Neural Underpinnings of Website Legitimacy and Familiarity Detection: An fNIRS Study

Ajaya Neupane
University of Alabama at Birmingham
aneupane@uab.edu

Nitesh Saxena
University of Alabama at Birmingham
saxena@uab.edu

Leanne Hirshfield
Syracuse University
lmhirshf@syr.edu

ABSTRACT

In this paper, we study the *neural underpinnings* relevant to user-centered web security through the lens of *functional near-infrared spectroscopy (fNIRS)*. Specifically, we design and conduct a fNIRS study to pursue a thorough investigation of users' processing of legitimate vs. illegitimate and familiar vs. unfamiliar websites. We pinpoint the neural activity in these tasks as well as the brain areas that control such activity. We show that, at the neurological level, users process the legitimate websites differently from the illegitimate websites when subject to phishing attacks. Similarly, we show that users exhibit marked differences in the way their brains process the previously familiar websites from unfamiliar websites. These findings have several defensive and offensive implications. In particular, we discuss how these differences may be used by the system designers in the future to differentiate between legitimate and illegitimate websites automatically based on neural signals. Similarly, we discuss the potential for future malicious attackers, with access to neural signals, in compromising the privacy of users by detecting whether a website is previously familiar or unfamiliar to the user.

Compared to prior research, our novelty lies in several aspects. *First*, we employ a neuroimaging methodology (fNIRS) not tapped into by prior security research for the problem domain we are studying. *Second*, we provide a focused study design and comprehensive investigation of the neural processing underlying the specific tasks of legitimate vs. illegitimate and familiar vs. unfamiliar websites. *Third*, we use an experimental set-up much more amenable to real-world settings, compared to previous fMRI studies. Beyond these scientific innovations, our work also serves to corroborate and extend several of the findings of the prior literature with independent methodologies, tools and settings.

Keywords

Phishing Detection, Privacy Attacks, fNIRS

WWW 2017, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4913-0/17/04.
http://dx.doi.org/10.1145/3038912.3052702

1. INTRODUCTION

Gaining insights into the innate behavior of end users, when they are faced with security threats while browsing the web, is an established line of research in computer security. Many prior studies have been conducted to evaluate users' *performance* in detecting web-based attacks and adhering to browser security warnings (e.g., [7, 17, 19, 22, 43, 48, 54]). A recent innovation in this research thread is a class of studies that measure users' low-level, *neural processes* underlying the security threat (e.g., [37, 50, 8, 36]). These studies have used traditional neuroimaging techniques, namely, functional magnetic resonance imaging (fMRI) and Electroencephalogram (EEG), to tap into users' brain signals, which boast to provide unique insights into users' behavior that may not be possible to capture through performance studies alone.

In this paper, we follow this latter class of studies to understand the neural mechanics in the context of user-centered web security through the lens of *functional near-infrared spectroscopy (fNIRS)*. fNIRS is a non-invasive imaging technique for brain activity measurement in naturalistic settings that uses light in the near infrared range (700-900 nm) to penetrate the skull and provide a measurement of changes in the ratio of deoxygenated (deoxy-Hb) and oxygenated hemoglobin (oxy-Hb) in active areas of the cortex of the brain. We carefully selected fNIRS as our study platform since it offers the spatial resolution better than EEG and similar to that of fMRI, while allowing for the measurement of brain activity in near-real-world conditions (not inside a scanner and not in a supine posture, unlike fMRI).

First, we consider the problem of *website legitimacy detection*, commonly referred to as *phishing detection*, in which the user has to determine whether a website the user is about to login to is legitimate ("real") or illegitimate ("fake", or a close replica of real). We dissect the human behavior underlying such real vs. fake website detection based on fNIRS neural signals.

Further, we also consider the problem of *website familiarity detection*, in which an attacker attempts to infer whether a given website the user is browsing is familiar to the user or not. This represents an attack against the privacy of the user since it allows the attacker to, for example, learn whether the user has an account with a website or has previously visited the website, which can be used to track the user online or later launch phishing and other social engineering attacks against the user. In this paper, we aim to study the differences in brain activation based on the familiarity of websites. This constitutes a form of *side channel attack*,

in which the information about the familiarity of the website can be extracted through *neural side channels*, which may be accessible to a malicious program running on the user's computer connected with the fNIRS device. Similar attack models have been studied by researchers using EEG-based brain computer interfaces, but focused on familiarity of people, locations or ATM machines [34], *not websites*.

Our Contributions and Novelty Claims: In this paper we study human behavior underlying website legitimacy detection and website familiarity detection based on the neurological phenomena captured by fNIRS. To this end, we design and conduct an fNIRS study to pursue a thorough investigation of users' processing of real vs. fake and familiar vs. unfamiliar websites. In particular, we pinpoint the neural activity in these tasks as well as the neural regions that control such activity. Our results show that, at the neurological level, users process the legitimate websites differently from the illegitimate websites when subject to *phishing* attacks. Similarly, our results show that users exhibit clear differences in the way their brains process the previously familiar websites from unfamiliar websites.

These insights drawn from our study may have important defensive and offensive implications. In particular, these differences may be used by the system designers in the future to differentiate between legitimate and illegitimate websites programmatically based on neural signals, offering an additional layer of security against phishing attacks beyond traditional phishing detection mechanisms. Similarly, we discuss the potential for future malicious attackers, with access to neural signals, in compromising the privacy of users by detecting whether a website is previously familiar or unfamiliar to the user.

Compared to prior research, our novelty lies in several aspects. *First*, we employ a neuroimaging methodology (fNIRS) not tapped into by prior security research for the problem domain we are studying. Prior to our work, fNIRS has been used to build user authentication mechanisms based on neural biometric patterns [44], but not website legitimacy and familiarity detection. *Second*, we provide a focused study design and comprehensive investigation of the neural processing of legitimate vs. illegitimate and familiar vs. unfamiliar websites. Prior neurological studies [37, 36] involved multiple security tasks (phishing detection and warnings adherence) in the same session, which may have resulted in participant fatigue and less number of trials per task. *Third*, we use an experimental set-up much more amenable to real-world settings, compared to previous fMRI studies [37, 8]. In the fMRI condition, the user has to perform the study tasks in the enclosure of the fMRI scanner and under a supine posture, which is not a realistic scenario for web browsing. *Fourth*, we discuss the potential for building automated website legitimacy detection mechanisms based on neural data.

Beyond the aforementioned innovations, one important scientific attribute of our work lies in corroborating and extending several of the findings of the prior literature based on independent methodologies, tools and settings. Notably, we identify several of the same neural signatures vis-a-vis real and fake websites as exhibited in the fMRI study of [37], but under more naturalistic experimental setting. Specifically, like the previous study, we found activation in frontopolar cortex and orbitofrontal cortex, the neural areas involved in making decisions and evaluation of trustworthiness in the phishing detection task. Similar to previous studies [37, 36], we also found neural differences in how users detect real and fake websites. This result is also in line with related neuroscience literature on identifying real and fake paintings [28].

Practicality of fNIRS: Although the fNIRS systems are currently generally expensive, we believe that our methodology is valuable for user-centered security investigations and can have applications in organizations with high security requirements, such as national defense. Also, in the near future, these devices are expected to become miniaturized and light-weight similar to the commodity Brain Computer Interface (BCI) headsets. In fact, small, portable and wireless versions of fNIRS [4, 5] are already available in the market. Nevertheless, the fNIRS probe cap we used in our study was light-weight, designed for flexibility and comfort to perform tasks seamlessly.

2. BACKGROUND & PRIOR WORK

2.1 fNIRS Overview

Emerging brain activity sensing devices, such as functional near-infrared imaging (fNIRS), allow the researchers to explore brain activation states in real-world and simulated real-world environments, safely and non-invasively [29]. fNIRS in particular holds great potential for non-invasive brain measurement in naturalistic settings due to its practical nature, ease of set-up, robustness to motion artifacts, and high spatial resolution [29, 27]. The fNIRS technology uses light in the near infrared range (700-900 nm) to penetrate the skull and provide a measurement of changes in the ratio of deoxygenated (deoxy-Hb) and oxygenated hemoglobin (oxy-Hb) in the cortex of the brain. Optical fibers are placed on the surface of the head for illumination while detection fibers measure light which reflects back (Figure 1).

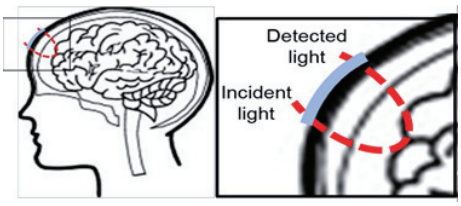

Figure 1: fNIRS functioning: Near infrared light is pulsed into the brain and light detectors measure the light reflected out of the cortex.

2.2 Related Work

Phishing is the act of luring people to reveal their private information using spoofed websites. Researchers have conducted a number of human-centered phishing detection studies (e.g., [17, 19, 22, 43, 48, 54]), which focus on users' task performance in identifying phishing scams and the effectiveness of anti-phishing toolbars, security indicators and phishing warnings. These studies have generally revealed that users do not pay attention to the browser-based phishing cues and often make incorrect choices.

There have been few recent studies which not only focus on the users' task performance but also on the underlying neural processes controlling users' decision making under security tasks (e.g.,[37, 50, 8, 36]). Related to our work is the fMRI study conducted by Neupane et al. [37], where users

were subjected to phishing detection and malware warnings adherence tasks. They reported that users exhibit higher activation in brain regions governing decision making, attention, and problem-solving, and may perceive real and fake websites differently. Neupane et al., in their other study [36], used EEG and eye-tracking to understand users neural and gaze metrics during phishing detection and malware warnings tasks in near-realistic environment. They reported that users may not heed the key areas of the website and may exhibit some differences while processing real and fake websites. In this paper, we corroborate several of the findings of these prior studies, and extend upon them in many significant ways by examining the aspect of neural differences in the processing of real vs. fake websites more closely following a focused study design, and by discussing the potential for automated detection based on such neural differences.

In another study, Vance et al. [50] reported that EEG based measure of users risk-taking behavior in Iowa Gambling Task could predict their security warning performance. Further, Anderson et al. [8] argued that polymorphic warnings can reduce the effect of warning habituation in their fMRI and mouse-tracking study. Martinovic et al [34] used EEG signal as a side-channel attack to reveal users private information, e.g., their locations, PIN codes, bank ATMs, and familiarity of people. This last study is relevant to the offensive website familiarity detection approach we are suggesting in this paper, but it used a different methodology (EEG-based Brain Computer Interface) compared to our work (fNIRS).

3. STUDY DESIGN & DATA COLLECTION

In this section, we describe the design of our experimental tasks to study website legitimacy (phishing) and familiarity detection, the study set-up involving fNIRS, and the protocol we followed for data collection with human participants.

3.1 Design of the Task

We designed an experimental task to investigate website legitimacy and familiarity detection based on neural mechanisms. To this end, we selected e-commerce, social networking, banking, web email and online storage websites drawn from the list of top fifty popular websites ranked by Alexa [1]. Out of Alexa top 50 websites, we selected 30 websites as unique real websites (denoted as "Real") in our experiment. Some of these websites (e.g., dropbox and facebook) were repeated in their fake versions in the experiment. We created spoofed or fake versions (denoted as "Fake") of the websites to emulate the phishing attack scenario by either modifying the logo, layout or URL, or combinations thereof. To build easy fake websites (denoted as "easy fake (EFake)") we used shortened URLs, IP addresses and completely different URLs. Also, we changed or removed website logo, or even modified website layout (e.g., by removing background image or css file). To create difficult fake websites (denoted as "difficult fake (DFake)"), we kept the layout intact and modified the URL using obfuscation techniques such as adding extra characters, replacing similar looking characters, and replacing website extension. EFake websites were assumed to be easier to detect while DFake websites were assumed to be harder to detect. In creating obfuscated URLs, we were inspired by real-life phishing URLs obtained from phishtank.com [3]. In total, we used sixty (30 Real, 15 EFake and 15 DFake) websites in this task. The EFake and DFake websites were hosted in our local webserver. An in-house software was developed to execute the task. For each of the Real website used in the task, at the end of our study, the participants were requested to provide their familiarity level as "familiar" and "unfamiliar". This information was later utilized in the familiarity detection analysis.

The basic design of this task is similar to previous phishing studies [17, 36, 37], except that we add a familiarity vs. unfamiliarity aspect to it. We used an event related design [42] for the task, where each trial is presented as an event with certain inter-stimulus interval. Longer trials are needed for fNIRS studies because it takes approximately 6-8 seconds for blood flow changes to reach the maximum value, similar to fMRI. In our task, a website was presented for 10 seconds, followed by a response page for another 10 seconds in which the participants had to answer whether they trusted the website by pressing yes/no button using mouse. Unlike [17, 36, 37], we did not explicitly ask the participants if the website was real or fake since our goal was to capture neural signals associated with implicit detection of real and fake websites in line with real-world conditions. The process was repeated for 60 trials with inter-stimulus interval (fixation) of 6 seconds. After every six trials, a rest event was presented for 12 seconds. During the fixation and rest events, the participants were instructed to relax. The websites in the tasks were randomly presented to the participants in "Firefox" browser. The total length of the experiment was around 31 minutes. During the trials, the participants were instructed to make as little head movement as possible. The goal of this experiment was to measure the participants' neural activity during the task. The sample stimuli and timing flow diagram of the task are shown in figure 2.

3.2 Repeated Measures and Experimental Set-up

In our study, we used an fNIRS device with 46 channels (18 sources and 15 detectors) developed by Hitachi Medical (ETG 4000). The frequency of the device was set to 10Hz and the inter-optode distance was set to 30mm.

We used the results from a previous fMRI study of phishing detection by Neupane et al. [37] for the purpose of fNIRS sensor configuration. Neupane et al. [37] had found that when participants identified websites as fake (contrasted with real), they activated right middle, inferior, and orbital frontal gyri, and left inferior parietal lobule. And, when participants identified real websites, they showed increased activity in left precentral gyrus, right cerebellum, left cingulate gyrus, and occipital cortex. So we created a custom probe design to cover the frontal cortex, right temporoparietal junction (TPJ), and left temporoparietal junction, which overlap with several of the brain regions from the Neupane et al. study. The frontal cortex had 22 channels, and the right and the left TPJ each had 12 channels. We used a comfortable skull cap to hold probes in precise locations on participants' head.

Unlike the fMRI experiment [37], the participants were presented with the task on a computer, which participants performed in an upright position in lab conditions, while their brain data was recorded in the background.

As our repeated measures in the experiment, we recorded changes in the concentration of oxygenated hemoglobin (oxy-Hb) and deoxygenated hemoglobin (deoxy-Hb) through the fNIRS device while the participants performed the tasks.

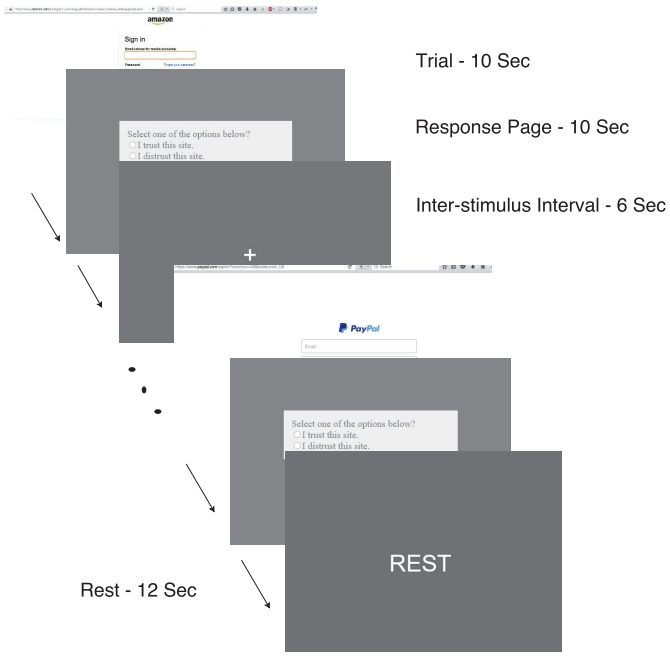

Figure 2: Sample Experimental Stimuli and Timing Flow Diagram.

3.3 Study Protocol

Ethical and Safety Considerations: Our study was approved by the Institutional Review Board (IRB) of our University. The participation in the study was strictly voluntary and it was ensured that the participants were comfortable during the experiment. A signed informed consent was collected from all participants prior to their participation in the study. The standard best practices were followed to protect the confidentiality and privacy of participants' data collected during the study.

Participant Recruitment and Pre-Experiment Phase: Twenty healthy university members (students and faculty members from diverse disciplines, and staff members) were recruited for our study. After providing informed consent, participants provided their demographic information (such as age, gender and education level). Our pool was comprised of 55% male and 45% female, 45% were above the age of 24 and belonged to fairly diverse educational levels (10% High School, 20% Undergrad, 35% Grad, 5% Doctorate, 30% Others). Our sample, especially in terms of age, was closer to the group of users who use the Internet frequently [2], and who are supposedly more vulnerable to phishing attacks [46] and hence are a good target of our study. Previous power analysis studies have found 20 to be an optimal number of participants for such studies. For instance, statistical power analysis of event-related design fMRI studies has demonstrated that 80% of clusters of activation proved reproducible with a sample size of 20 subjects [35]. Our participant sample is well-aligned with the samples used in related prior studies [17, 36, 37]. Also, in line with other prior studies [17, 36, 37], the participants were not told anything regarding the security relevance of the experiments, in order to avoid explicit security priming of the individuals which may impact their neural activity and task performance.

Task Execution Phase: The study followed a *within-subjects* design, whereby all participants were presented with the same set of (randomized) trials corresponding to the task designed to measure website legitimacy and familiarity detection. In the main experiment session, the fNIRS optode or sensor configuration was pre-established for each participant, as described in Section 3.2. Once the probe was placed in the correct location on probe cap, fnirs system was calibrated and a Patriot Polhemus 3d digitizer was used to measure the location of each optode/detector on that participant's head. This allowed for spatial localization of activated brain regions during fNIRS data analysis.

Post-Experiment Phase: During this phase, the participants were given a post-test questionnaire designed to determine their familiarity with the websites used in the experiment. The participants were asked to provide their familiarity with the websites on the scale of familiar and unfamiliar. After the completion of this phase, each participant was paid $10 for their participation in the study.

4. ANALYSIS METHODOLOGY

In this section, we describe the procedures and metrics employed for the analysis of the neural data captured during our experiment.

4.1 Neural Data Analysis

Hitachi's data acquisition software was used to collect and process the raw brain data during the experiment. The raw data was pre-processed to remove high frequency noise and motion artifacts. A low-pass filtering of the data, keeping frequencies between 0.01Hz and 0.5Hz, removed the signatures representing respiratory and cardiac fluctuations. The filtered raw data was then converted to the average concentrations of oxyenated Hemoglobin (oxy-Hb) and deoxygenated Hemoglobin (deoxy-Hb) using modified Beer-Lambert's Law [52]. When hemoglobin transports oxygen,

Table 1: Regions of Interest based on fNIRS channel locations

Area #	Acronym	Regions of Interest
40	RSGWA	Right Supramarginal gyrus part of Wernicke's area
6	RPSMC	Right Pre-Motor and Supplementary Motor Cortex
9	RDLPFC	Right Dorsolateral prefrontal cortex
22	RSTG	Right Superior Temporal Gyrus
21	RMTG	Right Middle Temporal Gyrus
6	LPMSMC	Left Pre-Motor and Supplementary Motor Cortex
40	LSGWA	Left Supramarginal gyrus part of Wernicke's area
9	LDLPFC	Left Dorsolateral prefrontal cortex
1,2,3	PSC	Primary Somatosensory cortex
42	PAAC	Primary and Auditory Association Cortex
21	LMTG	Left Middle Temporal Gyrus
22	LSTG	Left Superior Temporal Gyrus
10	FPA	Frontalpolar area
11	OPA	Orbitofrontal area

it is referred to as oxy-Hb and when it releases oxygen due to an increase in oxygen metabolism, it deoxy-Hb. When brain regions are active, regional cerebral blood flow and oxygenated hemoglobin in that area increases. However, during functional metabolism, oxygen consumption increases, and hence concentration of deoxy-Hb also increases [32]. The increase in oxy-Hb and decrease in deoxy-Hb are associated with activation of the brain area [14]. These hemodynamic changes are similar to the blood oxygenation level dependent (BOLD) signal measured during fMRI scans.

We had synchronized the brain data collected during the experiment with the trial presentation time and order. We designed an in-house software, which took synchronized brain data and trial presentation log, to extract the oxy-Hb and deoxy-Hb changes related to each website. We then computed the average oxy-Hb and average deoxy-Hb from each channel for each website. We wanted to analyze the brain data when participants were observing the websites, so the brain data related to response page was excluded from the analysis.

Hemisphere Analysis: As described in Section 3.2, two probe sets of 12 channels each were placed at right and left TPJ junction, and one probe set of 22 channels was placed at frontal cortex. In hemisphere analysis, we averaged the oxy-Hb and deoxy-Hb values measured from all channels located in each probe for each channel. We then compared the differences in oxy-Hb and deoxy-Hb for real, fake, easy fake and difficult fake websites at each of these three locations. Along with the neural differences in processing Real and Fake websites, we were interested in the differences in EFake and DFake websites, as these websites were substantially distinct.

Regions of Interest Analysis: Our preprocessed data included 46 channels of data, where each channel contained the rate of change in oxy-Hb and deoxy-Hb as measured at that location over time. We used Tsuzuki's 3D-digitizer-free method for the virtual registration of fNIRS channels onto the stereotactic brain coordinate system. Essentially, this method allows us to place a virtual optode holder on the scalp by registering optodes and channels onto reference brains. Assuming that the fNIRS probe is reproducibly set across subjects, the virtual registration can yield as accurate spatial estimation as the probabilistic registration method. So we identified the brain area represented by each channel and grouped these channels based on the majority of the Brodmann Area [11] they cover. The channels were hence grouped into 14 different regions of interest (ROIs) as shown in Table 1. For ROI analysis, we averaged oxy-Hb and deoxy-Hb measured by the channels grouped in these ROIs separately for each website. We then compared the differences in oxy-Hb and deoxy-Hb for real, fake, easy fake and difficult fake websites at each of these 14 ROIs.

4.2 Website Familiarity Analysis

To enable website familiarity analysis, we had showed participants, in the post-experiment phase, the images of the websites used in the experiment and asked them to answer if they were familiar or unfamiliar to these websites. We then grouped all the familiar websites and unfamiliar websites separately and measured differences in oxy-Hb and deoxy-Hb across them. We also measured differences in oxy-Hb and deoxy-Hb for real, fake, easy fake and difficult fake websites, which were familiar to users.

4.3 Statistical Testing

We used Kolmogorov-Smirnov test to measure normality of the data. Our data set was non-normal so we used Friedman's test and Wilcoxon Singed-Rank Test (WSRT) for measuring differences in the means of different groups underlying our analysis. Holm-bonferroni correction was used during post-hoc analysis for multiple comparisons. The effect size of WSRT was calculated using the formula $r = Z/\sqrt{N}$, where Z is the value of the z-statistic and N is the number of observations on which Z is based. The effect size is considered *medium* if it is >0.2 and *large* if it is >0.5.

5. NEURAL ANALYSIS AND RESULTS

For neural analysis, as described in Section 4.1, we performed hemisphere and ROI analysis for all websites, only familiar websites, and familiar vs. unfamiliar websites.

5.1 Website Legitimacy: All Websites

Hemisphere Analysis: In hemisphere analysis, we contrasted different categories of websites for the oxy-Hb and deoxy-Hb values measured from all channels located in each

hemispheres (frontal cortex, right TPJ and left TPJ) for all trials. When first looking at frontal cortext, using Friedman's test, we found statistically significant differences in deoxy-Hb among the different types of trials ($\chi^2(3)=242.7$, $p<.0005$). Further, upon contrasting the deoxy-Hb in Real trials with deoxy-Hb in Fake trials with WSRT, we saw statistically significant difference ($p=.017$) with a medium effect size ($r=.37$). When analyzing the right and left hemispheres, however, we did not find any statistically significant differences across the trials.

Table 2: Statistically Significant Results at ROIs for All Websites: Pairwise Comparisons of Real, EFake, DFake

#	ROI	Comparison	HbType	p-value	Effect Size
1	LSGPWA	DFake >EFake	deoxy-Hb	.012	.39
2	FPA	Real >Fake	deoxy-Hb	.017	.37
		Real >EFake	deoxy-Hb	<.0005	1.92
		Real >DFake	deoxy-Hb	<.0005	1.91
3	OFA	Real >EFake	deoxy-Hb	<.0005	1.37
		Real >DFake	deoxy-Hb	<.0005	1.34

Regions of Interest Analysis: To recall, for ROI analysis, we averaged oxy-Hb and deoxy-Hb measured by the channels grouped in 14 ROIs separately for each trial (Section 4.1). We contrasted the oxy-Hb and deoxy-Hb separately for different trials at each of these ROIs.

Friedman's test showed the presence of statistically significant difference in mean deoxy-Hb among different trials at left supramarginal gyrus part of wernicke's area ($\chi^2(3)=29.96$, $p<.005$). On further using WSRT, we saw statistically significant difference in deoxy-Hb in DFake websites as compared to EFake websites (Table 2, row 1 provides the details).

We also found that the differences in mean deoxy-Hb ($\chi^2(3)=254.9$, $p<.0005$) among different trials at frontopolar area, based on Friedman's test. On further using WSRT, we found several statistically significant pairwise results including: Real and Fake, Real and EFake and Real and DFake (Table 2, row 2 provides the details)

On using Friedman's test, we further noticed differences in mean deoxy-Hb ($\chi^2(3)=254.9$, $p<.0005$) among different trials at the orbitofrontal area. On further contrasting with WSRT deoxy-Hb across different trials, we found statistically significant pairwise results including: Real and EFake and Real and DFake (Table 2, row 3 provides the details)

Interpretation: Decrease in deoxy-Hb signifies increased metabolism and blood flow in the local brain area [14, 12]. For the Fake websites, we found higher activation (lower deoxy-Hb) in orbitofrontal area of participants. Dimoka [18] looked at trust and distrust in the brain by showing ebay seller profiles designed to have varying levels of credibility to participants. Their results suggested that *trust* was associated with lower activation in the orbitofrontal cortex. Fake websites having different URL and logos might have raised *distrust* in the participants brain, resulting in the increased activation in the orbitofrontal cortex we observed. The activation of frontopolar area, which is implicated in *working memory* and *cognitive workload* [9], suggesting that participants experienced higher cognitive load when trying to assess the validity of the fake websites when compared to the real sites. Our results are also in line with the findings of the previous phishing detection studies. Neupane et al. [37] in their fMRI study of phishing detection found higher activation in frontal and left parietal brain corresponding to fake websites. Establishing similar results in a more naturalistic environment (fNIRS vs. fMRI) suggests that deep-down neural patterns may persist irrespective of the environment. The fMRI study of real and fake Rembrandt paintings by Huang et al. [28] had also found higher activation in right middle frontal gyrus corresponding to fake paintings when contrasted with real paintings. The EEG-based study of Neupane et al. [36] also revealed similar differences, but were only limited to the distraction level between real vs. difficult fake websites. Overall, our analysis demonstrates the existence of marked neural differences in the processing of real and fake websites in key areas of the brain.

5.2 Website Legitimacy: Familiar Websites

In this analysis, we collected the websites familiar to the participants and contrasted corresponding oxy-Hb and deoxy-Hb for different trials at left TPJ, right TPJ and frontal cortex, and the 14 predefined regions of interest separately.

Hemisphere Analysis: In hemisphere analysis, we noticed statistically significant differences in deoxy-Hb ($\chi^2(3)=53.69$, $p<.0005$ for different trials at right TPJ, on using Friedman's test. On further contrasting deoxy-Hb among different trials using WSRT, we found statistically significantly higher deoxy-Hb in Real trials as compared to Fake trials $p<.0005$ with a large effect size ($r=.76$), EFake trials $p<.0005$ with a large effect size ($r=.76$) and DFake trials $p=.002$ with a medium effect size ($r=.44$).

Table 3: Statistically Significant Results at ROIs for Familiar Websites: Pairwise Comparisons of Real, EFake, DFake

#	ROI	Comparison	HbType	p-value	Effect Size
1	RDLPFC	Real >EFake	oxy-Hb	.003	.45
		Real >DFake	oxy-Hb	.011	.39
		Real >EFake	deoxy-Hb	<.0005	.93
		Real >DFake	deoxy-Hb	.002	.49
2	OFA	Real >EFake	deoxy-Hb	<.005	.78
		Real >DFake	deoxy-Hb	<.0005	.76
3	RSTG	Real>DFake	deoxy-Hb	<.0005	.63
		DFake >EFake	deoxy-Hb	.001	.50
4	RMTG	DFake >EFake	deoxy-Hb	.006	.42
5	LDLPFC	EFake >DFake	deoxy-Hb	.006	.49
6	LMTG	DFake >EFake	deoxy-Hb	.007	.42

Region of Interest Analysis: In the ROI analysis, we observed statistically significant difference in oxy-Hb ($\chi^2(3)=8.03$, $p<.035$) and deoxy-Hb ($\chi^2(3)=36.5$, $p<.0005$) among different trials at right dorsolateral prefrontal cortex on using Friedman's test. On further using WSRT, we found statistically significantly higher oxy-Hb and deoxy-Hb in Real trials as compared to EFake trials and DFake trials (Table 3, row 1). Similarly, we noticed, on using Friedman's test, statistically significant differences in deoxy-Hb ($\chi^2(3)=19.46$, $p<.0005$) among different trials at Orbitofrontal area. Table 3, row 2 depicts the corresponding pairwise statistically significant comparisons on using WSRT between Real and EFake trial, and Real and DFake trial.

At right superior temporal gyrus also, we noticed statistically significant difference in deoxy-Hb ($\chi^2(3)=18.11$, $p<.0005$) on using Friedman's test. We observed statistically significant differences on using WSRT for Real and DFake, and DFake and EFake trials (Table 3, row 3). We

Table 4: Statistically Significant Results at ROIs: Familiar (Fam) vs Unfamiliar (UnFam) Websites

#	ROI	Comparison	HbType	p-value	Effect Size
1	RDLPFC	Fam >UnFam	oxy-Hb	.007	.33
2	RSTG	Fam >UnFam	deoxy-Hb	.002	.49
3	RMTG	Fam >UnFam	deoxy-Hb	.013	.49

further noticed statistically significant difference in deoxy-Hb at right middle temporal gyrus ($\chi^2(3)=44.34$, $p<.0005$) on using Friedman's test. On using WSRT, we found statistically significant pairwise difference between DFake and EFake trials (Table 3, row 4).

Next, on using Friedman's test at left dorsolateral prefrontal cortex, we found statistical difference in deoxy-Hb among different trials ($\chi^2(3)=52.45$, $p<.0005$). We observed statistically significant differences on using WSRT between EFake and DFake trials (Table 3, row 5). Finally, at left middle temporal gyrus, we found that statistically significant difference exist in deoxy-Hb ($\chi^2(3)=15.62$, $p=.001$) among different trials, based on Friedman's test. Table 3, row 6 depicts statistically significant results between DFake and EFake upon using WSRT. No other statistically significant differences were found at these ROIs.

Interpretation: Dorsolateral prefrontal cortex (DLPFC), frontopolar cortex, and orbitofrontal cortex, are key neural regions, all of which interact together to play a critical role in decision making [20, 24, 38]. In particular, they have been implicated in making evaluation of trustworthiness [30, 41, 18, 31, 53]. The activation in these areas during the task suggests active participation of participants in our study. Left and right temporoparietal junctions are involved in making trustworthy decisions [53]. The activation of these areas shows that the users were actively involved in making important decisions on the *trustworthiness* of these websites. In ROI analysis, we saw decreased activation in orbitofrontal area during real websites suggesting the trust participants had in the real websites [53]. Increased activation in right dorsolateral prefrontal cortex implicates the use of *working memory* in decision making process [9]. Superior temporal gyrus is found to be activated during visual search and switching of choices by users [40, 23]. Middle temporal gyrus is associated with language processing and semantic memory [39]. The higher activation in these areas for fake websites suggests the level of suspicion users may have had while deciding upon the trustworthiness of the websites [13, 10].

5.3 Website Familiarity

We grouped the familiar and unfamiliar "real" websites, and analyzed the neural activity corresponding to them. We wanted to understand the changes in neural metrics (oxy-Hb and deoxy-Hb) when users were viewing websites familiar and unfamiliar to them.

Hemisphere Analysis: We tested for the difference in oxy-Hb and deoxy-Hb between familiar and unfamiliar trials at left TPJ, right TPJ and frontal cortex using WSRT. We observed that at Right TPJ, mean oxy-Hb ($p=.028$) with a medium effect size ($r=.31$)and deoxy-Hb ($p=.031$) with a medium effect size ($r=.33$) for familiar trials were statistically significantly higher as compared to unfamiliar trials.

Region of Interest Analysis: We used WSRT to contrast oxy-Hb and deoxy-Hb among familiar and unfamiliar trials at 14 different ROIs. At right dorsolateral prefrontal cortex, we noticed statistically significantly higher oxy-Hb for familiar trials than unfamiliar trials, upon using WSRT (Table 4, row 1). On contrasting deoxy-Hb in familiar trials vs. unfamiliar trials, at right superior temporal gyrus, we saw a statistically significant difference (Table 4, row 2). At middle temporal gyrus, on contrasting deoxy-Hb between familiar and unfamiliar trials, we noticed statistically significantly higher deoxy-Hb for familiar trials as compared to unfamiliar trials (Table 4, row 3).

Interpretation: Dorsolateral prefrontal cortex [15] is found to be associated with working memory and executive functions. The activation in dorsolateral prefrontal cortex during the task might be related to the decision making aspect of the task. The activation of superior temporal gyrus is related to visual attention [45, 23]. The left and right temporoparietal junction and middle temporal gyrus are found to be activated for familiar faces in previous neuroscience studies [49, 51, 33].Temporoparietal junction activations are also related to *long-term memory retrieval* [33]. Familiar websites being in long-term memory might have activated these areas in users. These results clearly indicate strong differences in the way people process familiar websites compared to unfamiliar websites.

6. DISCUSSION AND FUTURE WORK

In this section, we summarize and further discuss the main findings from our study. We also outline the strengths and limitations of our study.

6.1 Neural Underpinnings

The users showed increased brain activity in many areas of the brain associated with evaluating trustworthiness, decision-making, working-memory and visual search while deciding about the legitimacy or trustworthiness of websites presented to them. Our neural analysis shows extra activation in frontopolar area for fake websites, which is implicated in working memory and cognitive workload, suggesting that participants experienced higher cognitive load when trying to assess the validity of the fake websites when compared to the real sites. Our neuroimaging results also depict decreased activation in orbitofrontal area during real websites, suggesting the level of trust participants possessed in the real websites. Overall, there were concrete differences in the activation of key brain areas when participants were viewing real and fake websites, indicating that users are processing real and fake websites differently. Our work is well-aligned with a prior fMRI-based study [37], but highlights the existence of real vs. fake neural signatures using a different neuroimaging technique (fNIRS) which facilitated testing in a much more realistic scenario.

Another important aspect of our study was comparison of brain signals for familiar and unfamiliar websites. Our analysis in this regard demonstrated that familiar websites triggered significantly higher activity in regions primarily shown to be activated when users retrieve words or pictures from long-term memory. Observing this level of activity difference in familiar vs. unfamiliar websites is interesting, which sheds light as to how users' memory may be playing an important role in the context of web browsing.

6.2 Potential Defense Mechanisms

The current automated phishing detection schemes like blacklists are ineffective to protect users in real-time, since 47% - 83% of phishing websites appear on blacklists about twelve hours after the initial test [47], while the median lifetime of phishing websites is just few hours [6]. In this light, there is a need for a new design to detect and prevent phishing attacks. We explored the potential automated detection of phishing websites based on these characteristic differences in neural activities when viewing real and fake websites.

For each of the 46 channels, we processed the raw data (as described in Section 4) and converted to oxy-Hb and deoxy-Hb levels. We then normalized the oxy-Hb and deoxy-Hb data in each channel using z-score. Next, for each channel we computed several features, including standard deviation, slope, average, max, and min. We computed these features separately for the first and second half of the time series data corresponding to each 10-second long task of viewing a website trial, which occurred for 30 real websites and 30 fake websites (15 easy fake and 15 difficult fake websites). We also separated the data into 5 equal segments of data, and we took the average value of each of these segments for the oxy-Hb and deoxy-Hb datastreams.

Considering that the phishing attacks are generally personalized and are mostly launched on websites familiar to users, we built classification models only for websites familiar to participants. We then used 10-fold cross validation for estimation of the classification models built on eight different machine learning algorithms, namely, K Nearest Neighbor (with different number of nearest neighbors in classification - KNN-0, KNN-3, KNN-5, KNN-10), Support Vector Machine (SVM), SVM with Polynomial SVM Kernel, Naive Bayes (NB), and decision tree, following the methodology similar to prior research [25]. Previous studies have reported that fNIRS neural signals are unique for each user [44]. Taking this into account, we built a unique classifier model for each user. We tested these models with all sets of features (size n) and three different subsets (size $n/3$, $n/10$, and $n/100$) of selected features based on information gain [16].

We tested our models using two metrics: Accuracy and Area under the Curve (AUC). Accuracy represents the ratio of the total number of correctly identified instances to the total number of instances present in the classification model. Area under the curve (AUC) is the probability of correctly identifying which of the two stimuli is "real" and which is "fake" [26]. AUC is measured on the scale of 0-1 and higher value of AUC represents higher true positive rate and lower false positive rate [21]. A random classifier has an accuracy of 0.5 and an area under the curve of 0.5, while a perfect classifier (with no classification errors) has an accuracy of 1 and an area under the curve of 1. Our models achieved an average accuracy of 76% (AUC of 73%) for real and fake website classification. These results are significantly better than a random guessing model.

These results demonstrate that fNIRS-based neural data can be used in the development of an automated phishing detection tool, and provide foundation for building new mechanisms based on neural cues. This approach does not necessarily compete with other approaches, but can rather be seamlessly used in conjunction with other approaches and may serve to add another layer of security against phishing attacks. We believe that these are avenues for exciting future research.

6.3 Potential Attack Mechanisms

Several gaming and entertainment applications based on brain control are currently commercially available. These applications have unrestricted access to brain signals measured by such brainwave devices (applicable to BCI, EEG or fNIRs). Such an application may be malicious in nature, and the attackers may perform privacy attacks against the users based on their neural data.

Our study shows that there are differences in neural activities when people are viewing familiar websites and unfamiliar websites. Using such neural data, we studied the feasibility of building a classification model which can detect websites familiar to users versus those that are unfamiliar to them following the technique implemented in the potential defense mechanisms (previous subsection). Our classification model with the fNIRS data reveals familiarity vs. unfamiliarity of users to given websites with average accuracy of 73% (AUC is 77%). These results are promising. Related to this study, Martinovic et al. [34] used P300 to reveal 20-43% of personal information for example, bank cards, ATMs by showing them the images of different banks, ATMs, credit cards etc.

Such an offensive approach may enable attackers to launch targeted phishing attacks against users, and extract other private information from users' neural activity, for example, familiar faces or voices, familiar places, familiar cuisines, etc. Future work is necessary to study the efficacy of these attacks in practice and to come up with mitigation strategies.

6.4 Strengths and Limitations

We believe that our study has several strengths. We collected data in near realistic environment, where participants interacted with a popular browser and actual websites. This is in contrast to prior fMRI studies [37, 8], which only worked with static snapshots of websites. Our study design was closely focused at the web browsing scenario and was explicitly tailored for website legitimacy and familiarity detection. This allowed us to subject the participants to multiple trials without causing fatigue-related biases. In contrast, prior studies [37, 36] involved multiple security tasks (phishing and warnings) in a single session.

In line with any study involving human subjects, our study has certain limitations too. Our sample may not be representative of a wider population, although it is representative of active computer users and phishing-prone populations. The study was conducted in a lab-based environment. Although we tried to imitate the real-world scenario of web-browsing and computer usage in our study, the participants' performance might have been affected just due to the fact that their neural activity were being monitored. The fNIRS probe cap we used was light-weight, designed for flexibility and comfort to perform tasks seamlessly. Still, it may have caused some discomfort to some of the participants, which may have had an impact on the data collected. In our study, we presented sixty different websites to each participant in a span of around thirty minutes, which may not match the quantity of websites people usually browse in real-life within this time-frame. This represents a limitation of every neurophysiological study, as it needs repeated trials to establish a high signal-to-noise ratio. Future studies might be needed to investigate a more realistic task.

7. CONCLUDING REMARKS

In this paper, we presented an fNIRS study which investigated neural mechanics underlying website legitimacy and website familiarity detection. We found significant differences in neural activity in many key brain regions when users were processing real and fake websites. We also found differences in neural processes when users were subjected to familiar and unfamiliar websites. We discussed how the insights drawn from our study offers a potential for building a new line of defensive and offensive mechanisms based on neural mechanisms.

Acknowledgments

The authors thank: Jeremy Michael Bitten and Sarah Elaine Bratt for their participation in data collection. We would also like to thank Rajesh Kana, Maliheh Shirvanian, and Sagar Thapaliya for their feedback on a previous draft version of this paper, and WWW'17 anonymous reviewers for their constructive input.

8. REFERENCES

[1] Alexa. `http://www.alexa.com`. [5-15-2016].

[2] Internet Users. `http://www.pewinternet.org/data-trend/internet-use/latest-stats/`. [19-05-2016].

[3] Phishtank. `http://www.phishtank.com/`. [19-05-2016].

[4] Portalite fnirs system. `http://www.artinis.com/portalite/`. [7-28-2016].

[5] Portamon wireless fnirs system. `http://www.artinis.com/portamon/`. [7-28-2016].

[6] G. Aarin and R. Rasmussen. Global phishing survey 1h2014: Trends and domain name use. *Technical Report 1H2014, APWG*, 2014.

[7] D. Akhawe and A. P. Felt. Alice in warningland: A large-scale field study of browser security warning effectiveness. In *USENIX Security Symposium'13*.

[8] B. B. Anderson, C. B. Kirwan, J. L. Jenkins, D. Eargle, S. Howard, and A. Vance. How polymorphic warnings reduce habituation in the brain: Insights from an fMRI study. In *Conference on Human Factors in Computing Systems,*, 2015.

[9] J. V. Baldo and N. F. Dronkers. The role of inferior parietal and inferior frontal cortex in working memory. *Neuropsychology*, 20(5):529, 2006.

[10] M. A. Bhatt, T. Lohrenz, C. F. Camerer, and P. R. Montague. Distinct contributions of the amygdala and parahippocampal gyrus to suspicion in a repeated bargaining game. *Proceedings of the National Academy of Sciences*, 109(22):8728–8733, 2012.

[11] K. Brodmann. *Brodmann's: Localisation in the cerebral cortex*. Springer Science & Business Media, 2007.

[12] R. B. Buxton, K. Uludağ, D. J. Dubowitz, and T. T. Liu. Modeling the hemodynamic response to brain activation. *Neuroimage*, 23:S220–S233, 2004.

[13] A. W. Craig, Y. K. Loureiro, S. Wood, and J. M. Vendemia. Suspicious minds: Exploring neural processes during exposure to deceptive advertising. *Journal of Marketing Research*, 49(3):361–372, 2012.

[14] X. Cui, S. Bray, and A. L. Reiss. Functional near infrared spectroscopy (nirs) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *Neuroimage*, 49(4):3039–3046, 2010.

[15] C. E. Curtis and M. D'Esposito. Persistent activity in the prefrontal cortex during working memory. *Trends in cognitive sciences*, 7(9):415–423, 2003.

[16] J. Demšar, T. Curk, A. Erjavec, Č. Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, et al. Orange: data mining toolbox in python. *The Journal of Machine Learning Research*, 14(1):2349–2353, 2013.

[17] R. Dhamija, J. D. Tygar, and M. Hearst. Why phishing works. In *Conference on Human Factors in Computing Systems*, 2006.

[18] A. Dimoka. What does the brain tell us about trust and distrust? evidence from a functional neuroimaging study. *Mis Quarterly*, pages 373–396, 2010.

[19] S. Egelman, L. F. Cranor, and J. Hong. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *Conference on Human Factors in Computing Systems*, 2008.

[20] A. Etkin, T. Egner, and R. Kalisch. Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in cognitive sciences*, 15(2):85–93, 2011.

[21] T. Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[22] B. Friedman, D. Hurley, D. C. Howe, E. Felten, and H. Nissenbaum. Users' conceptions of web security: A comparative study. In *Extended abstracts on Human factors in computing systems*, 2002.

[23] A. Gharabaghi, M. F. Berger, M. Tatagiba, and H.-O. Karnath. The role of the right superior temporal gyrus in visual search – insights from intraoperative electrical stimulation. *Neuropsychologia*, 44(12):2578–2581, 2006.

[24] A. Golkar, T. B. Lonsdorf, A. Olsson, K. M. Lindstrom, J. Berrebi, P. Fransson, M. Schalling, M. Ingvar, and A. Öhman. Distinct contributions of the dorsolateral prefrontal and orbitofrontal cortex during emotion regulation. *PLoS One*, 7(11):e48107, 2012.

[25] V. Gottemukkula and R. Derakhshani. Classification-guided feature selection for nirs-based bci. In *Neural Engineering (NER), International IEEE/EMBS Conference on*, 2011.

[26] J. A. Hanley and B. J. McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36, 1982.

[27] L. M. Hirshfield, R. Gulotta, S. Hirshfield, S. Hincks, M. Russell, R. Ward, T. Williams, and R. Jacob. This is your brain on interfaces: enhancing usability testing with functional near-infrared spectroscopy. In *Conference on Human Factors in Computing Systems*, 2011.

[28] M. Huang, H. Bridge, M. J. Kemp, and A. J. Parker. Human cortical activity evoked by the assignment of authenticity when viewing works of art. *Frontiers in human neuroscience*, 5, 2011.

[29] K. Izzetoglu, S. Bunce, B. Onaral, K. Pourrezaei, and B. Chance. Functional optical brain imaging using near-infrared during cognitive tasks. *International*

Journal of human-computer interaction, 17(2):211–227, 2004.

[30] B. King-Casas, D. Tomlin, C. Anen, C. F. Camerer, S. R. Quartz, and P. R. Montague. Getting to know you: reputation and trust in a two-person economic exchange. *Science*, 308(5718):78–83, 2005.

[31] F. Krueger, K. McCabe, J. Moll, N. Kriegeskorte, R. Zahn, M. Strenziok, A. Heinecke, and J. Grafman. Neural correlates of trust. *Proceedings of the National Academy of Sciences*, 104(50):20084–20089, 2007.

[32] J. León-Carrión and U. León-Domínguez. Functional near-infrared spectroscopy (fnirs): principles and neuroscientific applications. *Neuroimaging methods. Rijeka, Croatia: InTech (2012): 47-74*, 2012.

[33] C. L. Leveroni, M. Seidenberg, A. R. Mayer, L. A. Mead, J. R. Binder, and S. M. Rao. Neural systems underlying the recognition of familiar and newly learned faces. *The Journal of Neuroscience*, 20(2):878–886, 2000.

[34] I. Martinovic, D. Davies, M. Frank, D. Perito, T. Ros, and D. Song. On the feasibility of side-channel attacks with brain-computer interfaces. In *USENIX Security Symposium*, 2012.

[35] K. Murphy and H. Garavan. An empirical investigation into the number of subjects required for an event-related fmri study. *Neuroimage*, 22(2):879–885, 2004.

[36] A. Neupane, M. L. Rahman, N. Saxena, and L. Hirshfield. A Multimodal Neuro-Physiological Study of Phishing and Malware Warnings. In *ACM Conference on Computer and Communications Security (CCS)*, 2015.

[37] A. Neupane, N. Saxena, K. Kuruvilla, M. Georgescu, and R. Kana. Neural signatures of user-centered security: An fMRI study of phishing, and malware warnings. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2014.

[38] T. A. Niendam, A. R. Laird, K. L. Ray, Y. M. Dean, D. C. Glahn, and C. S. Carter. Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective, & Behavioral Neuroscience*, 12(2):241–268, 2012.

[39] T. Onitsuka, M. E. Shenton, D. F. Salisbury, C. C. Dickey, K. Kasai, S. K. Toner, M. Frumin, R. Kikinis, F. A. Jolesz, and R. W. McCarley. Middle and inferior temporal gyrus gray matter volume abnormalities in chronic schizophrenia: an mri study. *American Journal of Psychiatry*, 2004.

[40] M. P. Paulus, J. S. Feinstein, D. Leland, and A. N. Simmons. Superior temporal gyrus and insula provide response and outcome-dependent information during assessment and action selection in a decision-making situation. *Neuroimage*, 25(2):607–615, 2005.

[41] M. L. Platt and S. A. Huettel. Risky business: the neuroeconomics of decision making under uncertainty. *Nature neuroscience*, 11(4):398–403, 2008.

[42] B. R. Rosen, R. L. Buckner, and A. M. Dale. Event-related functional mri: past, present, and future. *Proceedings of the National Academy of Sciences*, 95(3):773–780, 1998.

[43] S. E. Schechter, R. Dhamija, A. Ozment, and I. Fischer. The emperor's new security indicators. In *IEEE Symposium on Security and Privacy*, 2007.

[44] A. Serwadda, V. V. Phoha, S. Poudel, L. M. Hirshfield, D. Bandara, S. E. Bratt, and M. R. Costa. fnirs: A new modality for brain activity-based biometric authentication. In *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on*, 2015.

[45] K. Shapiro, A. P. Hillstrom, and M. Husain. Control of visuotemporal attention by inferior parietal and superior temporal cortex. *Current Biology*, 12(15):1320–1325, 2002.

[46] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs. Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. In *Conference on Human Factors in Computing Systems*, 2010.

[47] S. Sheng, B. Wardman, G. Warner, L. F. Cranor, J. Hong, and C. Zhang. An empirical analysis of phishing blacklists. In *Conference on Email and Anti-Spam (CEAS)*, 2009.

[48] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F. Cranor. Crying wolf: An empirical study of ssl warning effectiveness. In *USENIX Security Symposium*, 2009.

[49] M. J. Taylor, M. Arsalidou, S. J. Bayless, D. Morris, J. W. Evans, and E. J. Barbeau. Neural correlates of personally familiar faces: parents, partner and own faces. *Human brain mapping*, 30(7):2008–2020, 2009.

[50] A. Vance, B. B. Anderson, C. B. Kirwan, and D. Eargle. 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, 2014.

[51] O. Vartanian, V. Goel, E. Lam, M. Fisher, and J. Granic. Middle temporal gyrus encodes individual differences in perceived facial attractiveness. *Psychology of Aesthetics, Creativity, and the Arts*, 7(1):38, 2013.

[52] A. Villringer and B. Chance. Non-invasive optical spectroscopy and imaging of human brain function. *Trends in neurosciences*, 20(10):435–442, 1997.

[53] M. Watabe, H. Ban, and H. Yamamoto. Judgments about others' trustworthiness: An fmri study. *Letters on Evolutionary Behavioral Science*, 2(2):28–32, 2011.

[54] M. Wu, R. C. Miller, and S. L. Garfinkel. Do security toolbars actually prevent phishing attacks? In *Conference on Human Factors in computing systems*, 2006.