

Enlisting Principles from Cognitive Psychology to Create More Secure Interfaces

Awarded Research Grant on 12/7/18

General description of your Honors Thesis/Project topic.

Phishing is an umbrella term that describes fraudulent methods by which an attacker attempts to steal personal or sensitive data by posing as a trustworthy entity. This attack can occur over various channels, with the most common vectors being email and websites, and usually works in tandem to create a false representation of legitimacy for average users. The fraudulent link loads malware that steals sensitive information when clicked on. Deceptive phishing, the specific phishing attack I focus on through my thesis, refers to an attack by which attackers mimic a legitimate company and attempt to steal a customer's credentials (Bisson). These tactics boast a high success rate, and the effectiveness is due to its manipulation of human behavior. The specific language in a phishing email body uses emotionally-charged language to limit a user's critical thinking by raising levels of fear, sadness, or anger to compel one to take immediate action. For example, a bank phishing email may urge people to verify their account credentials via an online link to prevent something catastrophic from happening to their account. A panicked bank customer may overlook the fact that the redirected website is illegitimate and give up their information. Additionally, attackers make emails look more convincing by recreating typical scenarios one would encounter with an entity and using believable branding.

Hackers aim to manipulate a user's mental models and visual perception lapses. The more closely a deceptive phishing email resembles a legitimate company's official email, the higher the success rate of the phishing attack. Hence, users should verify all URLs to see if they redirect to an unknown website. They should also look out for generic greetings and grammatical errors throughout the email (Barkly). Users are expected to methodically follow these precautions when checking their email, but in reality, they pay very little attention to the email content in typical browsing scenarios. Regardless of their technical background, a user will not do this every single time they check email or use a web browser. People click on malicious URLs by accident when they do not take the time to look at the URL carefully.

Therefore, email software should implement proactive indicators that warn the user before clicking alongside existing reactive security indicators which only notify the user after they click a link. These are more in line with typical user behavior and therefore should be devised to combat phishing attacks and raise general security awareness.

Statement about the scientific problem or intellectual/creative intent. Description of objectives and hypotheses or other discipline-specific inquiry.

Google Chrome and similar modern browsers contain a built-in indicator that checks the security of a website. If a site is deemed to be secure, Google features a tiny gray lock in the left-most corner of the browser. For websites that it considers insecure, the indicator displays a red caution icon and "Not Secure" in red text in the browser. Compared to the overall website content, the security indicator has a small font and icon, resulting in the website content overpowering it. The disparity of effectiveness between browser security indicators demonstrates that some browsers present security information more compellingly than others

do, and therefore information presentation is a critical factor in determining effectiveness (Akhawe, Felt). Unlike browsers, emails provide no sign of whether there is any danger in a particular link, and the user has to know how to classify the URL as dangerous. Emails usually feature URLs in a hyperlink format, hidden under a button design or plain text. Traditionally, a hyperlink is an example of a mouseover (Paganini), which is an element that gets activated when a user moves or “hovers” their mouse pointer over the element’s trigger area. Usually, the trigger area is some underlined text colored either blue or black, and a user can carefully hover over the text with their mouse to uncover the destination of the URL. When they do this, a small pop up appears containing the full URL. Emails do not present this pop up in a way that prevents a user from accidental clicking, especially if a user is in a hurry. Constrained to a tiny box, the URL’s long string of characters may cause a user to overlook small details within it.

Through my thesis, I aim to redesign existing email interfaces to present pertinent information that can aid a user in classifying whether a redirected link is legitimate or not. I want my implemented features to improve the design and presentation of URLs within the email - this can be in regards to size, color, and many more aspects. Previous studies focus more on security messaging in the browser, but I aim to shift the focus to deceptive emails as users typically end up on malicious websites by clicking on hyperlinks within the malicious email. My design aims to prevent the user from even clicking on a hyperlink if there is any doubt of its legitimacy. I also intend to give the end user the final say in the legitimacy of a link as well as provide discrete choices to help them make an informed decision about the URL.

Despite existing indicators, at this stage, human users are still going to be the best at validating whether a URL is secure, so by presenting the URL in a more attention-grabbing way, the user has a more proactive indicator of the redirected website's legitimacy. This dynamic messaging can warn the user and prevent deceptive phishing attacks before they happen. I will leverage principles from visual and cognitive psychology research as the basis for creating these components as well as devising the experimental procedure. Cognitive psychology is the study of mental processes such as attention, perception, problem-solving, and more. This field provides insight into how users make decisions, prioritize information, or learn skills. Specifically, visual psychology research focuses on how the eyes and brain process information and drive interactions with the outside world. When using a web or mobile interface, humans rely on their vision to form judgments, process information and allocate attention to objects. Psychology lends an overall deeper insight on how humans interact and perceive information on an interface, and having this solid background allows me to create the redesign to be in line with a user’s typical behavior.

Relate the key literature to the problem or creative endeavor and explain the study's importance to the advancement of knowledge in the discipline.

Phishing can pose disastrous consequences to individual users, by causing identity theft, to Fortune-100 Companies, in the form of data breaches - recall Target and Verizon. Below are some statistics that illustrate the relevance of these attacks (Barkly):

- 76% of organizations say they experienced phishing attacks in 2017
- Users only reported 17% of phishing campaigns

- By 2017, the average user had 16 malicious emails per month

Many companies address phishing by incorporating employee phishing awareness training, but previous studies argue against the efficacy of these methods. After attending phishing awareness training, many employees are still unable to detect cleverly disguised phishing emails, especially if they mimic credible emails that employees expect to find in their inboxes (Paganini). Training that focuses on reactive education does not anticipate the typical user behavior under attack and does not protect them in practice.

Humans are easily distracted and tend to make irrational decisions. In his book *Thinking, Fast and Slow*, psychologist Daniel Kahneman sheds light on our irrational behavior by describing the brain's two-system thinking process in producing decisions. System 1 is very automatic and intuitive, while System 2 is rule-based and logical and has been known to correct System 1 when it is prone to violating rules (Kahneman). In circumstances that hinder logical thinking such as panic and stress, humans make errors as System 2 fails to intervene on behalf of System 1. The emotionally charged messaging found in phishing hinders System 2 and encourages the individual to give up some information or commit an act they would not usually do.

Therefore, a more feasible method of solving this problem is through technical interventions that prevent the user from attacks altogether. I hypothesize that my redesign, with its use of colors and messaging, will not only catch the user's attention but will calm them down and let them rationally pick an option. This design eliminates the user's need to recollect past training, saving them sizeable cognitive overload when they are already in a stressful situation.

Phishing is a rapidly growing attack, and two-thirds of organizations have reported receiving targeted and personalized attacks (Barkly). With the potential to destroy an organization's reputation and cause data losses worth millions of dollars, newer, proactive approaches to preventing a phishing attack are needed.

1. Akhawe, Devdatta, and Adrienne Porter Felt. *Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness*. 2013, pp. 257–72. www.usenix.org,

<https://www.usenix.org/conference/usenixsecurity13/technical-sessions/presentation/akhawe>.

2. The Barkly Team. "Must-Know Phishing Statistics 2018." Barkly Endpoint Security Blog, Barkly Protects, Inc., Aug. 2018, blog.barkly.com/phishing-statistics-2018.

3. Bisson, David. "6 Common Phishing Attacks and How to Protect Against Them." *The State of Security*, Tripwire, 5 June 2016, www.tripwire.com/state-of-security/security-awareness/6-common-phishing-attacks-and-how-to-protect-against-them/.

4. Daniel Kahneman (25 October 2011). *Thinking, Fast and Slow*. Macmillan. ISBN 978-1-4299-6935-2. Retrieved 8 April 2012.

5. Paganini, Pierluigi. "New Intel Security Study Shows That 97% of People Can't Identify Phishing Emails." *Security Affairs*, Security Affairs, 18 May 2015, securityaffairs.co/wordpress/36922/cyber-crime/study-phishing-emails-response.html.

Goals of Literature Review

Through the literature review, here are some areas I wish to study:

- The nature of phishing attacks: who gets targeted, how victims respond, attacker methodology, etc. to provide context on the effectiveness of such attacks. I investigate

existing security indicators within browsers and emails as well as previously proposed solutions to figure out areas of improvement.

- Past research in interface design and visual cognition to compile principles that justify my changes in the display of web pages and emails. I have already started gathering some principles, but I aim to get more detailed and relevant principles after completing the literature review:

How users interact with a web interface:

- Humans have a central-viewing bias when it comes to viewing screens (Bindemann).
- Inattention blindness: Because humans navigate a website based on a specific goal they have in mind, less prominent elements become overlooked. Usually, the URL is not the central aspect the user is trying to find and gets ignored (Rensink).
- Scanning: One of the most critical facts about web users is that they scan for the most relevant elements rather than read all the information (Krug).

These principles highlight that from the start, humans tend to not focus on the URL of a website. Thus, some principles below may lend some insight into how to redesign the browser to draw attention towards it:

- Exogenous orienting: A sudden change/movement in the periphery redirects attention towards it.
- Gestalt Similarity Principle: Items that look similar are perceived to belong together, but they also may be more easily confused with each other. It is hard to find distinctions between items with visually similar colors, sizes, or shapes. So, the eyes are more drawn to contrasting elements.

A significant component of my literature review is dedicated to identifying principles like the above mentioned and reviewing existing research to learn more about these specific topics.

1. Bindemann, Markus. "Scene and Screen Center Bias Early Eye Movements in Scene Viewing." *Vision Research*, vol. 50, no. 23, Nov. 2010, pp. 2577–87. *PubMed*, doi:[10.1016/j.visres.2010.08.016](https://doi.org/10.1016/j.visres.2010.08.016).
2. Krug, Steve. *Don't Make Me Think, Revisited: A Common Sense Approach to Web Usability*. Third edition, New Riders, 2014.
3. Rensink, Ronald A. "When Good Observers Go Bad: Change Blindness, Inattentional Blindness, and Visual Experience." *Psyche*, Aug. 2000, <http://cogprints.org/1050/>.

How are you conducting your research?

Using knowledge from my literature review, I will design prototypes that I hypothesize will reduce the prevalence of phishing attacks in an email. I will conduct user testing for my prototypes through an experiment, and I will analyze the results of the experiment and draw conclusions about the viability of my prototypes.

What procedures/techniques are you using to gather information/data or to create your project?

I am designing an experiment to conduct my user testing and evaluate my hypothesis. As security is not the primary objective of users when interacting with an interface, I will disguise the experiment as a usability study in which the participants will roleplay as a character checking their email. It will be a single-blind experiment as I will randomly divide participants into two groups. Group one will have emails with none of my design changes, while group two will have emails with my design changes. Both groups will get a random order of legitimate and malicious emails, and both groups will have emails with the same content presented to them (ex: fake Bank of America email, fake Discover card email). In a trial that contains a malicious email, I will define clicking on any of the hyperlinks as being “susceptible to a phishing email.” I will calculate the number of times a participant clicks on a bad link in comparison to the number of bad emails I present to them. After the experiment, I will debrief participants about the true intentions of the research. I will ask them questions about their computer knowledge to get a better demographic understanding of the participant group.

What resources or materials are you using in your research?

- Octave - An open source programming language for scientific computing
- Linux - Open-source operating system that the computers in the lab use.
- Sketch - Digital design software that I will use to create my redesign prototypes.
- Standard PCs in the lab
- Psychtoolbox Software - Octave/MATLAB functions specifically used for conducting vision and neuroscience research. I am using Psychtoolbox to control certain aspects of how my prototypes are projected on the computer screen, as well as measure and collect data on how the participants respond to my screenshots.

Completed training:

- CITI online ethics training
- PBS Department ethical training session

How often will you meet with your primary Committee Chair?

Completed training:

- Weekly 1-hour meetings/check-ins to discuss readings, insights gained, and the process of designing the experiment
- 1.5-hour general lab meetings (attended by everyone in the Visual Cognition & Attention lab) to give status updates on research work

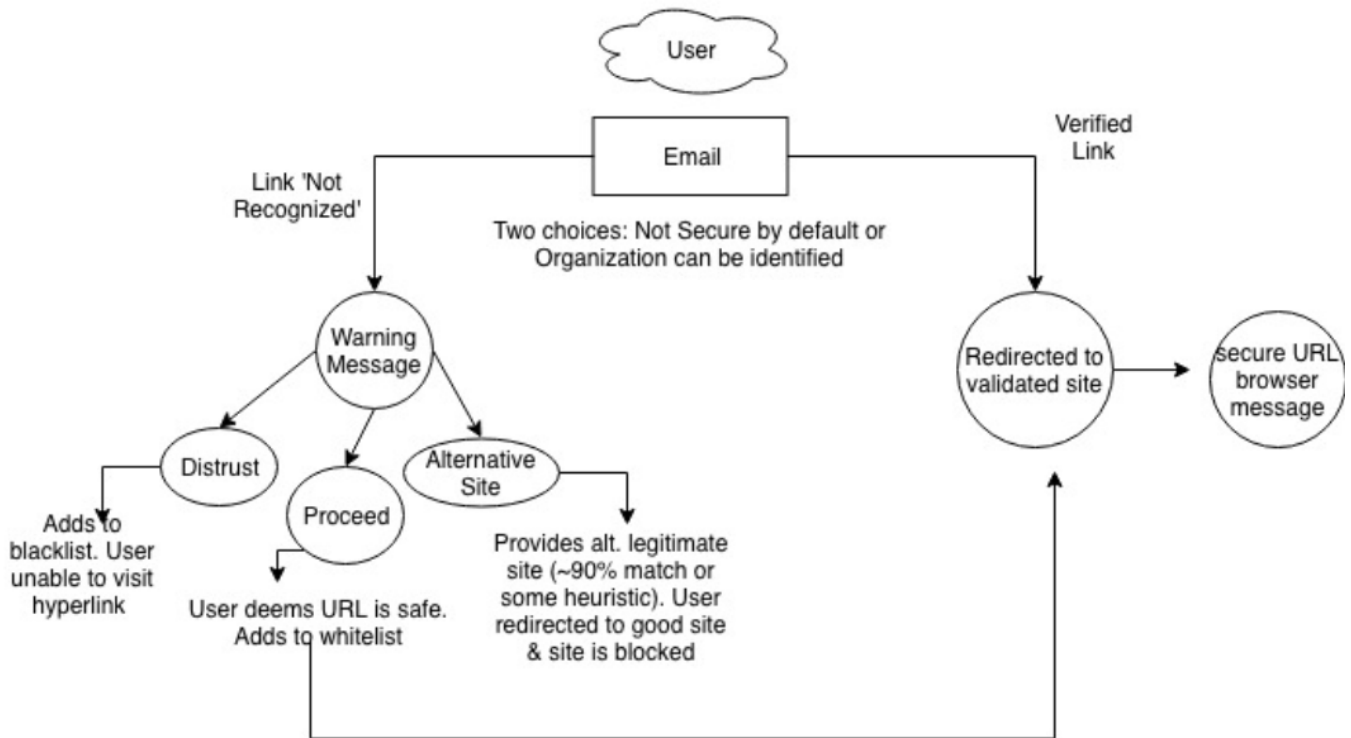
What are your committee chair's expectations of such meetings? What are your committee members' expectations about meeting with you?

- In every meeting, my committee chair expects me to demonstrate progress from previous meetings. I must bring relevant research and materials to discuss my ideas and we will spend our time addressing issues and determining an agenda for next week's meeting. Overall, they expect that I am not slacking off, keeping a journal of progress/activity, and raising any issues or concerns in a timely manner.
- Commit 9 hours of research each week

Thesis Prototypes


06

Current redesign as of 12/4/18



I created this chart to show the suggested flow of my design. I classify a link as 'recognizable' if it belongs to a major organization or if the user has previously trusted it. If a link is unrecognizable, then the email presents a generic warning. Upon reviewing the links, a user can 'Distrust' or 'Proceed' to the link. If the software can identify a legitimate alternative, it displays it.

Sample Case: Not Recognized

Email Title
 First Last
firstlast@gmail.com

This message contains unrecognizable links.
Double check links and be cautious.

Review Links

Dear First,

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

In the 'Link Not Recognized' case, a generic warning appears between the email sender information and the email body to give a proactive indicator of potential phishing. The 'Review Links' button displays any potentially malicious links with various options.



bankoffamerica.com	Recommended Alternative: bankofamerica.com	Redirect
xxxdaf-345.com	Distrust	Proceed