Decision Strategies and Susceptibility to Phishing

Julie S. Downs
Carnegie Mellon University
Social & Decision Sciences
Pittsburgh, PA 15213
1-412-268-1862
downs@cmu.edu

Mandy B. Holbrook
Carnegie Mellon University
Social & Decision Sciences
Pittsburgh, PA 15213
1-412-268-3249
holbrook@andrew.cmu.edu

Lorrie Faith Cranor
Carnegie Mellon University
Computer Science & EPP
Pittsburgh, PA 15213
1-412-268-7534
lorrie@cs.cmu.edu

ABSTRACT
Phishing emails are semantic attacks that con people into divulging sensitive information using techniques to make the user believe that information is being requested by a legitimate source. In order to develop tools that will be effective in combating these schemes, we first must know how and why people fall for them. This study reports preliminary analysis of interviews with 20 non-expert computer users to reveal their strategies and understand their decisions when encountering possibly suspicious emails. One of the reasons that people may be vulnerable to phishing schemes is that awareness of the risks is not linked to perceived vulnerability or to useful strategies in identifying phishing emails. Rather, our data suggest that people can manage the risks that they are most familiar with, but don't appear to extrapolate to be wary of unfamiliar risks. We explore several strategies that people use, with varying degrees of success, in evaluating emails and in making sense of warnings offered by browsers attempting to help users navigate the web.

Categories and Subject Descriptors
J.4 [**Social and Behavioral Sciences**]: *Psychology;* H.1.2 [**User/Machine Systems**]: *Software psychology;* K.4.4 [**Electronic Commerce**]: *Security*

General Terms
Security, Human Factors

Keywords
Phishing; qualitative methods; mental models

1. INTRODUCTION
Phishing attacks, in which victims get conned by spoofed emails and fraudulent web sites, pose a growing problem for both Internet users and for the companies whose brands are spoofed. Victims perceive that phishing emails are associated with a trusted brand, but in reality they are the work of con artists. These increasingly sophisticated attacks not only spoof email and web sites, but they can also spoof parts of a user's web browser, for example, to hide warnings and URL information. Attackers aim to capture users' passwords, bank account information, credit card numbers, or other personal information, or to con users into sending them money or unwittingly assisting them in carrying out a scam. A more recent form of phishing attack known as *spear phishing* involves personalized emails or emails sent to a specifically targeted group, such as employees of a particular organization [1]. These attacks can be more effective than non-personalized phishing attacks and particularly difficult for anti-phishing tools to catch [2].

According to the Anti-Phishing Working Group, the number of new phishing web sites increased dramatically in 2005, with 7197 new phishing sites detected in December 2005, up from the previous high of 5259 new phishing sites detected in August 2005 and 1707 phishing sites detected in December 2004 [3]. Direct phishing-related losses to US financial institutions are estimated at over a billion dollars per year [4].

Computer security attacks can be classified as *physical*, *syntactic*, or *semantic*. Physical attacks target the physical infrastructure of computer systems and networks, while syntactic attacks target software. Semantic attacks, on the other hand, are aimed at people. Rather than taking advantage of system vulnerabilities, semantic attacks take advantage of the way humans interact with computers or interpret messages. Phishing attacks are examples of semantic attacks. Much research has gone into mitigating syntactic attacks, as well as adapting approaches for combating syntactic attacks into similar approaches to combat semantic attacks — for example, developing filters that can detect the signature of a phishing email. However, much less research has been done to try to systematically understand and address the human side of semantic attacks. As Bruce Schneier put it, solutions in this area need "to target the people problem, not the math problem" [5].

Phishing attacks are successful when attackers are able to manipulate users into "forming inaccurate mental model[s] of an online interaction" [6]. Thus it is important to understand the types of mental models people use when reading email and browsing the web, and the degree to which they are susceptible to manipulation. Before we can address the "people" side of phishing attacks, we must develop a better understanding of why people fall for these attacks and the extent to which people take advantage of available cues that might help them identify fraudulent emails and web sites.

1.1 Anti-phishing tools
Much effort has gone into the development of automated tools for detecting phishing attacks. Anti-phishing services and tools are now provided by Internet service providers, built into mail servers and clients, and available as web browser "tool bars." However,

these services and tools do not effectively protect against all phishing attacks, as attackers and tool developers are engaged in a continuous arms race.[1] Furthermore, Internet users who are unaware of the phishing threat will be unlikely to install and use an anti-phishing tool, and may ignore warnings from anti-phishing tools provided by their ISPs.

Internet users who are aware of the phishing threat can now choose from over a dozen free anti-phishing toolbars that they can download and install into their web browsers. Some of these toolbars employ heuristics for identifying fraudulent web sites [7]. Some toolbars connect periodically (or with each web request) to a server that maintains a blacklist of known phishing URLs. Others maintain lists of web sites that a particular user has visited and assist the user in detecting sites that are similar to these sites and might be spoofs of a legitimate site.

Some research has focused on the development of tools to help users determine when they are interacting with a trusted site. Ye et al. [8] and Dhamija and Tygar [9] have developed prototype "trusted paths" for the Mozilla web browser that are designed to assist users in verifying that their browser has made a secure connection to a trusted site. Herzberg and Gbara have developed TrustBar, a browser add-on that uses logos and warnings to help users distinguish trusted and untrusted web sites [10]. While these tools may assist an alert and informed user in identifying spoofed web sites, they still require a degree of diligence on the part of the user to distinguish between the indicators provided by the tool and spoofed indicators that an attacker might provide.

Wu et al. conducted a study of three simulated anti-phishing toolbars to determine how effective they were at preventing users from visiting web sites that the tools had determined to be fraudulent [11]. They found that many of their participants ignored the passive toolbar security indicators and instead used the site's content to decide whether or not it was a scam. In some cases participants did not notice warning signals, and in other cases they noticed them but assumed the warnings were invalid. In a follow-up study the authors tested anti-phishing toolbars that produced pop-up warnings that blocked access to fraudulent websites until overridden by the user. These pop-up warnings reduced the rate at which users fell for fraudulent sites, but did not completely prevent all users from falling for these sites. The authors concluded that Internet users are not very good at interpreting security warnings and are unfamiliar with common phishing attacks and recommended online safety practices.

1.2 Cues

Expert users rely on a variety of cues to avoid falling for phishing attacks. These cues may be found within the headers or content of phishing email messages or within the content of fraudulent web sites. In addition, cues may be provided by browser-based security indicators and by anti-phishing tools.

A number of cues that an email may be fraudulent can be found within the email itself. Experts recommend that users treat with suspicion any email that asks them to follow a link to update account information, or threatens dire consequences for not immediately providing or updating personal information, such as closing an account. Messages from banks or other companies with which a user does not have a relationship should also be viewed with suspicion. Messages that claim to be from businesses but contain misspelled words or sloppy grammar are also suspicious. Likewise, business messages sent from a domain name other than the one usually used by that business are also suspicious, although sometimes businesses do outsource email services to third parties who use their own domain names. Experts recommend disabling the use of JavaScript in email clients and manually checking the URLs behind email hyperlinks (by mousing over them or viewing their raw source), or typing any hyperlinks directly into a web browser rather than clicking on them. URLs containing IP addresses or that appear similar to but not exactly the same as domains for well-known brands should be viewed with suspicion.

Once users arrive at a web site, there are additional cues they might use to determine whether it is fraudulent. They can check to see what URL they are visiting, again looking for IP addresses in the URL or addresses similar to popular domains. They should also look for misspelled domain names and subtle substitutions such as 0 for O or vv for w, although it may be unrealistic to expect even the most informed users to spot such subtle cues. They might also look for the presence of the SSL or TLS lock icon in the corner of the browser and click on it to verify that it has a certificate that matches the web site they believe they are visiting. Unfortunately, many users do not understand or have misconceptions about the meaning of the lock icon [12] and there are a number of ways that the lock icon can be spoofed or manipulated, so it is not a completely reliable indicator. In addition, many legitimate web sites do not use SSL/TLS except when transmitting form data, so the lock icon will not appear until after a user presses the submit button.

Anti-phishing tools provide additional indicators that a web site may be fraudulent, including warning icons, information about the country in which a domain has been registered and how long it has been in existence, and pop-up warnings.

Dhamija et al conducted a study in which they showed 22 participants 20 web sites and asked them to determine which were fraudulent. Participants made mistakes on this test set 40% of the time. The authors noted that 23% of their participants ignored all cues in the browser address bar and status bar as well as all security indicators [13]. This study did not present users with the email messages that might lead users to visit the web sites presented, so it provides no data on whether users pay attention to, or how they interpret, email cues.

1.3 Mental Models Approach

In this paper we report on a mental models interview study we conducted with 20 non-expert Internet users. We collected qualitative data on their awareness of phishing-related risks, sensitivity to phishing cues, and email decision strategies.

The theoretical approach to decision-making that is the basis of this research falls into the broad category of "mental models" approaches in cognitive psychology [14,15,16,17]. The strength of this approach comes from an extensive discourse with members of the intended audience, examining their understanding and underlying assumptions. Interventions based on the qualitative insights gained from mental models interviews provide

[1] Little information is currently available on the effectiveness of the available toolbars. A pilot study conducted by our research group found many of them to be quite disappointing. However, a more thorough examination is needed.

information in an accurate way and promote trust between audience and intervention [18,19].

These interviews are designed to gain understanding of how email users describe the decisions they make and the contexts in which they make them. The open-ended procedure helps to identify myths and misconceptions that are often missed by more structured surveys or user tests [20]. In this paper, we report a preliminary analysis of emerging themes, rather than a full mental models analysis. The 20 interviews discussed here are part of a larger mental models study involving 40 participants, including 35 users without experience in computer security, and five experts. A future report will discuss the results of the larger study.

2. METHODS

2.1 Recruiting methods

Twenty participants were recruited by posting paper flyers in various locations around Pittsburgh and posting online messages to Craigslist.org, a community bulletin board. These flyers and posts advertised an interview study at Carnegie Mellon, with no description of the topic. Respondents replied via email and received a reply email with a hyperlink that they could follow (or they could choose to reply via email), to ensure that they had some experience receiving and responding to email.

Those who replied to the advertisements were directed to a website with a set of demographic questions and three screening questions asking: 1) whether they had ever changed preferences or settings in their web browser, 2) whether they had ever created a web page, and 3) whether they had ever helped someone fix a computer problem. If they answered yes to any of these questions, they were asked to explain what they had done. Those whose answers involved security, such as changing security levels or reviewing cookies, were excluded from eligibility. Other requirements were being over 18 years old and having been online for at least one year. Thus, our sample may have a broad range of experience with computers, but were selected to have relatively little expertise in computer security.

2.2 Mental models interviews to gain insights

One-on-one interviews were conducted in person, and were audio taped to provide exact transcripts. At the outset, participants were told that the interview would be about "your computer use," and "how people make decisions while using their email and visiting web sites." The interview protocol had two segments: the *email and web role-play* section, in which participants read and responded to a set of emails, and the *security and trust decisions* section, in which participants talked about their own computer use, concepts relating to trust on the Internet, and their awareness of security measures online.

2.2.1 Email and web role play

Participants were given an identity to role play, complete with a wallet containing identification (without any picture), account information, login passwords for Amazon and PayPal written on the back of a business card, a Citibank credit card, a social security number, and a note with the name and phone number of a friend who would turn out to be one sender of email. Female participants were given a woman's wallet with identification for Patricia Jones, and male participants were given a man's wallet with identification for Patrick Jones; all other information referred to the identity as Pat Jones.

Participants used a PC or Mac laptop (their choice) to look at Pat's email, and were instructed to read and react to the messages as they normally would in their own life. A video camera used an over-the-shoulder perspective (focusing on the screen and keyboard) to visually record participants' actions.

Participants viewed eight emails in Pat's inbox (see Table 1). The first email was created to familiarize the participant with the procedure and set a very broad tone of interest to avoid raising early suspicions. It was a short message from the same domain name as Pat's email, which was the name of the company that Pat worked for according to the information in the wallet. This message, from the secretary of a manager, announced a time change for a meeting and requested an RSVP. Participants were instructed to treat this message however they normally would, whether that would be a telephone call, an email response, no reply, or any other action.

Table 1. Emails in Pat Jones' Inbox

Email	Legitimacy	Relevant features of email and sites
meeting	real	•regarding work details •no links in email
cool pic	real	•sender is known person •addressed to user •text of link: "this" •actual URL: antwrp.gsfc.nasa.gov/apod /astropix.html
Amazon	real	•web page doesn't ask for password •link: www.amazon.com/exec/obidos /sm/change/RF820KQA3VTJ •URL: same
Citibank	phishing	•urgent request •lock image in body of web page •link: www.citicard.com/verifyEmail •URL: www.citibank-accountonline.com /accountonline/AccountSummary.htm? verify=email
Great article	possible malware	•impersonal greeting •link: www.BestInsurance.com/SaveMoney.pdf •URL: 128.2.66.1/ws/SaveMoney.pdf.exe
PayPal	phishing	•ironic warning to protect password •broken image links •link: "Click here to activate your account" •URL: www.payaccount.me.uk/cgi-bin /wbscr.htm?cmd=_login-run
Amazon	phishing	•grammatical mistakes •link: www.amazon.com/exec/obidos/ sign-in.html •URL: www.amazonaccount.net/exec/ obidos/flex-sign-in.htm
Katrina	419 scam	•real CNN links •real company name (HSBC bank)

The next seven emails all contained links to web pages, progressing from legitimate to more and more obvious phishing schemes and scams. Participants were instructed to handle each in the same way: to do whatever they would normally do, and to look in Pat's wallet for more information if necessary. All web sites and email addresses were created specifically for this study, including domain names that we created for external web sites not hosted by legitimate companies.

For each email, participants talked aloud, describing what they were doing. They were probed briefly for their reasons for any action, such as deciding to delete an email or to click on a link. If they voiced suspicion about an email, it was neither confirmed nor denied by the interviewer, but merely probed for explanation, as with any other comment they made.

2.2.2 Security and trust decisions

The second segment of the interview protocol involved asking participants to describe their own online behaviors and their conception of what it meant for a website to be trustworthy. Participants were also prompted for awareness of specific cues about security. Table 2 shows sample questions, starting with general requests and moving toward more focused ones that ask respondents to elaborate on their thinking. This progression allows us to learn how computer users frame and think about issues of computer security, including myths and misconceptions, and the appropriate wording used to describe their beliefs. It also provides us with descriptive accounts of the different kinds of computer security threats subjects have experienced as well as how they would typically respond to them. Along with the open-ended questions, participants were asked to rate how bad five negative consequences of poor trust decisions would be: general consequences of complying with a suspicious email, having a credit card number stolen, having a bank account compromised, having a social security number stolen, and receiving a large increase of spam emails. These ratings were made on a scale ranging from 1 (*not bad at all*) to 7 (*major hassle or a disruption of one's life*). A few participants gave responses above the highest number of the scale, typically when they had previously given the maximum value but then wanted to indicate that a later question asked about an event that was even worse. This suggests that the labeling at the high end of the scale may have been too mild. To honor their intended relative ratings among questions, we transformed their answers to accommodate their responses within the 7-point range while preserving differences between responses.

Table 2. Sample Questions from Interview Protocol

Do you ever receive email messages from companies that you have an existing relationship with?
• Can you describe some emails messages you've gotten from these companies? [prompt: were they asking you to take some type of action?]
• What did you do when you received these email messages? [prompt: did you reply? did you follow the instructions?]
How can you tell when you can trust a web site?
• How does [what participant said] make it a trusted site?

Assessing awareness of security and trust cues was approached obliquely, to give participants as many chances as possible to volunteer what cues they look for before asking them outright about particular signals. First they were asked about their own computer-use behaviors. Second, they were shown four pop-up security messages. Third, they were directed to three websites, one legitimate and two phishing, and asked about cues for trustworthiness. For each, participants were asked to indicate whether they would look for anything on the screen to determine whether the site was trustworthy. If participants mentioned a security feature, they were asked what the feature meant about the trustworthiness or security of the site. Standard security features that were not volunteered during the interview were brought up explicitly at the end to ensure that nothing was missed in the participant's knowledge.

The four pop-up messages were shown as images alerting a user that: 1) information was to be sent over an unencrypted connection, 2) the user was leaving an encrypted page, 3) the user was about to visit an encrypted page, and 4) a certificate was signed by an unknown authority. Some of the pop-up images were not shown to a small number (10% or fewer for each) of participants due to occasional technical problems with the online images. The three websites that participants were asked to visit came from the emails that participants had already seen in Pat Jones' mailbox. They included a legitimate Amazon.com site and two phishing sites, spoofing Citibank and PayPal. The Citibank site was a relatively good spoof, including a lock image in the page (but was not https) and a reasonable-sounding .com URL. The PayPal site was a worse spoof, with broken images and a foreign URL (.uk).

2.3 Qualitative Content Coding

Interviews were transcribed verbatim. The transcripts were then coded according to several criteria. In the email and web role-play sections, behavioral responses were recorded for each of the email messages (along with their associated web sites, where appropriate), indicating whether the participant had refused to comply due to suspicion. In addition, codes were assigned for each cue that participants mentioned as reasons for trusting (or suspecting) websites. These codes did not include the explicit questions about cues in the later part of the interview, but were limited to those volunteered in the course of responding to email messages.

For the security and trust decisions segment of the interview, we coded most variables dichotomously to indicate whether the participant reported having experience with the concept in question. For example, participants would receive a score of 1 for the variable "ever shopped online" if they reported ever having shopped online. Similar variables were coded for other online behaviors, awareness of cues, fraud victimization, and use of security tools.

Several indices were calculated to combine multiple observations into single continuous variables, to aid analysis and interpretation: *phishing susceptibility* (total of the three phishing emails fallen for); *misplaced suspicions* (suspecting each of the two legitimate emails with links); *online activities* (e.g., online banking, up to four activities); *risky behaviors* (e.g., installing freeware, up to three behaviors, partially overlapping with the online activities); and *awareness of certificates* (having heard of certificates, and understanding what they are). In addition, the ratings for the four possible consequences of compromised information were averaged for an overall rating of consequences.

2.4 Statistical Analyses

Analyses were conducted using SPSS 11.0.4 for the Macintosh [21]. To sort the specific cues that people reported using during the email role play into smaller, meaningful combinations, a factor analysis was conducted. It sorted seven common signals that people mentioned in evaluating emails into three independent factors.

Analyses were performed on two types of variables: dichotomous (e.g., observations of whether a participant had mentioned a

concept or engaged in a behavior) and continuous (e.g., indices combining multiple dichotomous variables or 7-point ratings). All continuous variables were roughly normally distributed, so none required transformation. Pearson product-moment correlations examined relationships between continuous variables. Chi-square analyses examined the relationships between dichotomous variables. The Student t-test was used to compare groups distinguished by dichotomous variables on continuous outcome variables, and logistic regression was used to examine how continuous variables predicted dichotomous outcome variables, with odds ratios (ORs) reported. Repeated measures analysis of variance (ANOVA) was used to explore groups of participants across multiple, continuous variables.

3. RESULTS

3.1 Participants

A total of 56 respondents completed all the screening information necessary to determine whether they qualified for enrollment in the study by being sufficiently inexperienced in computer security. The most common reason for disqualification was adjusting security preferences on their computer, which 39% reported having done. An additional 14% were disqualified (25% of all respondents) for having helped another person with a computer problem involving security such as scanning for viruses. A further 11% did not respond to follow-ups to schedule an interview. Thus, only 36% of respondents qualified as members of a particularly security-naive subset of the general population and so were eligible for our study. Given this select sample, it is important to keep in mind that this study is designed to explore patterns of responses, such as relationships between perceptions of cues and decision strategies, rather than overall susceptibility to scams.

Participants ranged in age from 18 to 45 (mean = 27). Three quarters were female. Half of participants reported their race as white, 25% as African American, 15% as Asian and 10% other or declined to answer. One participant was a Mac computer user; the other 19 used PC computers. Participants reported Internet experience ranging from 6 to 16 years (mean = 10). Most (70%) used Internet Explorer as their primary browser; 20% used Firefox, and one each used Netscape and Safari. Most (80%) used a browser-based email client to read their email (e.g., Yahoo, Hotmail or GMail), 10% used Outlook and 10% used other programs. Participants reported receiving approximately 3-40 emails per day (mean = 15, median = 10). Participants reported a range of occupations, with 30% in the service industry, 30% students (mostly graduate students), 25% in professional positions, and 15% in administrative positions. About two-thirds (65%) had a college degree.

Respondents who did not qualify were similar to participants along many of these dimensions, including age range (18-55, mean = 25), education (63% with college degree), years on the Internet (6-14, mean = 10), use of a browser-based email client (80%), and emails per day (2-50, mean = 15). They appeared to differ slightly on some computer usage variables, such as being less likely to use Internet Explorer (43%) and more likely to use a Mac (13%). There were more students in this population (60%). They also tended to fall into different demographics, with more being male (67%) or Asian (37%) and fewer being African American (3%).

3.2 Experience Online

Nearly all participants (95%) reported having purchased something online at some point in their lives, and 70% had done online banking. All had entered correct information about themselves into a form on a web site at some point, and most (75%) also reported having entered incorrect information at some point. Typically they reported doing so when they felt that their personal information was not the business of the web site asking for it, e.g., "just to keep some anonymity if it wasn't, you know, crucial they know all that information." Or, as this participant describes:

> *But usually, I will only typically enter, not false but modified information if I feel that this website may not be as reliable as I feel it should be. Sometimes I will want to enter into some kind of free promotion and all that, and I may not give my full information or I may modify my information.*

Younger people reported having engaged in more online activities (correlation with age: r = -.49, $p<.05$), especially risky activities (r = -.65, $p<.01$), but age was not related to awareness of cues or risks, or behaviors in the email role play.

3.3 Awareness of Security Risks

Potential consequences of compromised information were rated quite negatively overall compared to merely receiving more spam (see Table 3). These consequences were rated as worse by people who engaged in more online activities (r = .44, $p=.05$: on average 4.9 for those who had only done one or two of the four online activities vs. 5.9 for those who had done 3 or 4 activities). But higher ratings of these consequences did not predict how people responded to the role-play emails.

Table 3. Ratings of Negative Consequences

Possible consequences	Rating
Following directions of suspicious email	5.9
Stolen credit card number	5.5
Bank account compromised	6.1
Social security number compromised	6.6
Large influx of spam	3.7

One-quarter of participants reported having been the victim of fraud (although not necessarily via the Internet), either through a stolen credit card number or social security number. (None reported having had their bank account compromised, although two answered by saying, "not yet," suggesting a possible sense of inevitability about this kind of fraud.) These individuals did not rate the consequences of such fraud any differently than others did (5.8 vs. 6.1, $t(18) = 0.58$, $p=.57$).

Although they do not seem to be overly alarmed about the hassle of having their information compromised, these prior fraud victims are not complacent: all participants rated the consequences as quite negative. One might wonder, though, whether they would be more wary of schemes aiming to defraud them again. In fact they were marginally *more* likely to fall for at least one of the phishing attacks in our role play (100% vs. 60%, $\chi^2 = 2.86$, $p=.09$). This gullibility might be what put them at risk

in the first place, but it does seem to be the case that past experience is no guarantee of preventing it from happening again.

Suspicion about email is certainly not lacking. Nearly all of our participants (95%) had heard the term 'spyware' before. Several believed incorrectly that it was something that protected their computer, but this still indicates awareness that spyware is a term related to computer security. In contrast, only about half of participants had heard the word 'phishing,' with others generally unable to guess what the term meant. A common guess, "something to do with the band Phish, I take it," was not helpful. Moreover, awareness of this term was unrelated to any other measure of awareness or behavior in the interview.

3.4 Sensitivity to Phishing Cues

Our participants reported having previously seen several cues that might alert a user to be suspicious, including:

- *spoofing "from" addresses (*95% of participants).

Nearly all participants reported having received emails with addresses in the sender line that were not the true sender. Participants reported noticing such things as generic names, additional letters in the address or a different domain extension, e.g., "I get some of these crazy advertisements, and it seems like they try to use these standard names like Jessica Jones or something. And I am like assuming that some girl named Jessica Jones didn't send that to me."

- *secure site lock icon* (85% of participants).

Most participants had seen lock images on a web site, and knew that this was meant to signify security, although most had only a limited understanding of what that meant or how to interpret locks, e.g., "I think that it means secured, it symbolizes some kind of security, somehow." Few knew that the lock icon in the chrome (i.e., in the browser's border rather than the page content) indicated that the web site was using encryption or that they could click on the lock to examine the certificate. Indeed, only 40% of those who were aware of the lock realized that the lock had to be within the chrome of the browser. Rather, they discussed the appearance of a lock anywhere on the page, e.g., "Basically it gives the consumer, the facade that it's secure. It might not be, but at least the consumer feels a little bit more safe on the website." Several mentioned explicitly that a lock in the page itself did not guarantee security, without indicating any awareness that a lock elsewhere had meaning, e.g., "No I am sure they just probably put it there. Maybe it is secure maybe it is not, I don't really think that it means anything."

- *broken images on web page* (80% of *participants*).

Most participants had noticed broken images, including red Xs, question marks or blank spaces where it looks like there should be an image. However, this was not immediately taken as a cue that a web site might be suspicious. Rather, participants thought that a problem with their own computer or Internet connection might be the cause. Some augmented this explanation with inferences about broken images as cues about professionalism, e.g., "It either means that my browser didn't load it right and something was screwed up on my end, or if I reload it and it's still that way, something is obviously amiss with the website itself and either they didn't code properly to make that image appear, which also makes you more worried about the website because whoever made it was, if they were inept enough to miss that then maybe they're, you know, maybe they're not someone that, whose website you should be visiting I suppose." Alternatively, some harbored no suspicions at all, e.g., "I think it means that image where it is housed is moved and the page has not been updated."

- *unexpected or strange URL* (55% of participants).

About half had noticed a URL that was not what they expected, or that looked strange. For some, this was a reason to be wary of the website, e.g., "I've been to one or two in this situation and I just close it off. I just thought something was wrong with the engine." For others, it was an annoyance, but no cause for suspicion, e.g., "It seems like a lot of times the things that frustrates me about websites is that you'll click on one thing, and they give you something completely different." Others expressed awareness but not necessarily suspicion, e.g., "If it wasn't one of my standard dot-com, dot-edu, dot-us or some country code, then I would be really curious what that meant." Since none of these participants made any note of some quite suspicious URLs in the role play, one can assume that the remaining 45% of participants may have paid no attention at all to URLs.

- *"https"* (35% of *participants*).

Only about a third of our participants had noticed this text indicating Hypertext Transfer Protocol over Secure Socket Layer or Transport Layer Security. Some who reported having noticed both "http" and "https" did not think that the "s" indicated anything. But those who were aware of the security connotation of this cue tended to take it as a fairly reliable indication that it is safe to enter information, e.g., "If I'm online and filling in things and you want to verify, then that's another thing because it has that https, and you know its a secure website and all that." This extra security was often enough to get people beyond their initial trepidations about sharing sensitive information, e.g., "I feel funny about putting my credit card number in, but they say it is a secure server and some of them say 'https' and someone said that it means it's a secure server."

Overall, awareness of these general security cues did not appear to translate into caution in interpreting that information in emails. The only one of the cues that related to behavior was the security lock icon, and its effect was not particularly helpful. Those who were aware of the lock were marginally more likely to suspect at least one of the legitimate emails, which did not use SSL/TLS, of being possibly fraudulent (47% vs. 0%, $\chi^2 = 3.59$, $p=.06$), but were just as likely to fall for at least one of the phishing emails (73% vs. 60%, $\chi^2 = 0.32$, $p=.57$).

Cues that could alert someone to particular scams appeared to be more helpful in a scam-specific way. Merely asking for sensitive financial information (such as bank account or social security numbers) was mentioned by 55% of participants as a red flag that an email should be examined carefully. Not surprisingly, those mentioning this cue were less likely to give their social security numbers in response to the phishing email allegedly from Citibank (9% vs. 50%, $\chi^2 = 4.00$ $p<.05$). However, suspicion about financial information did not translate into wariness about other sensitive information; these participants were actually *more* likely to give out their Amazon.com passwords in response to the phishing email allegedly from Amazon (73% vs. 25%, $\chi^2 = 4.23$ $p<.05$). Given the prevalence with which passwords are used by websites, it is perhaps not surprising that they are treated with less concern.

Suspicion about giving any information out over email might be expected to lead to more cautious behavior. Only three people were concerned by emails asking for a password, two of whom had also been suspicious of requests for financial information. A repeated measures ANOVA explored how these people, versus the remaining 17 participants, reacted with suspicion to the phishing and to the legitimate emails, and found an interaction between the groups of participants and type of email (the repeated measure). None of those who were concerned about giving out passwords gave their information on any of the phishing emails (whereas the others fell for an average of 1.3 of the three emails), but they also had more misplaced suspicions, on average suspecting 0.7 vs. 0.4 of the two genuine emails to be fake, $F(1,18) = 3.78$, p=.07. This group is too small to draw strong conclusions from and the interaction is only marginal, but it appears that their suspicions may have only shifted their response to be more avoidant, while not reflecting much better ability to discern genuine from fraudulent emails. Although this study was not designed to test the negative effects of broad suspicion, it is important to consider the risks of missing important communications.

3.5 Email Decision Strategies

In the email and web role play, people mentioned various signals that they used in deciding how to respond to emails, typically revolving around what they needed to do with the information provided, and whether to trust the email or to be suspicious of it. Although we did not specifically ask about the trustworthiness of the emails in explaining the role play to participants, it was clear that this was the main dimension along which decisions were being made.

Table 4. Suspicion about Legitimate and Phishing Emails

Email	Legitimacy	Percent expressing suspicion
meeting	real	0%
cool pic	real	15%
Amazon	real	25%
Citibank	phishing	74%
Great article	possible malware	85%
PayPal	phishing	70%
Amazon	phishing	47%
Katrina	419 scam	95%

Table 4 shows how many respondents were suspicious about each email. Everybody expressed suspicion about at least two of the eight emails, with an average of four emails being found suspicious and a maximum of seven (with this participant suspecting all except the initial email about the meeting time change). Not surprisingly, the legitimate emails were suspected of being fraudulent or malicious far less often than the illegitimate ones, although 35% of participants expressed some suspicion about at least one of the legitimate emails. In rare cases, a respondent was too vague to determine whether they were suspicious. It is important to note that the particular emails used in this study are not necessarily representative of phishing emails in general, but were based on the range of emails received by colleagues. They were chosen to include various kinds of cues and deceptions, which allow us to explore the relationships between cues and strategies.

Three factors emerged from the factor analysis as general strategies that people use in describing their responses to the role-play emails: 1) this email appears to be for me (e.g., if it is personalized, the grammar is good and the sender is known, versus, to some extent, a familiar email that has been seen often), 2) it is normal to hear from companies that you do business with (e.g., it's OK as long as one has an account with a company or, to some extent, if an email is familiar, versus being suspicious of unexpected emails), and 3) reputable companies will send emails (e.g., email is OK if it's from a reputable company and if it looks familiar). Factor scores were calculated for each participant to represent the degree to which their reports of cues followed each strategy.

3.5.1 Strategy 1: This email appears to be for me

The first strategy was strongly correlated with awareness of certificates ($r = .67$, p<.01), suggesting that this may be one way that people respond to the idea of scams. However, it was generally unrelated to the ability to detect (and avoid) phishing schemes in the role play ($r = -.23$, p=.33, accounting for only 5% of the variance).

Those who had not been online as many years were especially likely to be suspicious of emails that were not personalized (OR = 0.43, p<.05), reasoning that, "Its just like, they just don't have anything to do with me. So, I don't want to have anything to do with them." Participants using this strategy tended to judge emails by their face credibility, such as whether it's a mass email e.g.,

> *I probably wouldn't even have opened them unless it had, unless it had something more convincing that it was from a human than "great article."*

Others judged the professionalism, e.g., "I mean, most franchises don't have misspelling come out."

3.5.2 Strategy 2: It's normal to hear from companies you do business with

In contrast, the second strategy, which was unrelated to any measure of online behavior or demographic, was highly predictive of *falling for* more phishing emails ($r = .69$, p<.01). These participants paid attention to the likelihood that an email was really for them, using criteria such as whether they had an account with the business in question, e.g., "I would respond to this if I really had an account there [Laugh], but if not, I would delete it." This purposeful strategy might make these people particularly vulnerable: phishing schemes depend on reaching the minority of people who have accounts with the spoofed brand, and these respondents' belief that they are being careful may lead them to have unwarranted confidence in their screening abilities, and thus to let their guard down.

> *I've used PayPal before, umm, I click on the email and like I see what they're, and you know, if I had an email address and a password, then, anyone I had an account with, then I… [Interviewer: Then you would log in?] Yeah.*

In contrast were those who didn't regard having an account as

informative, but rather used context-specific criteria in making their decisions, such as whether an email was unexpected, e.g.,

Um, I would probably just trash it, maybe not necessarily junk mail because I do have an account with Amazon, but it wouldn't be to my work mail it would be to my home mail.

These individuals were unlikely to accept a generic email from a company they had an account with as sufficient cause for responding. Rather, they used offline information to make their decisions, e.g.,

If I'm not buying anything, right now. There's no reason for you to call me and ask me what kind of update do I need.

They appeared to take unexpected email itself as a cue, e.g.,

In fact I don't recall having companies sending me emails asking me to update my personal information. The point is, how do they know the information is inaccurate?

3.5.3 Strategy 3: Reputable companies will send emails

The third strategy was only weakly related to experience online, specifically to receiving fewer emails, although this relationship was only marginally statistically significant ($r = -.38$, $p<.10$). This strategy did not predict overall susceptibility to phishing, but did marginally predict whether the participant opened the .exe file that was posing as a .pdf of an article describing car insurance (OR = 3.47, $p=.09$). These users may be particularly naive in their interpretation of email. They didn't appear to have high suspicions, nor did they have high confidence in their ability to avoid problems. Although this naive strategy did not help them spot phishing scams, nor did it leave them more vulnerable than others.

Ok, all right, they [Amazon] are reputable sales thingy. I am going to delete it, because I don't need to make any more changes. [Interviewer: Is that what you normally do, just read them and then delete them?] Unless it is something I want to buy.

These individuals did not receive as much email, and so had less well formed ideas of how to respond to such requests, e.g.,

I will probably give them the information that they asked for. And I would assume that I had already given them that information at some point so I will feel comfortable giving it to them again.

In short, none of these strategies particularly helped people to identify the well-constructed scams. Those choosing the opposite of the second strategy — that is, relying on the context of an email rather than the relationship with a business — may benefit from their suspicion of unexpected emails. They are relying on information outside the context of email, which will be particularly hard for scammers to manipulate. However, in some instances the scams will still get lucky – or pay off from casting a wide net – in hitting someone just when they are expecting another message. To the extent that more people have accounts with the spoofed company and are in the midst of a transaction, these occurrences may become more common. Moreover, spear-phishing techniques identifying people who would be expecting certain emails, e.g., to people who have items currently listed for sale on e-bay, might be particularly effective against this strategy. This study did not test any of these techniques.

Past experience with particular scams appeared to be the biggest factor in identifying similar (nearly identical) scams in the study. For example, the Katrina scam was based on a real email in circulation and is similar to what is commonly known as the Nigerian 419 scam. All but one participant dismissed it as a hoax, most saying that they had seen things like that before, many specifically mentioning Nigeria or Africa. Only one person did not spot the scam, which is not enough to warrant statistical analyses. However, just descriptively it is interesting to look at this individual's behavior. She described the email as "more professional," although she said she was not interested in doing business on the matter and might delete the message for that reason. But then she said, "or maybe [deleting the message] is not so nice from my part. Maybe I'll reply, and I will try to answer. A note, but a polite note." To anyone familiar with these scams, this sounds like a fairly naive response. However, this participant was not particularly gullible in other situations. For example, she was wary of the email containing the hidden .exe file, saying, "I don't usually trust so much the things that come through like this."

3.6 Pop-up Messages

When asked about warnings generally, only about half of participants recalled ever having seen a warning before trying to visit a web site. Their recollections of what they were warned about were sometimes vague, e.g., "sometimes they say cookies and all that," or uncertain, e.g., "Yeah, like the certificate has expired. I don't actually know what that means." When they remembered warnings about security, they often dismissed them with logical reasoning, e.g., "Oh yeah, I have [seen warnings], but funny thing is I get them when I visit my [school] websites, so I get told that this may not be secure or something, but it's my school website so I feel pretty good about it."

Other times, they try to do what they want to do without making much sense of the message:

First, when I click on some website, and I got a warning that the name looks strange, or unsecured. After I enter the website, I don't know, there are some websites that I got a lot of spam, when I'm clicking. So, then I got some messages, like blocking or something. That blocking, or, because I was not so sure what is there, because if I try to, if I close, and after that re-open, I couldn't enter sometimes. [Interviewer asks if participant still enters the website after a warning like that.] Depends on what kind of message. Because I can get a message that won't allow me to go further, or I get messages that tell me, it is a, I don't know exactly, when I try to download them or something. I'll ask my, ask the administrator access, or, there are some things that I cannot go further.

Others respond to warnings with caution, although perhaps not based on a thorough understanding, e.g.,

[Interviewer: Have you ever received any warnings when you have tried to visit a site]? Yes, it has said that this is not a secure site.[Interviewer: Was it clear about what the message meant?] No I just thought it was not a good idea to go there. [Interviewer: How do you usually respond?] I don't proceed. [Interviewer: Do you usually go into the site after these messages?] No.

When shown images of the common pop-up messages, 80% reported having seen at least one of them before. This higher recognition than recall of warnings is typical of familiar but poorly understood stimuli. Responses to the pop-up messages were consistent with this confusion. Table 5 indicates how many participants reported having seen each type of message before, how many would go on after seeing it, would stop, or would behave differently depending on various factors, such as whether they had visited the site before or if they trusted the site.

Table 5. Responses to Messages

Pop-Up Message	Seen	Go on	Stop	Depends
Leaving secure site	71%	58%	0%	42%
Insecure form	65%	45%	35%	20%
Self-signed certificate	42%	32%	26%	42%
Entering secure site	38%	82%	0%	18%

3.6.1 Pop-up message: Leaving secure site

Nearly half of participants reported considering factors about the site or their needs in deciding whether to go on after a warning, but more than half would go on without any apparent concern for the possible risk upon seeing that they were leaving a secure site. Some revealed a fairly complex understanding of what this message meant, e.g.,

> *Basically you were on an encrypted page and you might have just entered your name and your password, and that's still encrypted. But where you're about to go is not an encrypted place, so if you're going to read about something like your bills or whatnot, other people can easily read about your bills.*

Others were not as sure what this message meant, with reactions ranging from simply uninformed, e.g., "Huh, I'm really not certain, but I'm intrigued by it," to misdirected efforts to make sense of the message, e.g.,

> *Well, I mean, I'm figuring like, based off of what it seemed like an encrypted page kind of, I don't know, like walks out or crypts into the circle so that it can't be read.*

3.6.2 Pop-up message: Insecure form

This message explains that information submitted could easily be read by a third party, language that participants may have used to interpret the message if they had not previously paid attention to it. Many repeated back this explanation of the risk, although nearly half said that they would go ahead and submit their information without bringing other factors to bear.

Participants' descriptions of their behavior underscores this uncertainty, e.g., "I guess I'm not fully sure what encrypted means. So that's why I'm like, ok, whatever. I probably need to learn what encrypted means." Others express their uncertainty with even less indication that they will use the information in the message, e.g., "I would probably experience some brief, vague sense of unease and close the box and go about my business."

3.6.3 Pop-up message: Self-signed certificate

Participants appeared to be especially uncertain what to make of certificates, revealing some confusion between security and trust. Many respondents specifically said that they did not know what certificates were, and made inferences about how to respond to such a mysterious message. Some inferred that certificates were a formality, e.g.,

> *Basically that it's kind of like the elevator certificate. For whatever reason, they don't have it. But at that point sometimes when you go into the elevators you can see if their certificate is up to date or if it's not current. And that's kind of what that meant for me.*

Others imbued the message with more authority, e.g., "The wording says that it could not be verified as a trusted site, then I would not take the chance." Many, though, dismissed the warning, e.g., "I clicked yes because I felt it was safe enough, but I didn't really know what it was and I didn't want to check." Some used previous experience as their basis for ignoring it, e.g., "I have no idea [what it means], because it's saying something about a trusted website or the certificate hasn't, but I think I've seen it on websites that I thought were trustworthy."

For some, as with the other messages, this message prompted them to realize their limited knowledge, e.g.,

> *You know I'm not sure because I'm wondering what, what authority certifies websites as being secure to begin with. Like that's what question this box prompts me to ask. And I don't know the answer to that question, so I really can't say what it means. [Interviewer: OK, so again would you do anything differently if you got this message?] If I was already comfortable with going to it I would click ok and accept it.*

3.6.4 Pop-up message: Entering secure site

In contrast to the tendency to ignore messages that warn about possible risk, several participants in this study indicated that they would hesitate even upon receiving the message about entering a secure site, suggesting that the mere presence of a pop-up message sends a negative signal that users may not know how to interpret. Some appeared to interpret the fact of a pop-up box as a warning, misinterpreting the information that it presented, e.g.,

> *[This message means] that I wanted a secure website, and, the website has been verified as being authentic, but it's not secure…. I will be really wary about entering any information…. I mean, I will probably enter the website, but, depending on, again, like I said, on the information and the, I'll be really wary about it.*

Others correctly interpreted the information, but were reluctant to take it at face value, e.g.,

> *No [I have not seen this before]. And you know I don't know if I would believe it…. It is just like, ahhh everything is just wonderful and perfect security…. I would still click ok but I will wait before I entered any, you know, sensitive information.*

Others voiced concerns about encryption that belied a comprehensive understanding, e.g.,

> *I wouldn't be that concerned because it is an encrypted page. Unless the encrypted page is like, it could be dangerous if in the encryption they put like a virus or something in there.*

4. Future Work

The preliminary themes described in this work will be further fleshed out in the full mental models analysis of all forty

participants in the larger study from which these data were drawn. In addition to a larger sample, that analysis will include more in-depth coding of concepts to better represent the working model that naive computer users have of Internet security and trust. Furthermore, a small sample of computer users with expertise in security will be included as a comparison group.

The insights gained from these interviews provide valuable launching points for both further descriptive and corrective efforts. The results from this and the full mental models analysis will be followed up with a set of written surveys administered to a larger, more representative sample, to determine how prevalent these various beliefs are, and for more statistical power to determine how different beliefs and behaviors are correlated.

Furthermore, these results will help to guide development of tools to assist computer users in identifying and protecting themselves from phishing schemes. Further research might explore the perceived consequences and costs of these tools, to examine the trade-off between the risks of semantic attacks and the costs of intervention. Evaluation of the tools would provide external validity for the relationships between variables described in these interviews, and will provide valuable information about the causality of the links described here. We can tentatively propose that the links between email decision strategies and susceptibility to phishing schemes reflect the effectiveness of these strategies. A controlled experiment, such as the evaluation of tools enabling more promising strategies or guarding against ineffective ones, is required to test such causal hypotheses. Such an experiment, using participants' own email and information, would also allow more realism in testing their susceptibility to phishing, compared to the false information used in the role-play segment of this study.

5. DISCUSSION

5.1 Summary of findings

In general, these participants, selected to be relatively naive about computer security, were aware that there were risks associated with using the Internet and that they needed to protect their computer from problems like malware. However, they appeared to be less aware of social engineering attacks aimed at eliciting information directly from them.

All participants had noticed various cues that they might use to determine whether an email or web site was trustworthy, such as obviously false addresses in the "from" line, a lock icon, or broken images on the page. However, they did not necessarily interpret these cues appropriately. For example, few knew that a lock in the content of a web page did not mean the same thing as a lock in the browser's chrome, and many interpreted broken images as problems with their computer rather than an indication about the source of the site. Fewer noticed cues in URLs, and those who did were not particularly savvy about how to interpret them.

Participants used various strategies to make decisions about the trustworthiness of email, mostly centered around interpreting the text of the email rather than any more reliable cues in headers or URLs associated with links. None of these strategies appeared to be particularly effective in helping these naive users avoid falling for scams. Familiarity with very particular scams seemed to be the best predictor for spotting similar ones, but this benefit did not seem to extend to unfamiliar scams. Using off-line context to make decisions also appeared to be somewhat helpful in protecting people from possible scams in this context, although it is exactly this kind of information that is exploited by spear phishing, making it a potentially unreliable strategy.

Finally, participants had some difficulty making sense of standard pop-up messages, especially those that appeared to be warning about something that did not require action. Prior experience with such messages may have contributed to participants' tendency to feel comfortable ignoring warnings that they did not understand.

5.2 Limitations

Given this small and non-representative sample, we can't extrapolate prevalence of beliefs to the general population. We purposefully selected participants who were more naive than the average, in order to understand how those without a good understanding of security make sense of Internet risks. The sample was not much different from the broader population of those who responded to our advertisements except along a few dimensions that may correlate with computer expertise, namely choice of computer platforms, sex and racial background. It is important to recognize that those selected by our criteria had been using computers for as many years and received as many emails as those whose expertise disqualified them. Thus, this population is at particularly high risk for phishing attacks, and so is especially important to study.

The motivation of mental models interviews is to generate insights into how people view the decision problem. With reasonable sharing of beliefs among individuals living in the same society and sharing common experiences, a sample of 15 will be likely to reveal (at least once) any belief held by 10 percent or more of the population [19]. Our sample of 20 individuals varying in experience and demographics, but with a shared environment of computer security risks, should be sufficient to elicit many important beliefs held by the population. Our age range is limited to younger adults, so we have no data about children, adolescents or elderly computer users, where there might be more evident differences. Overall, our participants were more educated than the national average (65% had college degrees). However, the patterns of knowledge, awareness & behaviors are likely to generalize. A survey will follow up to establish prevalence and replicate patterns.

5.3 Relationship to previous work

Consistent with previous research, we found that users in this study were not particularly savvy about distinguishing between legitimate emails and semantic attacks. In particular, many users do not interpret pop-up messages in useful ways [9] and many miss cues that could be found in the address bar and status bar [13]. This study recruited participants who were particularly naive about security, targeting a population that may be at particularly high risk in this domain.

One element that this research adds to the literature is the role of awareness and experience in these decisions. Whereas many studies have distinguished between naive and expert computer users, revealing the former to be relatively unaware of security cues and information, our data suggest that even quite naive computer users vary considerably in the type of decision strategies that they use, some of which can be quite complex and are based in part on the particular experiences that they have had in the past. The fact that these strategies tend not to be very effective may be tied to a lack of basic understanding of the Internet.

Previous research has shown that people tend to prefer cues in a site's content rather than more authoritative tools [9], which is also consistent with our findings. Warnings and toolbars may use terms that are often not understood, but which people attempt to make sense of using their imperfect understanding, if only to make messages go away or to do what they're trying to do. If naive computer users do not differentiate between good and bad cues, or — worse yet — find the less reliable cues to be easier to make sense of, then that leaves them particularly vulnerable to semantic attacks.

5.4 Implications for development of tools

These interviews show that Internet users have little awareness of phishing and limited knowledge about effective strategies for detecting fraudulent emails and web sites. The strategies they employ to protect themselves are at times ineffective or even counter-productive. Security warnings and indicators provided by web browsers seem to have little or no meaning to many Internet users, and concepts such as encryption are poorly understood. In particular, pop-up messages may be assumed to be providing information that needs to be acted upon, and thus perhaps should only be used in such circumstances. When developing anti-phishing toolbars, it is important to keep in mind that most users are unlikely to understand the implication of information related to domain registration, certificates, or other technical concepts. Security tools need to explain to users how the information they provide is relevant to them or, better yet, recommend a course of action based on this information. For example, alerting a novice user about a 'self-signed certificate' may be less meaningful than warning that 'your information will be transmitted safely, but you should take a look at where it's going.'

There appears to be a need to educate Internet users about how to avoid falling for phishing scams. However, it is important not to assume very much existing knowledge about phishing and to try to communicate this information in ways that will help people generalize it and apply it as they are exposed to the ever-changing strategies of attackers. Simply teaching people to look for a security indicator is unlikely to lead to their understanding the implications of that indicator. Thus, people trained to look for indicators are likely to fall for unsophisticated spoofs of such indicators. Likewise, our interviews suggest that people who are suspicious about emails that request financial information may feel confident complying with emails that request non-financial information, even if those passwords would grant access to the use of credit cards. It is not readily apparent to some users that there are risks associated with logging into a scam artist's web site even if the only information captured is their user name and password. Education therefore needs to begin at a very basic level and to explain the intuition behind recommended strategies in a non-technical way.

Although our interviews did not ask specifically about spear-phishing attacks, they do raise concerns that people will have great difficulty avoiding such attacks as the few strategies people currently employ are likely to be largely ineffective against such personalized attacks.

Our group has begun exploring the development of educational games as well as email clients and anti-phishing toolbars that can provide educational information on avoiding phishing scams. The results of our interviews will be important in guiding those efforts.

6. CONCLUSION

The pattern of results emerging from this small interview study suggests that merely being aware of phishing or of cues is not enough to protect people from scams, especially new ones about which they might not be aware. Mere experience with scams appears to be associated with ineffectual strategies. More contextualized strategies might be more effective, relating emails to specific offline information. For example, merely having a PayPal account does not legitimate an email that says it's from PayPal. Learning to be wary of unexpected emails may be a more effective approach. A browser tool that can help guide users to bring message-specific (as opposed to merely sender-specific) offline information into their decision might be a particularly effective feature for helping people to identify phishing ploys.

Some of the cues that appear to be most effective in helping people avoid phishing schemes are the most extreme, e.g., merely being wary of any request for a password. However, this cue appears to work not by helping people to better spot the scams but merely by shifting people's responses so that they are more cautious in response to any emails. The benefits of avoiding phishing may outweigh the (smaller) risk of suspecting a real email. However, this avoidant approach is unlikely to work for most people, who gain great benefits from conducting business on the web and are unlikely to want to subject all emails to such scrutiny. Indeed, more than one member of our research team has suffered an inconvenience (interruption of telephone service) from ignoring what appeared to be a suspicious message.

7. ACKNOWLEDGEMENTS

We gratefully acknowledge support from National Science Foundation grant number 0524189 entitled "Supporting Trust Decisions," and from the Army Research Office grant number DAAD19-02-1-0389 entitled "Perpetually Available and Secure Information Systems." We would also like to thank Alessandro Acquisti, José Brustoloni, Sven Dietrich, Jason Hong, Norman Sadeh, Serge Egelman, Ian Fette, Mark Huneke, Faisal Jawdat, Ponnurangam Kumaraguru, and Steve Sheng for their helpful insights in development of the interview protocol and thoughtful comments aiding interpretation of the interview results.

8. REFERENCES

[1] O'Brien, T.L. Gone spear-phishin'. 2005. *The New York Times* (4 December) http://www.nytimes.com/2005/12/04/business/yourmoney/04spear.html?pagewanted=1&ei=5088&en=2f313fc4b55b47bf&ex=1291352400&partner=rssnyt&emc=rss

[2] Jagatic, T., Johnson, N., Jakobsson, M., Menczer, F. Social Phishing. *Commun. ACM. To appear.* http://www.indiana.edu/~phishing/social-network-experiment/phishing-preprint.pdf

[3] Anti-Phishing Working Group (APWG). 2005. Phishing Trends Report, December 2005. http://www.antiphishing.org/reports/apwg_report_DEC2005_FINAL.pdf

[4] Emigh, A. Online Identity Theft: Phishing Technology, Chokepoints and Countermeasures. Identity Theft Technology Council Report. October 3, 2005. http://www.antiphishing.org/Phishing-dhs-report.pdf

[5] Schneier, Bruce. 2000. Semantic Attacks: The Third Wave of

Network Attacks. *Cryptogram Newsletter* (October 15). http://www.schneier.com/crypto-gram-0010.html

[6] Miller, R. and Wu, M. 2005. Fighting Phishing at the User Interface. In L. Cranor and S. Garfinkel, eds., *Usable Security*. O'Reilly 2005.

[7] Chou, N., Ledesma, R., Teraguchi, Y., and Mitchell, J.C. 2004. Client-Side Defense Against Web-Based Identity Theft. *11th Annual Network and Distributed System Security Symposium (NDSS '04), San Diego, CA*. http://www.isoc.org/isoc/ conferences/ndss/04/proceedings/Papers/Chou.pdf

[8] Ye, Z., Smith, S., and Anthony, D. 2005. Trusted paths for browsers. *ACM Trans. Inf. Syst. Secur.* 8, 2 (May. 2005), 153-186. DOI= http://doi.acm.org/10.1145/1065545.1065546

[9] Dhamija, R. and Tygar, J. D. 2005. The battle against phishing: Dynamic Security Skins. In *Proceedings of the 2005 Symposium on Usable Privacy and Security* (Pittsburgh, PA, July 06 - 08, 2005). SOUPS '05, vol. 93. ACM Press, New York, NY, 77-88. DOI= http://doi.acm.org/10.1145/ 1073001.1073009

[10] Herzberg, A., and Gbara, A. 2004. TrustBar: Protecting (even Naïve) Web Users from Spoofing and Phishing Attacks. Cryptology ePrint Archive, Report 2004/155. http://eprint.iacr.org /2004/155.

[11] Wu, M., Miller, R.C., and Garfinkel, S.L. Do security toolbars actually prevent phishing attacks? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montreal, Quebec, Canada, April 22 - 28, 2006). CHI '06.

[12] Friedman, B., Hurley, D., Howe, D., Felten, E., Nissenbaum, H. 2002. Users' conceptions of web security: a comparative study. In *CHI '02 extended abstracts on Human factors in computing systems, Minneapolis, Minnesota*, 746-747.

[13] Dhamija, R., Tygar, J.D., and Hearst, M. 2006. Why phishing works. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montreal, Quebec, Canada, April 22 - 28, 2006). CHI '06.

[14] Gentner, D., & Stevens, A. L. (1983). Eds. *Mental Models*. Hillsdale, NJ: Erlbaum.

[15] Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge, MA: Harvard University Press.

[16] Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

[17] Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin, 100*, 349-63.

[18] Fischhoff, B., & Downs, J. (1997). Accentuate the relevant. *Psychological Science, 8*, 1-5.

[19] Morgan, M. G., Fischhoff, B., Bostrom, A., & Atman, C. (2001). *Risk communication: The mental models approach*. New York: Cambridge University Press.

[20] Bruine de Bruin, W. & Fischhoff, B. (2000). The effect of question format on measured HIV/AIDS knowledge in detention center teens, high school students, and adults. *AIDS Education and Prevention, 12*, 187-198.

[21] SPSS for Mac OS X, Rel 11.0.4 2002. Chicago: SPSS Inc.