavior supports Heath and Tversky's (1991) competence hypothesis. The research shows that the perceived familiarity of an assertion or opinion (such as a product brand) results in a corresponding increase in its perceived validity (rather than actual validity) and a decrease in perceived uncertainty. As such, one can become overconfident (Alba & Hutchinson, 2000) because familiarity-based illusions of truth are particularly seductive: one can find it difficult to identify the source of an assertion's familiarity, and the actual validity of the assertion may not be obvious (Alba & Hutchinson, 2000). Since beliefs are precursors to decisions, as Alba and Hutchinson (2000) argue, misperceived validity will result in overly confident decisions. In addition, prior studies (Berger & Mitchell, 1989) also suggest that one can boost one's judgmental confidence via direct experience with an object or repeated exposure to the object, such as receiving the same messages repeatedly. Overall, though, little research has investigated how confidence changes with experience and familiarity (Tsai, Klayman, & Hastie, 2008).

## 2.3 Overconfidence in Phishing Detection

In judgment under uncertainty (such as in detecting phishing emails), one often cannot observe decision accuracy until the event occurs (e.g., the "phisher" steals the individual's personal information). Thus, decision makers often rely on confidence in their judgment as an indicator of accuracy to take subsequent actions (Camerer & Lovallo, 1999; Chuang & Lee, 2006; Hirshleifer & Luo, 2001). Because they do not know what they do not know (Berner & Graber, 2008), overconfident individuals fail to elicit complete and accurate information when judging something and fail to recognize the significance of data (such as misinterpreting information cues) and synthesize it (Graber, Franklin, & Gordon, 2005). Overconfidence also misguides people to take excessive risk under uncertainty (Camerer & Lovallo, 1999; Chuang & Lee, 2006; Hirshleifer & Luo, 2001). By recognizing overconfidence in phishing email detection, researchers can design mechanisms such as training programs (Kumaraguru et al., 2007) to mitigate individuals' overconfidence and enhance their judgment.

Researchers have explored overconfidence in phishing detection in only a limited way. In an experiment with 53 undergraduates, Hong et al. (2013) found that approximately 92 percent of their participants misclassified phishing emails even though 89 percent indicated they were confident of their ability to identify phishing emails. In a similar study on 30 subjects, Kumaraguru et al. (2007) asked subjects to rate, on a scale of 1 to 7, how confident they were when making judgments on 19 emails. They found that participants claimed they knew about phishing and knew how to protect themselves but still fell for the phishing scams regardless. In another study on phishing website detection, Dhamija et al. (2006) asked 22 subjects (university students and staff) to judge if a website was legitimate or not and to state their confidence in their judgments on a five-point Likert scale. They found a mismatch between the confidence rating and actual judgmental accuracy. We can see that most of the studies have used a limited demography (as most subjects are university students), have employed small sample sizes, and have focused on prospective confidence (reflected by self-efficacy beliefs). Research has yet to address what causes retrospective overconfidence when individuals attempt to detect phishing.

### 2.3.1 Measurement of Overconfidence

A practical issue in empirical studies on overconfidence concerns its measurement (Moore and Healy 2008). We adopt two metrics of overconfidence. The first, most dominant measurement is "overprecision" (Moore & Healy, 2008). It is the difference between the mean confidence (or mean subjective probability) and mean accuracy (i.e., the frequency or proportion of correct answers) of judgments in a task (Keren, 1991; Lichtenstein et al., 1982; McClelland & Bolger, 1994):

$$\text{Overprecision} = \bar{f} - \bar{d}, \qquad (1)$$

where $\bar{f} = \frac{1}{N}\sum_{n=1}^{N} f_n$ and $\bar{d} = \frac{1}{N}\sum_{n=1}^{N} d_n$, with N denoting the number of judgments in the task, $f_n$ denoting subjective probability (or the confidence) in each judgment, and $d_n$ denoting the accuracy of each judgment. We measured confidence ($f_n$) after each email judgment; its value could range between 50 percent and 100 percent, where 50 percent means by chance and 100 percent means very certain (Juslin & Olsson, 1997). We set the value of $d_n$ to 1 when one correctly judges an email as a phishing email or not and 0 for incorrectly judging it as a phishing email or not (Yates, 1982).

The second measurement of overconfidence is "overestimation". It is the difference between perceived accuracy (i.e., self-estimated frequency of correct judgments, which we call "self-estimated correctness" in this study and denote it with $\tilde{d}$) and actual accuracy ($\bar{d}$):

$$\text{Overestimation} = \tilde{d} - \bar{d}, \qquad (2)$$

In our study, we measured perceived accuracy or self-estimated correctness ($\tilde{d}$) after each subject judged all emails as phishing emails or not.

The literature shows that overprecision and overestimation are positively related because they share a common psychological basis in subjective feelings of competence (Larrick, Burson, & Soll, 2007). They also have distinct characteristics: overestimation is based on individuals' overall estimation of their performance (i.e., frequency of correct answers) in *N* judgments, while overprecision is based on the average of the subjective probabilities (i.e., $f_n$) assigned to each of the *N* judgments. Based on the probabilistic mental model (PMM), Gigerenzer, Hoffrage, and Kleinbölting (1991) argue that there is a systematic difference between probability judgment and frequency judgment: people do not estimate the frequency of correct responses by averaging the probabilities across all items but by judging the items' overall difficulty, and the overall difficulty is a product of reflecting on those items. Therefore, the estimate of the overall performance of the N judgments is generally more accurate, though the average of the probabilities commonly exceeds the overall accuracy (Gigerenzer et al., 1991). We include and compare both in our empirical test.

## 3 Hypothesis Development

Guided by the conceptual framework we discuss in Section 2.2., we recognize cognitive and motivational factors that may influence overconfidence in phishing detection. We use "cognitive effort" and "variability in attention allocation" among email judgments to capture cognitive processes that a person may engage with in detecting phishing emails (Alba & Hutchinson, 2000; Palmer et al., 2013; Johnson & Payne, 1985; Payne, 1982). We consider both personal traits related to overconfidence and perceptions regarding one's knowledge or competence in phishing detection as motivational factors. To capture personal traits associated with overconfidence, we include "dispositional optimism" (Scheier et al., 1994). To capture perception regarding one's knowledge or competence in phishing detection, we include "perceived familiarity with the business entities" (Alba & Hutchinson 2000; Li, 2013) and "perceived self-efficacy of detecting phishing" (Alba & Hutchinson, 2000; Stone, 1994). Figure 2 illustrates the relationships between these predictors and overconfidence.

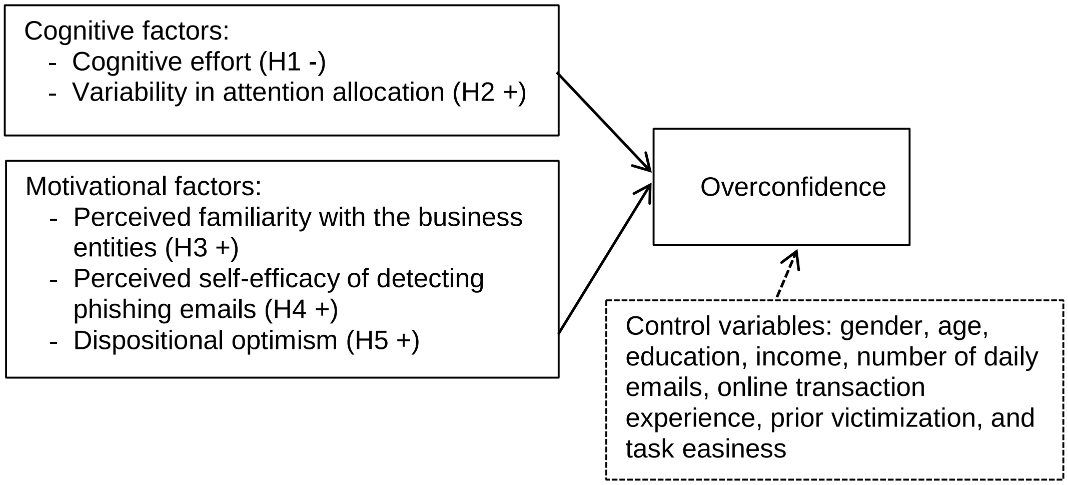

Cognitive factors:
- Cognitive effort (H1 -)
- Variability in attention allocation (H2 +)
Motivational factors:
- Perceived familiarity with the business entities (H3 +)
- Perceived self-efficacy of detecting phishing emails (H4 +)
- Dispositional optimism (H5 +)
Overconfidence
Control variables: gender, age, education, income, number of daily emails, online transaction experience, prior victimization, and task easiness

**Figure 2. Research Model**

## 3.1 Impacts of Cognitive Factors

From the information processing perspective, we expect that the more cognitive effort an individual spends on examining an email to reach a judgment, the more accurate and confident the judgment will be, which will yield less overconfidence. A rapid judgment may lead to substantial decision errors because individuals may not be able to examine a wider range of conceivably relevant cues. With increased effort in processing emails, individuals may be more likely to discover inconsistent cues and notice the abnormality of a phishing email and raise suspicion and, therefore, increase their accuracy in identifying such emails.

Researchers have often used time spent to make a decision to reflect one's cognitive effort (Bettman, Johnson, & Payne, 1990; Garbarino & Edell, 1997; Johnson & Payne, 1985; Payne, 1982). Studies on eyewitness identification decisions provide evidence about the impact of time to judgment on overconfidence. For example, Weber and Brewer (2004) found that exposure duration of suspects' photos influenced eyewitness' overconfidence. In particular, shorter exposure durations decreased accuracy more than it decreased confidence and, thus, increased overconfidence; longer exposure durations increased accuracy more than it increased confidence and, thus, reduced overconfidence; in both scenarios, though, the authors observed some extent of overconfidence in their experiment. Palmer et al. (2013) repeated the eyewitness experiment and observed similar findings; that is, shorter exposure durations yield greater overconfidence than longer exposure durations.

In terms of detecting phishing emails, we expect that cognitive effort will have the same effect as research has shown exposure duration to have. First, cognitive effort influences to what extent an individual adequately analyzes information cues in an email to make a judgment. With the advances in technology, phishing emails are becoming increasingly deceptive and demand that users systematically examine various elements in the emails such as the senders, subject lines, and email bodies (Vishwanath, Herath, Chen, Wang, & Rao, 2011).  Doing so requires extensive effort. Nevertheless, with the increasing use of email as a popular means of communication, people feel more and more time pressure in dealing with information overload and, thus, may not spend enough time or pay enough attention to examine the emails, which may lead to misjudgment. For example, they may ignore standard security indicators such as the actual addresses in the emails (Dhamija et al., 2006; Pattinson, Jerram, Parsons, McCormac, & Butavicius, 2012) or disregard security warnings (Vance, Anderson, Kirwan, & Eargle, 2014). Therefore, more cognitive effort has a positive impact on judgmental accuracy. In terms of confidence, similarly, more effort spent on judging an email helps one find evidence or information supporting one's judgment, which increases one's confidence (Bjorkman, 1994; Gigerenzer et al., 1991). Studies in the decision making literature have well documented the effect of information on confidence (Peterson & Pitz, 1988). Taken together, we expect that both judgmental confidence and accuracy will improve with cognitive effort (such as time to decision) and that the correspondence between the two will improve as well (Weber & Brewer, 2004) due to the systematic investigation of information and increased certainty in one's judgment. Therefore, we hypothesize that:

**H1:** Cognitive effort reduces overconfidence in phishing email detection.

Given a sequence of judgments on emails' authenticity (such as processing one's emails in the mailbox), individuals need to decide how to spend their limited mental resources across the emails. Individuals may choose to spread their effort across all their emails or focus on only a certain number. We argue that individuals who more greatly vary in how much attention they pay to a set of emails is associated with their using heuristic decision rules that requires less or incomplete information cues when judging them for authenticity (Payne, 1976). Such an attention-allocation strategy focuses on the more important, relevant, or confirming information from the decision maker's point of view (Alba & Hutchinson, 2000; Payne, Bettman, & Luce, 1996), such as email layout, sender's address, grammar, embedded URLs, or other information cues presented in emails. Such variations in attention result in individuals' selectively retrieving supporting evidence and neglecting disconfirming evidence, which contributes to overconfidence (Alba & Hutchinson, 2000; Koriat et al., 1980). In other words, individuals may engage in such a strategy if they notice that an email cue confirms their initial feeling, and, thus, they may make an immediate judgment on the email's authenticity. Individuals may spend more time on the emails where they cannot spot the cues confirming their initial hypothesis. Other studies show that individuals who are less selective in processing information are less susceptible to overconfidence bias and are more accurate in their judgments (Creyer, Bettman, & Payne, 1990; Stone & Schkade, 1994). In detecting phishing emails in particular, individuals may examine all necessary cues (several could exist) for all emails before reaching their judgment on them. Therefore, we hypothesize that:

**H2:** Variability in attention allocation in an email increases overconfidence in phishing email detection.

## 3.2 Impacts of Motivational Factors

Following the competence hypothesis (Heath & Tversky, 1991), we study perceived familiarity with the business entities in the emails and perceived self-efficacy to capture individuals' perceptions regarding their knowledge or competence in detecting phishing. We define perceived familiarity with the business entities as the degree to which one feels familiar with the business entities related to emails. On the one hand, we argue that perceived familiarity with the business entities may improve judgmental accuracy because it allows people to learn about the business entities (and related communication) and various practices in e-commerce, which helps them to discern between a legitimate and a malicious activity. Such experience helps one understand information cues in an email and judge whether the information cues point to a phishing email or a genuine email. For instance, many people may assume that governmental websites in the US end up with .gov, but, in fact, some of them end up with other top-level domains. Therefore, experience and familiarity with those websites and domains helps a person to better judge emails from the websites, which improves their judgmental accuracy. Note that, in a conceptual paper, Alba and Hutchinson (2000) argue that experience leaves accuracy unchanged.

On the other hand, familiarity may lead to an excessive increase of confidence that exceeds the increase of accuracy and, thus, result in overconfidence in phishing detection. Bhandari and Hassanein (2012) explain that, due to familiarity, individuals become too wedded to their familiar views and may under-react to potentially important information. Indeed, a large number of studies have found evidence for such a phenomenon. For instance, the psychological literature shows that the more one is familiar with a field, the more likely one is at risk of being overconfident about their knowledge (Fischhoff, Slovic, & Lichtenstein, 1977; Heath & Tversky, 1991). In the financial market, Blavatskyy (2008) argues that experienced traders place more weight on their own (private) signals than inexperienced traders, which leads them to be more prone to overconfidence (Glaser, Nöth, & Weber, 2004). In the online environment, Li (2013) suggests that personal familiarity with a website changes consumer behavior; that is, it induces good feelings about a website and reduces their privacy concerns. All these studies make the point that familiarity with an entity, such as the business in the email, may boost one's overconfidence in judging it. Therefore, we hypothesize that:

**H3:** Perceived familiarity with the business entities in emails increases overconfidence in phishing email detection.

We define perceived self-efficacy of detecting phishing as individuals' belief in their own ability to recognize phishing emails. Perceived self-efficacy of detecting phishing reflects one's expected performance in phishing detection. While Alba and Hutchinson (2000) argue that the inflated feelings of self-efficacy may have a motivational effect that ultimately results in higher levels of performance, Stone (1994) suggests that the initial, "first-impression" self-efficacy judgments are biased toward overestimates of personal ability.

Further research shows that self-efficacy may increase the likelihood of one's committing logic errors in analytic games and leads to overconfidence (Vancouver, Thompson, Tischner, & Putka, 2002). Such a positive self-efficacy expectation is likely to increase post-decision perceptions of performance more than it increases the actual performance and, thus, leads to overconfidence (Moores & Chang, 2009).

We expect that self-efficacy will have the same effect on phishing email detection. On the one hand, self-efficacy of detecting phishing may decrease one's chance of being phished. Because many phishing emails pretend to introduce a problem and then offer a solution in the same message, such as recovering the password, releasing space in mailbox, checking the status of a package, and so on, such a scheme may work for low self-efficacy individuals (Wright & Marett, 2010). High self-efficacy individuals may have relevant knowledge and skills to recognize those messages as phishing. On the other hand, we expect self-efficacy to have a positive effect on confidence in phishing detection (see Figure 1). From the motivational theory perspective, Vancouver et al. (2002, p. 507) cite Bandura and Jourden (1991) and argue that self-efficacy "creates little incentive to expend the increased effort needed to attain high levels of performance". In other words, once individuals spend their amount of expected effort, they may feel confident in the output and discontinue the effort. Individuals with higher self-efficacy of detecting phishing may expect phishing detection to be effortless and, therefore, may feel confident in their task performance. Taken together, we suggest that, although self-efficacy increases both detection accuracy and confidence, it increases confidence even more, which leads to overconfidence. Therefore, we hypothesize that:

**H4:** Perceived self-efficacy of detecting phishing emails increases overconfidence in phishing email detection.

Prior literature also suggests that personal traits such as dispositional optimism, a generalized positive expectation for the future (Scheier et al., 1994), results in overconfidence in judgment under uncertainty (Libby & Rennekamp, 2011). From a positive perspective, research has demonstrated that dispositional optimism generally improves both an individual's physical (Peterson & Bossio, 2001) and psychological well-being (Diener, Sub, Lucas, & Smith, 1999). However, a high level of dispositional optimism can also lead to one's relying on biases and heuristics that result in overconfidence (Hayward, Shepherd, & Griffin, 2006), which research has shown to undermine one's decision making ability in an uncertain environment (Gilovich, Griffin, & Kahneman, 2002; Hmieleski & Baron, 2009; Lovallo & Kahneman, 2003). Specifically, dispositional optimism influences a person's desirability for certain outcomes and the confidence in self-ability to achieve those outcomes, but its effect on actual results is unfounded (Alba & Hutchinson, 2000). Following this logic, in the scenario of detecting phishing, individuals with high dispositional optimism may have high confidence in dealing with phishing attacks, but their detection accuracy may not be necessarily better. Therefore, we hypothesize that:

**H5:** Dispositional optimism increases overconfidence in phishing email detection.

In addition to the above research constructs, we included a few control variables in the study, including online transaction experience (i.e., the degree to which an individual uses the Internet as a channel to carry out business transactions such as buying products online, paying bills and using online banking, and trading stocks online), gender, age, education level, income, number of daily emails, and prior victimization related to identity theft (Pattinson et al., 2012; Vishwanath et al., 2011). We also included task easiness (i.e., how easy it is to recognize the nature of a given email set, legitimate or not, for an average person) as a control variable due to the hard-easy effect in judgmental tasks (particularly for overprecision; Juslin & Olsson 1997; Keren, 1997). In Section 4, we describe the research method we employed to empirically test the model.

# 4 Research Method and Data Collection

## 4.1 Experimental Design

We developed a survey experiment using Qualtrics Research Suite to collect data for model testing. In this experiment, each participant judged 16 emails (presented as images of the emails) randomly chosen from a pool of 50 phishing emails and genuine business emails. We measured their judgmental confidence, accuracy, and cognitive and motivational factors. Appendix A describes how we developed the email pool.

In the recent past, some researchers have studied phishing detection by relying on individuals' responses to images of phishing emails (or websites): these studies presented participants with images of emails (or websites) mixed with phishing and genuine business ones and asked how they would respond if they saw such emails or websites (Downs et al., 2006; Downs, Holbrook, & Cranor, 2007; Pattinson et al., 2012; Sheng, Holbrook, Kumaraguru, Cranor, & Downs, 2010; Vishwanath et al., 2011; Wang, Herath, Chen, Viswanath, & Rao, 2012) or alternately queried if the e-mails or websites were genuine (Anandpara, Dingman, & Jakobsson, 2007; Dhamija et al., 2006; Furnell, 2007). Research has proven this approach to be the most efficient in collecting large samples of data in phishing detection research given the ethical concerns on collecting data using other approaches (Pattinson et al., 2012). Industry has also used such an approach to help laypeople become aware of their ability to detect phishing attacks (see http://www.sonicwall.com/furl/phishing/).

Understanding that participants cannot interact with the email images in this survey experiment, we selected emails that contained various detection cues revealed in the images besides clear message content and email layout. For example, some email images showed a mouse-over URL at the bottom if the image hid the URL. Some emails contained an attachment (see the example in Appendix A). In addition, all the emails had the senders' email address visible. We modified the receivers' addresses if it was private. We did not change the receivers' address if the email had no receiver address shown, was sent to a group (for example, undisclosed recipients), or was clearly not a private email (for example, Paypal as the receiver). In the consent form of the survey, we emphasized that all these emails were the actual ones except where we removed private information.

The experiment proceeded as follows. First, participants filled out a consent form that described the objective of the study and instructions for the experiment. We then presented them with items that measured perceived self-efficacy of detecting phishing. Subsequently, each participant finished the email judgment task that sequentially presented 16 email images that they judged as legitimate or not. In the consent form of the survey, we stated that, if the participant believed the email was truly sent from the business entity it claimed to be (i.e., a genuine business email), the participant should choose “yes", and, if the participant believed that the email was from someone pretending to be what the email claimed to be (i.e., a phishing email), the participant should choose "no". We also asked the participants to indicate, on a scale from 50 (by chance) and 100 (very confident), how confident they were for each judgment (i.e., $f_n$). We then asked the participants how familiar they were with the business entity indicated in the email. After the participants judged all the 16 emails, they each estimated the percentage of correct answers they made (i.e., $\tilde{d}$), which we used as an alternative measure of confidence (Moore & Healy, 2008).

Figure A1 in the Appendix shows an example of the included emails and the four questions associated with each judgment. All 16 emails followed the same presentation format. At the end, the participants completed other measurements including dispositional optimism, gender, age, income, education, the number of emails received daily, online transaction experience, and prior victimization.

We pretested the survey protocol with a group of faculty members, PhD students, undergraduate students, and administrative staff from a research university in southwestern US. Subsequently, we pilot tested it with a group of undergraduate students from the same university before actually collecting data. The pretest helped to ensure the clarity and content validity of the instrument, and the pilot test helped to verify the response from the participants. We found that 16 emails were ideal for the study, which allowed each participant to complete the experiment in about 15 minutes. We made minor changes to the survey following the feedback from the pretest and the pilot test.

## 4.2 Measurement of Independent Variables

We measured the two cognitive factors with objective data. We recorded time elapsed between showing the email image and the first click the participant made (which was the click needed to judge whether the mail was legitimate or not) and used the sum of the time spent for the 16 judgments to measure cognitive effort (H1) (Bettman et al., 1990; Garbarino & Edell, 1997; Johnson & Payne, 1985; Payne, 1982). Following prior literature in behavioral decision making (Brucks, 1985; Payne, 1976; Stone, 1994), we calculated the coefficient of variation (CV) of the time expended on the emails to measure variability in attention allocation (H2), where high CV indicated a high variability.

We measured the other independent variables with self-reported data (see Table 2). First, we measured perceived familiarity with business entity in an email (H3) with a single item. Since the unit of analysis of overconfidence was at an individual level (instead of individual-judgment level) as prior research has shown (Brenner, Koehler, Liberman, & Tversky, 1996; Griffin & Brenner, 2004; Klayman et al., 1999; Lichtenstein et al., 1982; Lichtenstein & Fischhoff, 1977; Pulford, 1996), we aggregated measures of business entity familiarity from the 16 emails. We adapted two items from prior literature (Chen, Wang, Herath, & Rao, 2011;

Herath et al., 2012) to measure perceived self-efficacy of detecting phishing (H4). We calculated the latent score of perceived self-efficacy by loading these two items together based on a factor analysis that exhibited satisfactory reliability and convergent validity. We assessed dispositional optimism (H5) by the aggregated score of the 10-item Life Orientation Test Revised (LOT-R) (Scheier et al., 1994).

**Table 2. Survey Items**

| Constructs | Items |
|---|---|
| Perceived familiarity with business entities | How familiar do you think you are (not at all, a little, some, much, or a lot) with the business entity indicated in the email?<br>We aggregated the above item from 16 emails for the construct. |
| Perceived self-efficacy of detecting phishing (Chen et al., 2011; Herath et al., 2012) | Please indicate to what extent (strongly disagree, disagree, neither agree nor disagree, agree, or strongly agree) you agree with each of the following statements:<br>• I can recognize phishing emails.<br>• I can differentiate phishing emails from legitimate ones. |
| Dispositional optimism (Scheier et al., 1994) | Please indicate to what extent (strongly disagree, disagree, neither agree nor disagree, agree, or strongly agree) you agree with each of the following statements:<br>• In uncertain times, I usually expect the best.<br>• It's easy for me to relax. (Filler item; discarded from analysis)<br>• If something can go wrong for me, it will. (Reverse-worded)<br>• I'm always optimistic about my future.<br>• I enjoy my friends a lot. (Filler item; discarded from analysis)<br>• It's important for me to keep busy. (Filler item; discarded from analysis)<br>• I hardly ever expect things to go my way. (Reverse-worded)<br>• I don't get upset too easily. (Filler item; discarded from analysis)<br>• I rarely count on good things happening to me. (Reverse-worded)<br>• Overall, I expect more good things to happen to me than bad. |
| Online transaction experiences | How often  (never, rarely, sometimes, most of the time, or always) do you engage in the following online activities?<br>• Buying and selling stocks or mutual funds online.<br>• Buying products or services online with a credit card, a debit card, or a payment service such as PayPal.<br>• Paying bills (such as electronic, utility, credit cards, or loans) online.<br>• Accessing bank accounts (such as checking, savings, or mortgage) online. |
| Prior victimization of identity theft | Have you ever experienced the following situations? (Yes or no)<br>• Someone used or attempted to use your personal information without permission to obtain new credit cards or loans, run up debts, open other accounts, or commit other fraud.<br>• Someone used or attempted to use your credit cards without permission.<br>• Someone used or attempted to use your accounts such as your wireless phone account, bank account, or debit/check cards without your permission. |

We measured control variables such as gender, age, education level, income, and the number of daily emails with single items. We measured online transaction experience with the sum of four items (see Table 2); each captured an aspect of the experience. We measured prior victimization of identity theft as the sum of three items (see Table 2). We also used another control variable, task easiness, to measure the overall easiness (or difficulty) of judging the 16 emails randomly assigned to a subject because research has found that people tend to be overconfident in dealing with difficult tasks rather than easy tasks, known as the hard-easy effect (Juslin & Olsson, 1997; Keren, 1997). We calculated it based on the mean value of the easiness of the emails that a subject received, and we derived the easiness of each email from the proportion of subjects who judged the same email correctly.

## 4.3 Survey Administration

We collected data relying on the professional survey service company, Qualtrics, who draw data from a sample from the US population[2]. Compared with a lab experiment in which one invites participants to finish an experiment in a lab room at a given time, using Qulatrics with a survey allows one to access a

[2] A brochure explaining how Qualtrics recruits their participants is available on request from the first author.

broader demographic base and a larger sample size economically and to provide a better generalizability of our results. We filtered out those respondents who did not have any online transaction experience (answered “never” to all four items measuring online transaction experience) given they were not relevant to the phishing attacks in our study (which we designed to solicit online financial account, credit number, and other personal information). We collected 600 valid responses from 47 States in US. Thirty-six percent of the respondents were male. The mean age was 52, with a range from 19 to 89. Seventy-six percent (76.3%) of respondents had an education of some college or more. Most respondents were white/not Hispanic (84%). Forty-four percent (44.3%) of respondents had a family income higher than US$50,000. Sixty percent of the respondents had more than one credit card in their wallet or purse.

# 5 Data Analysis and Results

## 5.1 Descriptive Analysis

We first conducted a descriptive analysis of overconfidence in phishing detection based on the two measurement metrics (see Equations 1 and 2) to learn its extent. Table 3 presents the ranges, means, and standard deviations of the metrics, and Figure 3 illustrates their statistical distributions. At an aggregated level, the results suggest the prevalence of overconfidence in phishing detection in the overprecision measure only (mean = .14, standard deviation = .17), but not in the overestimation measure (mean = 0.00, standard deviation = .24). In particular, 80 percent of the participants had a score of overprecision greater than 0, while 45 percent had a score of overestimation greater than zero.

**Table 3. Descriptive Statistics of the Overconfidence Metrics (overprecision and overestimation) and Related Measures**

| Metric | Minimum | Maximum | Mean | Standard deviation |
|---|---|---|---|---|
| Overprecision | -.31 | .72 | .14 | .17 |
| Overestimation | -.75 | .63 | .00 | .24 |
| Accuracy ($\bar{d}$) | .25 | 1.00 | .67 | .15 |
| Subjective probability ($\bar{f}$) | .50 | 1.00 | .81 | .12 |
| Self-estimated correctness ($\tilde{d}$) | .00 | 1.00 | .68 | .21 |

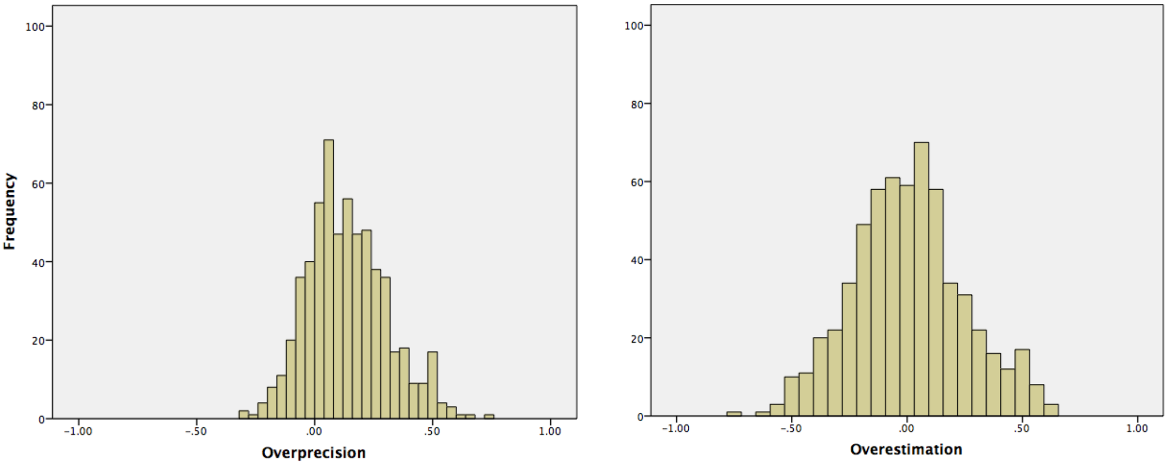


**Figure 3. Histogram of the Overconfidence Metrics**

The proportion of participants exhibiting overestimation was significantly less than that exhibiting overprecision, which supports Gigerenzer et al.’s (1991) view that frequency estimates (i.e., overestimation) are generally more accurate than probability estimates (i.e., overprecision). We also compared the difference between subjective probability ($\bar{f}$) and self-estimated correctness ($\tilde{d}$) in the same person and found a significant difference based on a paired t-test. The results seem to confirm the systematic difference between probability judgment and frequency judgment in the line of Gigerenzer et al. (1991).

Table 4 presents the bivariate correlations between the variables in our research model. Overprecision and overestimation were highly correlated (r = .628), which echoes prior studies (Larrick et al., 2007).

**Table 4. Correlation Coefficients**

| | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|
| (1) Overprecision | .628** | -.723** | .474** | .196** | -0.07 | .154** | .224** | 0.061 | .168** |
| (2) Overestimation | -- | -.515** | .219** | .778** | -0.054 | .111** | .088* | .086* | .169** |
| (3) Accuracy | -- | -- | .266** | .138** | .169** | -.165** | 0.032 | .159** | -0.02 |
| (4) Subjective probability | -- | -- | -- | .449** | .118** | 0.004 | .354** | .288** | .210** |
| (5) Self-estimated correctness | -- | -- | -- | -- | 0.062 | 0.007 | .126** | .216** | .181** |
| (6) Cognitive effort | -- | -- | -- | -- | -- | .330** | 0.02 | 0.033 | 0.036 |
| (7) Variability in attention | -- | -- | -- | -- | -- | -- | 0.03 | -0.018 | -0.002 |
| (8) Entity familiarity | -- | -- | -- | -- | -- | -- | -- | .151** | .089* |
| (9) Self-efficacy | -- | -- | -- | -- | -- | -- | -- | -- | .139** |
| (10) Disp. optimism | -- | -- | -- | -- | -- | -- | -- | -- | -- |

* Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed).

## 5.2 Hypotheses Testing

To test the hypotheses, we conducted an OLS regression analysis on each dependent variable with SPSS (ver. 21). To examine potential multicollinearity among the constructs in the regression models, we performed the variance inflation factor (VIF) test. The VIFs were all well below 3.3, which indicates that multicollinearity was not a cause of concern (Petter, Straub, & Rai, 2007). We also plotted the histogram and the normal P-P plot of standardized residuals for all regression models, and both did not present evidence of violating the normality assumption required for OLS regressions.

**Table 5. Regression Results (Standardized Coefficients)**

| Independent variables | Dependent variables | | | | |
|---|---|---|---|---|---|
| | Over-precision | Over-estimation | Accuracy | Subjective probability | Self-estimated correctness |
| Cognitive effort (H1) | -0.137*** | -0.102** | 0.239*** | 0.113** | .058 |
| Variability in attention (H2) | 0.189*** | 0.141*** | -0.243*** | -0.046 | -.014 |
| Familiarity with entities (H3) | 0.167** | 0.044 | 0.046 | 0.291*** | .084* |
| Self-efficacy (H4) | 0.019 | 0.075^ | 0.145*** | 0.210*** | .193*** |
| Disp. optimism (H5) | 0.145*** | 0.147*** | -0.043 | 0.147*** | .138*** |
| Task easiness | -0.142*** | -0.125** | 0.223*** | 0.086* | .020 |
| Transaction experience | 0.077^ | -0.017 | -0.064 | 0.025 | -.066 |
| Age | -0.007 | 0.019 | 0.019 | 0.016 | .036 |
| Gender | -0.095** | -0.171*** | 0.133*** | 0.036 | -.100** |
| Number of emails | -0.015 | 0.014 | 0.058 | 0.052 | .059 |
| Income | -0.030 | 0.002 | 0.068^ | 0.045 | .053 |
| Education | 0.003 | 0.043 | 0.012 | 0.011 | .058 |
| Prior victimization | -0.076^ | -0.053 | 0.079 | -0.006 | -.004 |
| $R^2$ | .153 | .110 | .189 | .230 | .106 |
| Adjusted $R^2$ | .134 | .090 | .171 | .213 | .086 |

^ p-value < 0.10 (2-tailed) ; * p-value < 0.05 (2-tailed) ; ** p-value < 0.01 (2-tailed); *** p-value < 0.001 (2-tailed)

Table 5 presents standardized coefficients of the regression results. The results show that cognitive effort reduced overprecision ($\beta$ = -.137, p < .001) and overestimation ($\beta$ = -.102, p < .01). In particular, the decrease of overconfidence arises due to a stronger increase in judgmental accuracy ($\beta$ = .239, p < .001) than in confidence (β =.113, p < .01 for subjective probability, and β = .058, p > .10 for self-estimated correctness). Thus, our results support H1.

High variability in attention allocation increased overprecision (β = .189, p < .001) and overestimation *(β* = .141, p < .001). In particular, the decrease in accuracy caused an increase in overconfidence (β = -.243, p < .001), and, at the same time, high variability in attention allocation did not significantly influence confidence (β = -.046, p > .1 for subjective probability, and β = -.014, p > .1 for self-estimated correctness). Thus, our results support H2.

The results also show that familiarity with business entities increased overprecision significantly (β = .167, p < .01) but not overestimation (β = .044, p >.10). In particular, familiarity with business entities significantly increased subject probability (β = .291, p < .001) but had no significant impact on accuracy (β = .046, p > .10), which resulted in increased overprecision. It also increased self-estimated correctness significantly (β = .084, p < .05) but not large enough to result in overestimation. The non-effect on accuracy seems to support Alba and Hutchinson's (2000) argument that experience leaves accuracy unchanged. Thus, our results partially support H3.

Furthermore, self-efficacy in phishing detection marginally increased overestimation *(β* = .075, p < .10), and its effect on overprecision was insignificant. Examining its impacts on accuracy and confidence showed that self-efficacy in phishing detection increased both accuracy (β = .145, p < .001) and confidence (β = .210, p < .001 for subjective probability, and β = .193, p < .001 for self-estimated correctness). Therefore, our results only very weakly support H4.

Finally, dispositional optimism increased overprecision (β = .145, p < .001) and overestimation (β = .147, *p* < .001). In particular, dispositional optimism significantly influenced confidence (β = .147, *p* < .001 for subjective probability and β = .138, *p* < .001 for self-estimated correctness), and, as expected, it had no significant impact on accuracy (β = -.043, *p* > .1). Thus, our results support H5.

For the control variables, gender had a significant impact on overprecision (β = -.095, *p* < .01) and overestimation (β = -.171, *p* < .001), which suggests that women are better than men in detecting phishing emails. We found that task easiness reduced overprecision (β = -.142, *p* < .001) and overestimation (β = -.125, *p* < .01). Such outcomes (see Table 5) result from its strong impacts on accuracy (β = .223, *p* < .001) but a relatively weak impact on confidence (β = .086, *p* < .05 for subjective probability and β = .020, *p* > .10 for self-estimated correctness), which is consistent with the hard-easy effect shown in prior studies (particularly for overprecision; Juslin and Olsson 1997; Keren 1997). None of the other control variables had significant impacts on the overconfidence metrics. The R$^2$s of the dependent variables ranged from .110 to .230.

## 5.3 Predicting Accuracy with Judgmental Confidence

We ran a sequence of regressions using both judgmental confidence (i.e., subjective probability and self-estimated correctness) and perceived self-efficacy to predict accuracy. Table 6 shows the results for the regression models. We can see that all the R$^2$s were very low, which suggests that, overall, all these confidence beliefs are poor predictors of accuracy, so that confidence is not a reliable indicator of judgment accuracy. The results highlight the issue of relying on judgmental confidence to guide one's behavior in phishing detection.

**Table 6. Predicting Accuracy with Confidence Beliefs (Standardized Coefficients)**

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|---|---|---|---|---|---|---|
| Self-efficacy | .159*** | -- | .090* | -- | .136*** | .089* |
| Subjective probability | -- | .266*** | .240*** | -- | -- | .234*** |
| Self-estimated correctness | -- | -- | -- | .138*** | .108** | .013 |
| R$^2$ | .025 | .071 | .078 | .019 | .037 | .078 |

* p-value < 0.05 (2-tailed) ; ** p-value < 0.01 (2-tailed); *** p-value < 0.001 (2-tailed)

These results also suggest that, in terms of predicting phishing email detection accuracy, the subjective probability measure of confidence performs better than the other two measures (i.e., both self-efficacy and self-estimated correctness). Models 1 to 3 compared subjective probability with perceived self-efficacy. With self-efficacy alone (β = .159, p < .001), Model 1 explained 2.5 percent of variance in detection accuracy; with subjective probability alone (β = .266, p < .001), Model 2 explained 7.1 percent of variance in accuracy; with both self-efficacy (β = .090, p < .05) and subjective probability (β = .240, p < .001), Model

3 explained 7.8 percent of variance in accuracy. These three models indicate that perceived self-efficacy only marginally contributed to judgmental accuracy and that subjective probability can partial it out. Models 1, 4 and 5 compared self-estimated correctness with self-efficacy. The results show that both factors had similar impacts on accuracy. Model 6 compared all three confidence beliefs simultaneously. It shows that subjective probability (β = .234, p < .001) had the strongest effect on detection accuracy, while the effect of self-efficacy (β = .089, p<.05) was marginal and the effect of self-estimated correctness ( $\beta$ = .013, p>.1) was insignificant.

# 6 Discussion and Conclusions

Extending prior literature in behavioral decision making and judgment under uncertainty, we examined overconfidence in phishing email detection. We distinguished between retrospective overconfidence (caused by judgmental confidence such as the subjective probability of an answer's correctness) from prospective overconfidence (caused by self-efficacy beliefs), which highlights the importance of studying the former. We examined the impacts of five antecedents on retrospective overconfidence. In a survey experiment on 600 online users, we found that cognitive effort decreased overconfidence and that variability in attention allocation, dispositional optimism, and familiarity with the business entities in emails all increased overconfidence in phishing email detection. Perceived self-efficacy increased both judgmental confidence and accuracy, but its effect on overconfidence was marginal. Further, all confidence beliefs poorly predicted detection accuracy and poorly explained its variance, which highlights the issue of relying on judgmental confidence to guide one's behavior in detecting phishing. In comparing the various types of confidence beliefs (see Table 6), we found that judgmental confidence, in the form of subjective probability, had a relatively stronger power to predict detection accuracy than both self-efficacy beliefs and self-estimated correctness.

The marginal effect of self-efficacy on retrospective overconfidence deserves further scrutiny. As we can see from Table 5, self-efficacy belief was positively associated with judgmental accuracy and retrospective confidence (which includes subjective probability and self-estimated correctness). However, its relationship to overconfidence is quite weak. In other words, self-efficacy belief did not increase retrospective confidence more than it increased task performance. Busey et al. (2000) argue that retrospective confidence (especially subjective probability) and accuracy are based on information from the same sources, such as information in one's memory that one recalls to judge something, but prospective confidence (or self-efficacy) is based on one's expectation of capability in achieving the task. Therefore, in phishing detection, cognitive processes (measured by cognitive effort and variability in attention allocation in our study) or situational motivational factors (such as perceived familiarity with business entities) rather than the pre-task self-efficacy beliefs more significantly influence the correspondence between retrospective confidence and accuracy. This result further justifies the need to distinguish between retrospective confidence and prospective confidence.

## 6.1 Contributions and Implications

The study has two potential contributions. First, to the best of our knowledge, this study is the first one on retrospective overconfidence in phishing email detection. Its distinction from prospective overconfidence (see Table 1 and Figure 1) helps advance theories in both behavioral decision making and phishing detection. As we mention above, prior studies on individuals' phishing detection abilities overemphasize prospective overconfidence, which ignores the fact that retrospective overconfidence, caused by judgmental confidence such as subjective probability, is more directly related to judgments and subsequent behaviors (Tang et al., 2014). Following recent literature on the distinction between self-efficacy and judgmental confidence (Busey et al., 2000; Cramer et ., 2009), our empirical test results show that, while both play a role in predicting phishing detection accuracy, judgmental confidence partials out the effect of self-efficacy (see Table 6), which explains why self-efficacy only has a marginal effect on overconfidence. This finding suggests that future research on behavioral decision making and on phishing detection should incorporate judgmental confidence as an important factor to consider.

On the other hand, one should interpret the importance of judgmental confidence with caution due to the prevalence of overconfidence in judgments, especially overprecision (see Figure 3a). Our study reveals that overconfidence is quite common in phishing email detection. This finding suggests that we need to recognize sources of overconfidence and help devise mechanisms to reduce such bias. Our results (Table 5) show that three factors (i.e., variability in attention allocation, dispositional optimism, and familiarity with the business entities) increase and that one factor (i.e., cognitive effort) reduces overconfidence. As such, our results present a starting point to conduct further research to investigate other motivational and/or cognitive factors that may influence overconfidence. The conceptualization and measurement of overconfidence (see Equations 1 and 2) enables further research on this topic.

Note that we studied two forms of overconfidence. We found at the aggregate level that subjects exhibited overprecision (not overestimation). We explain in Section 5.1 that probability judgment (i.e., overprecision) and frequency judgment (i.e., overestimation) differ and that the latter tends to be more accurate than the former (Gigerenzer et al., 1991): individuals may not know whether they picked the right answer to each question but may know how many questions they answered correctly. The finding has implications for behavioral decision making research in which researchers continue to debate judgmental bias or simply the manipulation of task difficulty causes overconfidence (Keren, 1997). Although it goes beyond our scope here to resolve the issue, the unbiased overestimation measure in our study (see Figure 3-b) suggests that one may use it as a benchmark to verify the overall difficulty of the judgmental tasks in research: if there were overestimation, the tasks might be overly difficult for the subjects and vice versa.

Second, our recognizing the antecedents to overconfidence, along with some significant control variables, enables researchers and practitioners to devise practical mechanisms to help users reduce such bias and improve judgment. Klayman et al. (1999, p. 218) point out that “we are often called upon to make a choice between two alternatives, and then our subjective confidence in that choice determines how much we commit to one course, how much we seek further information, and how much we hedge our bets”. Our study shows that one’s subjective confidence may not provide an accurate guidance in phishing detection. Therefore, when we educate individuals about the knowledge and skills to detect phishing, we need to make them aware of such judgmental bias and teach them the strategies to reduce or eliminate the bias (Fischhoff, 1982; Larrick, 2004). One may do so by offering warnings and feedback (Moores & Chang, 2009; Sharp et al., 1988) based on the overconfidence metrics we introduced: for example, one could assess users for their overconfidence before and after training programs to improve their self-awareness of the bias and willingness to learn to reduce the bias via training.

The training programs should focus on the cognitive and motivational factors we recognized, such as spending more effort in inspecting various aspects of the emails, allocating time evenly across emails (rather than selectively processing emails), and overcoming biases caused by perceived familiarity with the business entities and by dispositional optimism. Performance feedback about one’s judgmental confidence and accuracy may also be necessary to help decrease overconfidence.

We emphasize the importance of formal training in helping people fight against phishing attacks because our study shows that prior victimization of identity theft does not automatically translate to improved awareness of the risks or readiness to detect phishing. The impact of task easiness on overconfidence then suggests that the difficulty of training programs is important to produce the best training outcomes. The fact is that easy tasks tend to reduce overconfidence and hard tasks tend to increase overconfidence, which implies that, if the training program is too simple (e.g., introducing basic concepts of phishing without challenging the participants’ actual abilities), the participants may become overconfident when processing real-world phishing emails that are more tricky and difficult. On the other hand, if individuals receive tough training, they will exhibit less overconfidence when dealing with real-world emails because those emails would be relatively easier to detect. Therefore, the training program should comprise challenging tasks but that are not too challenging to destruct one’s self-efficacy beliefs because our study also shows that, although self-efficacy does not help bring down overconfidence, it improves both judgmental confidence and accuracy, which is valuable to phishing detection. Finally, our study suggests that, due to gender difference in overconfidence, male users should receive adequate training to combat phishing.

## 6.2 Limitations and Future Studies

Our study has several limitations. First, the experimental setting may not fully capture a person’s responses to phishing attacks in the real word because other factors, such as one’s existing email load at the time of attack, may have an impact on judgment (Vishwanath et al. 2011). As we mention above, the experimental setting is the most efficient for collecting large samples of data in phishing research. In addition, behavioral decision making studies have also studied judgmental confidence almost exclusively with lab experiments (e.g., Hong et al., 2014; Palmer et al., 2013; Tang et al., 2014; Weber & Brewer, 2004). If possible, though, one may employ mock attacks to test individuals’ actual responses to phishing emails, but one should carry out such methods with caution.

Second, our using images (instead of presenting emails via an email client) may have limited individuals' ability to correctly judge the emails due to the loss of information cues and their lack of interactivity with the emails. Tang et al. (2014) show that visualization and interactivity together reduce overconfidence in judgment. Future research could address this limitation by presenting emails via an email client (but not pictures of them). Nevertheless, to minimize the impact on judgments, we carefully selected emails that contained various detection cues, such as mouse-over URLs at the bottom, clear message content and email layout, and senders' email addresses. These cues are the typical information cues people use to detect phishing emails.

Third, we followed other studies in phishing literature (Downs et al., 2006; 2007; Pattinson et al., 2012; Sheng et al., 2010; Vishwanath et al., 2011; Wang et al., 2012) to explicitly ask participants to determine whether an email image was legitimate or not. While we assured the participants in our consent form we anonymously collected data and were interested only in their candid thoughts and opinions, such a priming may affect participants' choices and lead them to misclassify a business email as a phishing email (Pattinson et al., 2012).

The fact that we selected five antecedents can perhaps be considered to be a limitation. As we mention above, we selected the five antecedents because prior studies have recognized these cognitive and motivational factors. Our efforts present a starting point to empirically testing retrospective overconfidence in phishing detection. Future research may include other factors. For example, one could further investigate how anti-phishing tools, such as visual email authentication services (Herath et al., 2012; Wang, Chen, Herath, & Rao, 2009) or other detection tools (Dinev & Hu, 2007; Zahedi, Abbasi, & Chen, 2015), and education programs may change one's judgment process and, thereby, one's overconfidence in detecting phishing.

## Acknowledgments

We thank Senior Editor Mikko Siponen and the anonymous referees for their critical comments that have greatly improved the paper. This research was supported by the National Science Foundation under grants SES-1227353 and in part by SES-1420758 and SES-1419856. The third author also thanks AT&T Inc. endowment at UTSA for support. The usual disclaimer applies.

## References

Alba, J. W., & Hutchinson, J. W. (2000). Knowledge calibration: What consumers know and what they think they know. *Journal of Consumer Research, 27*(2), 123-156.

Anandpara, V., Dingman, A., Jakobsson, M., & Liu, D. (2007). Phishing IQ tests measure fear, not ability. In *Proceedings of the 11th International Conference on Financial cryptography* (pp. 362-366).

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A., & Jourden, F. J. (1991). Self-regulatory mechanisms governing the impact of social comparison on complex decision making. *Journal of Personality and Social Psychology*, *60*, 941-951.

Berger, I. E. (1992). The nature of attitude accessibility and attitude confidence: A triangulated experiment. *Journal of Consumer Psychology*, *1*(2), 103-123.

Berger, I. E., & Mitchell, A. A. (1989). The effect of advertising on attitude accessibility, attitude confidence and the attitude-behavior relationship. *Journal of Consumer Research, 16*, 269-279.

Berner, E. S., & Graber, M. L. (2008). Overconfidence as a cause of diagnostic error in medicine. *The American Journal of Medicine, 121*(5), S2–S23.

Bettman, J. J., Johnson, E. J., & Payne, J. W. (1990). A componential analysis of cognitive effort in choice. *Organizational Behavior and Human Decision Processes, 45*(1)*,* 111-139.

Bhandari, G., & Hassanein, K. (2012). An agent-based debiasing framework for investment decision-support systems. *Behaviour & Information Technology, 31*(5), 495-507.

Bjorkman, M. (1992). Knowledge, calibration, and resolution: A linear model. *Organizational Behavior and Human Decision Processes, 51*, 1-21.

Bjorkman, M. (1994). Internal cue theory: Calibration and resolution of confidence in general knowledge. *Organizational Behavior and Human Decision Processes, 58*, 386-405.

Blavatskyy, P. R. (2008). *Betting on own knowledge: Experimental test of overconfidence*. *Journal of Risk and Uncertainty*, *38*(1), 39-49.

Brenner, L. A., Koehler, D. J., Liberman, V., & Tversky, A. (1996). Overconfidence in probability and frequency judgments: A critical examination. *Organizational Behavior and Human Decision Processes, 65*(3), 212-219.

Brucks, M. (1985). The effects of product class knowledge on information search behavior. *Journal of Consumer Research, 12*(1), 1-16.

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review*, *7*, 26-48.

Camerer, C., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *The American Economic Review, 89*(1), 306-318.

Cramer, R. J., Neal, T. M. S., & Brodsky, S. L. (2009). Self-efficacy and confidence: theoretical distinctions and implications for trial consultation. *Consulting Psychology Journal: Practice and Research*, *61*(4), 319-334.

Chen, R., Wang, J., Herath, T., & Rao, H. R. (2011). An investigation of email processing from a risky decision making perspective. *Decision Support Systems, 52*(1), 73-81.

Chuang, W.-I., & Lee, B.-S. (2006). An empirical evaluation of the overconfidence hypothesis. *Journal of Banking & Finance, 30*(9), 2489-2515.

Creyer, E. H., Bettman, J. R., & Payne, J. W. (1990). The impact of accuracy and effort feedback and goals on adaptive decision behavior. *Journal of Behavioral Decision Making, 3*, 1-16.

Dhamija, R., Tygar, J. D., & Hearst, M. (2006). Why phishing works. In *Proceedings of the Conference on Human Factors in Computing Systems.*

Diener, E., Sub, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin, 125*(2), 276-302.

Dinev, T., & Hu, Q. (2007). The centrality of awareness in the formation of user behavioral intention toward protective information technologies. *Journal of the Association for Information Systems, 8*(7), 386-408.

Downs, J. S., Holbrook, M. B., & Cranor, L. F. (2006). Decision strategies and susceptibility to phishing. In *Proceedings of the 2nd Symposium on Usable Privacy and Security* (pp. 79-90). Pittsburg, PA: ACM Press.

Downs, J. S., Holbrook, M., & Cranor, L. F. (2007). Behavioral response to phishing risk. In *Proceedings of the Anti-phishing Working Groups 2nd Annual eCrime Researchers Summit* (pp. 37-44).

Fazio, R., & Zanna, M. P. (1978). On the predictive validity of attitude: The role of direct experience and confidence. *Journal of Personality*, *46*(2), 228-243.

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422-444). New York: Cambridge University Press.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 552-564.

Furnell, S. (2007). Phishing: Can we spot the signs? *Computer Fraud & Security, 2007*(3), 10-15.

Garbarino, E. C., & Edell, J. A. (1997). Cognitive effort, affect, and choice. *Journal of Consumer Research, 24*(2), 147-158.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, *46*(2), 107-119.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*(4), 506-528.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases*. Cambridge: Cambridge University Press

Glaser, M., Nöth, M., & Weber, M. (2004). Behavioral finance. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 527-546). Hoboken, NJ: Wiley-Blackwell.

Graber, M. L., Franklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of Internal Medicine, 165(*13), 1493-1499.

Griffin, D., & Brenner, L. (2004). Perspectives on probability judgment calibration. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 177-199). Hoboken, NJ: Wiley-Blackwell.

Hayes, T., Tanner, M., & Schmidt, G. (2012). Computer security threats: Small business professionals' confidence in their knowledge of common computer threats. *Advances in Business Research*, *3*(1), 107-112.

Hayward, M. L. A., Shepherd, D. A., & Griffin, D. (2006). A hubris theory of entrepreneurship. *Management Science, 52*(2), 160-172.

Heath, C., & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty, 4*(1), 5-28.

Herath, T., Chen, R., Wang, J., Banjara, K., Wilbur, J., & Rao, H. R. (2012). Security services as coping mechanisms: An investigation into user intention to adopt an email authentication service. *Information Systems Journal, 24*(1), 61-84.

Hirshleifer, D., & Luo, G. Y. (2001). On the survival of overconfident traders in a competitive securities market. *Journal of Financial Markets, 4*, 73-84.

Hmieleski, K. M., & Baron, R. A. (2009). Entrepreneurs' optimism and new venture performance: A social cognitive perspective. *Academy of Management Journal, 52*(3), 473-488.

Ho, S. Y., & Bodoff, D. (2014). The effects of Web personalization on user attitude and behavior: An integration of the elaboration likelihood model and consumer search theory. *MIS Quarterly*, *38*(2), 497-520.

Hong, J. (2012). The state of phishing attacks. *Communications of the ACM, 55*(1), 74-81.

Hong, K. W., Kelley, C. M., Tembe, R., Murphy-Hill, E., & Mayhorn, C. B. (2013). Keeping up with the Joneses: Assessing phishing susceptibility in an email task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting September, 57*(1), 1012-1016.

Johnson, E. J., & Payne, J. W. (1985). Effort and accuracy in choice. *Management Science, 31*(4), 395-414.

Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. *Psychological Review, 104*(2), 344-366.

Kassin, S. M., Rigby, S., & Castillo, S. R. (1991). The accuracy-confidence correlation in eyewitness testimony: Limits and extensions of the retrospective self-awareness effect. *Journal of Personality and Social Psychology, 61*(5), 698-707.

Keren, G. (1991). Calibration and probability judgements: Conceptual and methodological issues. *Acta Psychologica, 77*(3), 217-273.

Keren, G. (1997). On the calibration of probability judgments: Some critical comments and alternative perspectives. *Journal of Behavioral Decision Making, 10*(3), 269-278.

Klayman, J., Soll, J. B., & González-Vallejo, C. (1999). Overconfidence: It depends on how, what, and whom you ask. Organization Behavior and Human Decision Processes, *79*(3), 216-247.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 107-118.

Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L. F., & Hong, J. (2007). Protecting people from phishing: The design and evaluation of an embedded training email system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 905-914).

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2010). Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology, 10*(2), 1-31.

Larrick, R. P. (2004). Debiasing. In D. J. Koehler & N. Harvey (eds.), *Blackwell handbook of judgment and decision making* (pp. 316-337). Hoboken, NJ: Wiley-Blackwell.

Larrick, R. P., Burson, K. A., & Soll, J. B. (2007). Social comparison and confidence: When thinking you're better than average predicts overconfidence (and when it does not). *Organizational Behavior and Human Decision Processes, 102*, 76-94.

Li, Y. (2013). The impact of disposition to privacy, website reputation and website familiarity on information privacy concerns. *Decision Support Systems, 57,* 343-354.

Libby, R., & Rennekamp, K. (2011). Self-serving attribution bias, overconfidence, and the issuance of management forecasts. *Journal of Accounting Research, 50*(1), 197-231.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20*(2), 159-183.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases* (pp. 306-334). Cambridge, UK: Cambridge University Press.

Lovallo, D., & Kahneman, D. (2003). Delusions of success: How optimism undermines executives' decisions. *Harvard Business Review, 81*(7), 56-63.

McClelland, A., & Bolger, F. (1994). The calibration of subjective probability: Theories and models 1980–94. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 453-482). New York: Wiley.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502-517.

Moores, T. T., & Chang, J. C. J. (2009). Self-efficacy, overconfidence, and the negative effect on subsequent performance: A field study. *Information & Management, 46*(2), 69-76.

Palmer, M. A., Brewer, N., & Horry, R. (2013). Understanding gender bias in face recognition: Effects of divided attention at encoding. *Acta Psychologica*, *142*(3), 362-369.

Pattinson, M., Jerram, C., Parsons, K., McCormac, A., & Butavicius, M. (2012). Why do some people manage phishing e-mails better than others? *Information Management & Computer Security, 20*(1), 18-28.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance, 16*(2), 366-387.

Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin, 92*(2), 382-402.

Payne, J. W., Bettman, J. R., & Luce, M. F. (1996). When time is money: Decision behavior under opportunity-cost time pressure. *Organizational Behavior and Human Decision Processes, 66*(2), 131-152.

Peterson, C., & Bossio, M. L. (2001). Optimism and physical well-being. In E. C. Chang (Ed.), *Optimism & pessimism: Implications for theory, research, and practice* (pp. 127-145). Washington, DC: American Psychological Association.

Peterson, D. K., & Pitz, G. F. (1988). Confidence, uncertainty, and the use of information. *Journal of Experimental Psychology: Learning: Memory, and Cognition, 14*(1), 85-92.

Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly, 31*(4), 623-656.

Pulford, B. D. (1996). *Overconfidence in human judgement (*doctoral thesis). University of Leicester.

RSA. (2012). *Phishing and the social world.* Retrieved from http://www.emc.com/collateral/fraud-report/online-fraud-report-1012.pdf

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology, 67*(6), 1063-1078.

Sharp, G. L., Cutler, B. L. & Penrod, S. D. (1988). Performance feedback improves the resolution of confidence judgments. *Organizational Behavior and Human Decision Processes, 42*, 271-283.

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J. (2010). Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the 28th International Conference on Human factors in Computing Systems* (pp. 1-10).

Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., & Cranor, L. F. (2007). Anti-phishing Phil: The design and evaluation of a game that teaches people not to fall for phish. *Carnegie Mellon University.* Retrieved from http://repository.cmu.edu/isr/22

Siegel-Jacobs, K., & Yates, J. F. (1996). Effects of procedural and outcome accountability on judgment quality. *Organizational Behavior and Human Decision Processes, 65*(1), 1-17.

Stone, D. N. (1994). Overconfidence in initial self-efficacy judgments: Effects on decision processes and performance. *Organizational Behavior and Human Decision Processes, 59*(3), 452-474.

Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review*, *12*(4), 437-475.

Stone, D. N., & Schkade, D. A. (1994). Effects of attribute scales on process and performance in multiattribute choice. *Organizational Behavior and Human Decision Processes, 52*(2*)*, 261-287.

Symantec. (2014). *Internet security threat report 2014*. Retrieved from http://www.symantec.com/content/en/us/enterprise/other_resources/b-istr_main_report_v19_21291018.en-us.pdf

Tang, F., Hess, T. J., Valacich, J. S., & Sweeney, J. T. (2014). The effects of visualization and interactivity on calibration in financial decision-making. *Behavioral Research in Accounting, 26*(1), 25-58.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A Social psychological perspective on mental health. *Psychological Bulletin*, *103*(2), 193-210.

Tetlock, P. E., & Kim, J. I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology, 52*(4), 700-709.

Tsai, C. I., Klayman, J., & Hastie, R. (2008). Effects of amount of information on judgment accuracy and confidence. *Organizational Behavior and Human Decision Processes*, *107*(2), 97-105.

Vance, A., Anderson, B. B., Kirwan, C. B., & Eargle, D. (2014). Using measures of risk perception to predict information security behavior: Insights from electroencephalography (EEG). *Journal of the Association for Information Systems*, *15*(10), 679-722.

Vancouver, J. B., Thompson, C. M., Tischner, E. C., & Putka, D. J. (2002). Two studies examining the negative effect of self-efficacy on performance. *Journal of Applied Psychology, 87*(3), 506-516.

Vishwanath, A., Herath, T., Chen, R., Wang, J., & Rao, H. R. (2011). Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems, 51*(3), 576-586.

Wang, J., Chen, R., Herath, T., & Rao, H. R. (2009). Visual e-mail authentication and identification services: An investigation of the effects on e-mail use. *Decision Support Systems, 48*, 92-102.

Wang, J., Herath, T., Chen, R., Vishwanath, A., & Rao, H. R. (2012). Phishing susceptibility: An investigation into the processing of a targeted spear phishing email. *IEEE Transactions on Professional Communication, 55*(4), 345-362.

Weber, N., & Brewer, N. (2004). Confidence–accuracy calibration in absolute and relative face recognition judgments. *Journal of Experimental Psychology: Applied,* 10(3) 156-172.

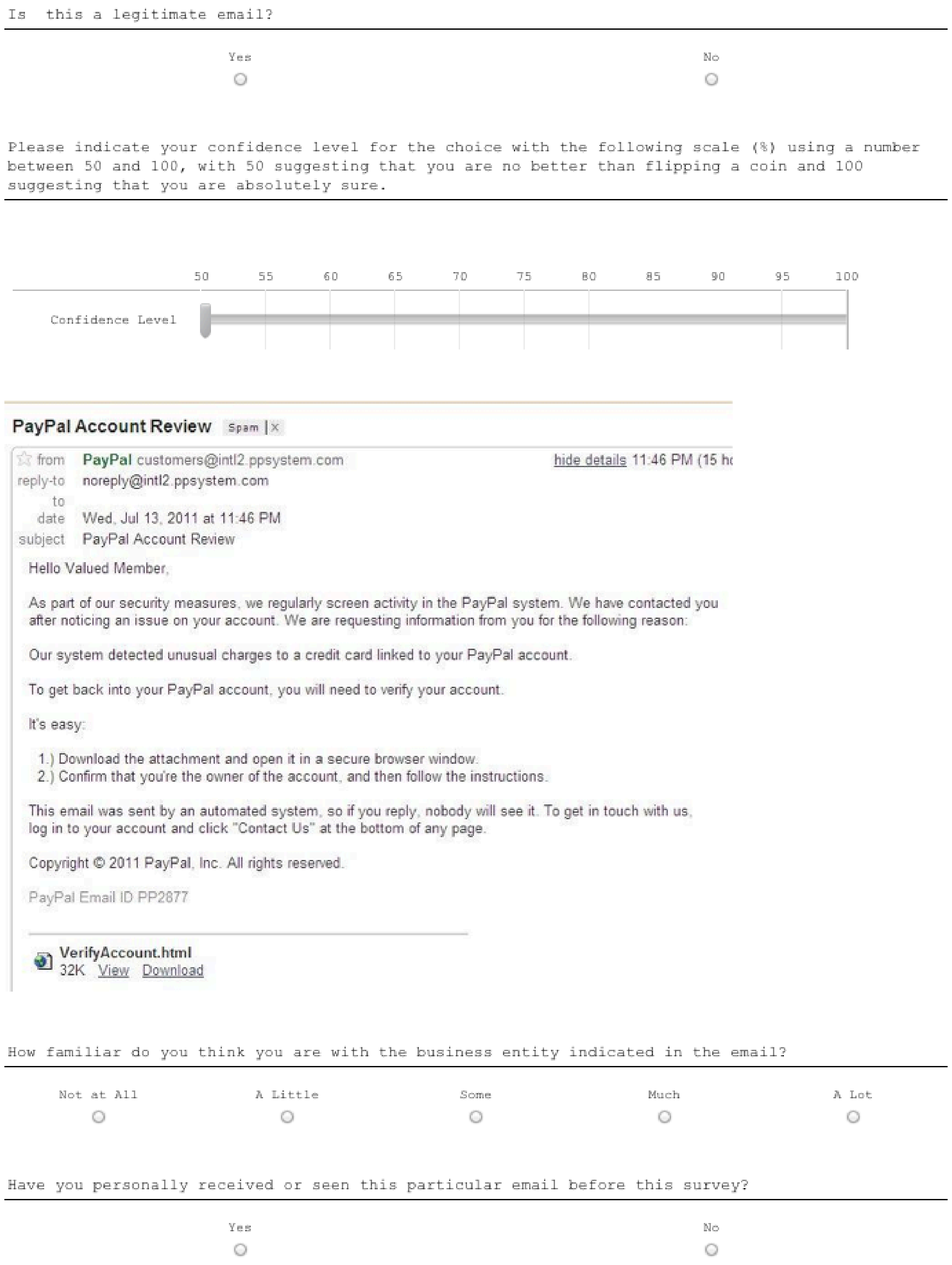
Wright, R. T., & Marett, K. (2010). The influence of experiential and dispositional factors in phishing: An empirical investigation of the deceived. *Journal of Management Information Systems, 27*(1), 273-303.

Yates, J. F. (1982). External correspondence: Decomposition of the mean probability score. *Organizational Behavior and Human Performance*, *30*, 132-156.

Zahedi, F. M., Abbasi, A., & Chen, Y. (2015). Fake-website detection tools: Identifying elements that promote individuals' use and enhance their performance. *Journal of the Association for Information Systems, 16*(6), 448-484.

# Appendix: Instrument Design

Following prior studies on phishing email detection (Downs et al., 2006, 2007; Pattinson et al., 2012; Sheng et al., 2010; Vishwanath et al., 2011; Wang et al., 2012), we used images of phishing emails and genuine business emails in the survey experiment. We randomly selected the emails from a pool of 50 images and presented them in a random order to avoid confounding with the email presentation order and learning effects. Half of the fifty email images were legitimate business emails sent by banks or financial institutions in the US and half were phishing emails targeted at customers of banks or financial institutions in the US. Customers of banks and financial institutions have been traditionally heavily targeted by identity thieves for their personal information, bank account, credit card number, and online banking login information in the past several years (Hong, 2012; RSA, 2012). We collected these emails from the Internet and banking colleagues' and our email inboxes. When we included the emails from the public domain in our survey, their text needed to be clearly legible (because we made most of the emails available in the form of image files), together with the sender, the receiver, the time sent, and title exactly as shown in the email that reached one's inbox. Different emails present varying informational cues and content. We changed the receiver's email address to a fictional email address if an email originally had it properly displayed and set the name to a fictional name if someone's name appeared in the original email. In the consent form of the survey, we stated that all these emails were actual ones that had arrived in someone's mailbox known to the researchers. Figure A1 illustrates an example.



Is this a legitimate email?

Yes ○ No ○

Please indicate your confidence level for the choice with the following scale (%) using a number between 50 and 100, with 50 suggesting that you are no better than flipping a coin and 100 suggesting that you are absolutely sure.

**PayPal Account Review** Spam | ×

from PayPal customers@intl2.ppsystem.com hide details 11:46 PM (15 h
reply-to noreply@intl2.ppsystem.com
to
date Wed, Jul 13, 2011 at 11:46 PM
subject PayPal Account Review

Hello Valued Member,

As part of our security measures, we regularly screen activity in the PayPal system. We have contacted you after noticing an issue on your account. We are requesting information from you for the following reason:

Our system detected unusual charges to a credit card linked to your PayPal account.

To get back into your PayPal account, you will need to verify your account.

It's easy:

1.) Download the attachment and open it in a secure browser window.
2.) Confirm that you're the owner of the account, and then follow the instructions.

This email was sent by an automated system, so if you reply, nobody will see it. To get in touch with us, log in to your account and click "Contact Us" at the bottom of any page.

Copyright © 2011 PayPal, Inc. All rights reserved.

PayPal Email ID PP2877

VerifyAccount.html
32K View Download

How familiar do you think you are with the business entity indicated in the email?

Not at All ○ A Little ○ Some ○ Much ○ A Lot ○

Have you personally received or seen this particular email before this survey?

Yes ○ No ○

**Figure A1. Email Judgment in the Survey**

## About the Authors

**Jingguo Wang** is an associate professor of information systems at the University of Texas at Arlington. He received his PhD in Management Science and Systems from the State University of New York at Buffalo. His current research interests are in the areas of cybercrime and information security, information search, and decision-making. His work has been published in *MIS Quarterly, Information Systems Research, Journal of Management Information Systems, the Journal of the Association for Information Systems, IEEE Transactions on Systems, Man and Cybernetics (Part C), European Journal of Operational Research, Decision Support Systems,* among others. His research has been funded by National Science Foundation and the University of Texas at Arlington.

**Yuan Li** is an associate professor of business in the Division of Business, Mathematics and Sciences at the Columbia College in Columbia, South Carolina, US. He received his PhD in Management Information Systems from the University of South Carolina. His research focuses on information and knowledge management, end-user computing, and online information privacy and security. His research appears in *the Journal of the Association for Information Systems, European Journal of Information Systems, Decision Support Systems, the Journal of Organizational and End User Computing, and the Communications of the Association for Information Systems*, among others.

**H. Raghav Rao** is AT&T Distinguished Chair of ISCS at University of Texas at San Antonio. Earlier, he was SUNY Distinguished Service Professor at UB. He is the co Editor-in-Chief of *Information Systems Frontiers*, SE of *MISQ*, Advisory Editor of *DSS*, and AE of *ACM Transactions on MIS*. He has published over 150 archival articles and is a graduate of Purdue University and the FBI citizens' academy.