This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2016.2566265, IEEE Transactions on Information Forensics and Security

Neural Markers of Cybersecurity: An fMRI Study of Phishing, and Malware Warnings

Ajaya Neupane, Nitesh Saxena, Jose O Maximo, and Rajesh Kana

[1]***Abstract*— The security of computer systems often relies upon decisions and actions of end users. In this paper, we set out to investigate users' susceptibility to cybercriminal attacks by concentrating at the most fundamental component governing user behavior – the human brain. We introduce a novel neuroscience-based study methodology to inform the design of user-centered security systems as it relates to cybercrime. Specifically, we report on an fMRI study measuring users' security performance and underlying neural activity with respect to two critical security tasks: (1) distinguishing between a legitimate and a phishing website, and (2) heeding security (malware) warnings. We identify neural markers that might be controlling users' performance in these tasks, and establish relationships between brain activity and behavioral performance as well as between users' personality traits and security behavior.**

Our results provide a largely positive perspective on users' capability and performance vis-à-vis these crucial security tasks. *First*, we show that users exhibit significant brain activity in key regions associated with decision-making, attention, and problem-solving (phishing and malware warnings) as well as language comprehension and reading (malware warnings), which means that users are actively engaged in these security tasks. *Second*, we demonstrate that certain individual traits, such as impulsivity measured via an established questionnaire, are associated with a significant negative effect on brain activation in these tasks. *Third*, we discover a high degree of correlation in brain activity (in decision-making regions) across phishing detection and malware warnings tasks, which implies that users' behavior in one task may potentially be predicted by their behavior in the other. *Fourth*, we discover high functional connectivity among the core regions of the brain while users performed the phishing detection task. *Finally*, we discuss the broader impacts and implications of our work on the field of user-centered security, including the domain of security education, targeted security training, and security screening.

I. INTRODUCTION

Computing has become increasingly common in many spheres of users' daily lives. At the same time, the need for securing computer systems has become paramount. To enable secure on-line interactions, actions performed and decisions made by human users need to be factored into system design – a principle sometimes referred to as "human in the loop" [9]. Two such prominent *user-centered security* tasks are: (1) distinguishing between a legitimate and a fake website (*phishing detection task*), and (2) heeding warnings provided by modern browsers when connecting to potentially malicious websites (*malware warnings task*).

The field of user-centered security has received considerable attention recently but is still in its infancy. As such, researchers' understanding of end user performance in real-world security tasks is neither very precise nor very clear. Previous computer lab-based studies focusing on security warnings and security indicators (e.g., [10, 12, 13, 14, 15, 16, 17]) have concluded that users do not perform well at these tasks and may often ignore them. This general wisdom has been called into question however by a large-scale field study of browsers' tasks relating to phishing, SSL and malware warnings [11] which showed a high likelihood users actually heeded the warnings they received.

User attitudes, perceptions, acceptance and use of information technology have been long-standing issues since the early days of computing. Users' personal characteristics are also identified as one of the important factors affecting phishing detection and malware warnings interactions (e.g., [60, 61, 62, 63]). In this light, it is important to understand users' behavior and personality characteristics pertaining to the execution of security tasks, and users' potential susceptibility to attacks.

Our goal in this paper was to enhance current knowledge in, and address fundamental questions pertaining to, user-centered security from a *neuropsychological* standpoint. The primary questions driving our research included: (1) what brain regions are activated and functionally connected while performing security tasks?; (2) how well do users perform these tasks?; (3) do certain personality traits (like impulsivity, or attention control) influence users' security behavior and performance?; and (4) are users' behavior in one security task related to their behavior in another.

To answer these inquiries, we developed a novel methodology for studying user-centered security that involves *neuroimaging*. Using this general methodology, our overarching goal was to delineate the nature of cognitive and neural processes that underlie user-centered security decisions and actions. This specific goal was achieved via fMRI (functional Magnetic Resonance Imaging) scanning. fMRI is a Blood Oxygen Level Dependent function measure, and is derived from a combination of stimulus-induced changes in the local cerebral blood flow, local blood volume, and local oxygen consumption rate [5,6]. fMRI provides a unique opportunity to examine in-vivo brain responses mediating user decisions during human-computer security interactions. As a first line of investigation into our novel methodology, our

[1] Ajaya Neupane (Email: aneupane@uab.edu) and Nitesh Saxena (Email: saxena@uab.edu) are currently associated with Department of Computer and Information Sciences; and Jose O Maximo (Email: omaximo@uab.edu) and Rajesh Kana (Email: rkana@uab.edu) are associated with Department of Psychology at University of Alabama at Birmingham.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2016.2566265, IEEE Transactions on Information Forensics and Security

fMRI-based study sheds light on end users' behavior and performance with respect to the important tasks of *phishing detection* and responding to *malware warnings*.

Contributions: Our main contributions in this paper are summarized as follows:

1. Novel Methodology to Study User-Centered Security: We propose a new methodology for studying neurological patterns governing users' performance and behavior with respect to user-centered security tasks.

2. fMRI Study of Phishing, and Malware Warnings: As a specific use case of our methodology, we designed and developed in-scanner fMRI experiments for phishing detection and malware warnings tasks (*Section III*), and conducted a user study by recruiting and scanning 25 individuals performing these tasks. (*Section IV*)

3. Comprehensive Neural and Behavioral Analysis: We provide a comprehensive analysis of neuroimaging and behavioral data, not only evaluating the phishing and malware warnings experiments independently but also contrasting them with each other. We also perform *functional connectivity analysis* to identify the interaction among different brain regions corresponding to tasks relating to phishing detection and responding to malware warnings. (*Section V-VIII)*

This paper is an extension and consolidation of our NDSS 2014 paper [7]. From our previous analysis, we identified the regions of interest (ROI) -- brain areas activated when completing tasks in phishing detection and control, and responding to malware warnings. In our extension, we systematically investigated the functional connectivity among these ROIs (see Section VI). We performed (1) *whole-brain analysis*, where the functional connectivity of one ROI with the rest of the brain was examined; (2) *region of interest analysis*, in which we examined functional connectivity among ROIs, and (3) *brain-behavior analysis*, which examined the functional connectivity of each ROI and impulsivity as a co-variate. We found strong functional connectivity in the phishing detection task compared to the phishing control task. This result confirms findings of our original analysis. The stronger level of functional connectivity suggests greater coordination among brain areas while identifying phishing websites. We did not find any statistically significant results during analyses of responses to malware warnings, however.

Finally, we discuss the broader impacts and implications of our work for the field of user-centered security, including the domain of security education, targeted security training, and security screening. (*Section VIII*)

II. Related Work

Our study centers on phishing detection and malware warnings. Most closely relevant to the phishing component of our study is the lab study reported by Dhamija et al. [10] with 22 participants who were asked to distinguish between real and fake websites. Results indicated that users do not do well at this task as they made incorrect choices 40% of the time. Our behavioral data yielded similar results. However, our neuroimaging data show that users exhibited significant brain activation during the fake or real website identification task. This suggests that although the outcome of participants' efforts to differentiate between fake and real websites may not be good (perhaps because they did not know what to look for on the sites to make a decision), they seemed to be undertaking considerable effort in solving the puzzles as reflected by activity in appropriate brain regions during the decision-making process.

A recent large scale field study reported by Akhawe and Felt [11] used modern browsers' telemetry frameworks to record users' real-world behavior when interacting with malware, as well as phishing and SSL, warnings. Unlike previously conducted lab-based studies of security warnings and security indicators (see below), this study demonstrated that users heeded warnings most of the time. Specifically, Akhawe and Felt found that users ignored Chrome's and Firefox's phishing and malware warnings between 9% and 23% of the time, and ignored Firefox's SSL warnings 33% of the time. These results are very much in line with results of our study, which provides neurological proof of users' ability to process and heed malware warnings.

For over a decade, many lab studies have focused on different browser security indicators (passive indicators, and active warnings for phishing and SSL attacks) [12, 13, 14, 15, 16, 17]. All of these studies suggested that users seldom act upon warnings and security indicators. (We refer to Akhawe and Felt [11] who provide an excellent survey of the results of these studies). Akhawe and Felt [11] attributed the stark difference in the results of prior lab studies focusing on warnings, and their own field study mainly to changes in the nature of browser warnings.

Users' personal characteristics are also identified as one of the important factors affecting their susceptibility to phishing attacks [60, 61, 62, 63]. Viswanathan et al. [59] argued that different attributes of email messages such as source, body content, attention to urgency, attention to title, computer self-efficacy, and amount of emails received, affect detection of phishing emails. The Communication-Human Information Processing model proposed by Wogalter [60] defines the sequence of warnings effect, and assumes attention, memory, attitudes, motivation and behavior as several factors affecting it. The information processing model process studied by Mayhorn et al. [61] showed that personality factors like impulsivity, trust/distrust, anxiety, and calmness measured using standard questionnaires, affect detection of phishing emails. Pattison et al. [62] found that less impulsive individuals are better at identifying and managing phishing emails. Both of these studies used a role-based method [48] to study phishing detection. Wogalter and Mayhorn [63] discussed the need to tailor warnings to accommodate differences in individual characteristics, situations, experience, and skill level. In our study, we wanted to see how neural

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2016.2566265, IEEE Transactions on Information Forensics and Security

responses of users with different personal characteristics differ while identifying phishing websites.

A previous neuroimaging study somewhat relevant to our work was performed by Craig et al. [18]. This study aimed at understanding users' behavior when viewing advertisements, including the level of suspicion aroused by deceptive advertising. Their study found activation of the precuneus and superior temporal sulcus brain regions while participants processed different levels of deceptive stimuli. This has relevance to user-centered online security interactions, as users may become suspicious when they encounter phishing sites or connect to malware-prone websites. While Craig et al. point to the cognitive dangers associated with moderately deceptive materials, our phishing task presented participants with a "real life" online security scenario where they had to determine whether the website was malicious or real.

There have been other studies that applied neuroscience principles to computer security problems, [19, 20, 52, 53]. Bojinov et al. [19] proposed a neuroscience-inspired approach to coercion-resistant authentication. Thorpe et al. [52], and Chung et al. [53] explored user authentication using EEG devices. Martinovic et al. [20] explored the feasibility of side channel attacks with commodity brain-computer interfaces.

TABLE I. SAMPLE LIST OF WEBSITES USED IN THE PHISHING EXPERIMENT

Website	URL
Amazon	http://www.amazon.1click.com/exec/flex-sign-in.com.ch
WellsFargo	www.vvellsfargo.com
eBay	http://91.109.13.183/~ebay/security/
Twitter	https://twitter.login.com
Facebook	http://securitycenter.3dn.ru/facebook/warning/account/suspend/index.html
Gmail	https://accounts-google.com/servicelogin?service=mail

III. DESIGN OF EXPERIMENTS

Our phishing detection and malware warnings tasks were implemented using *E-Prime* software (Psychology Software Tools Inc., Pittsburgh) [2].

A. Phishing Detection and Phishing Control

Phishing is the act of deceiving people by presenting a fake website that resembles a real one. For this experiment, we identified popular websites and took snapshots of the sites' login pages. We then modified the login pages to create fraudulent replications and took snapshots of them as well. The snapshots were then categorized into two types: "real" and "fake." The fake website snapshots were further divided into two categories: "easy" and "difficult." The "easy" sites were those for which we modified both the URL and the logo of the companies; keeping the layout of the webpages intact; or we changed the URL of the webpages to an IP address. The "difficult" sites were those for which we modified just the URL keeping the security icons and parameters intact. Table I provides a sample list of the websites used in the experiment along with their URLs (we obtained some of the URLs from the website www.phishtank.com). The design of fake websites, for this experiment, was similar to the design adopted in the previous study on phishing detection reported by Dhamija et al. [10]. Figure 1 provides a sample of a fake website.

Fig 1: Sample Easy Fake (logo and URL different compared to real)

1) Experiment Design (Phishing): The phishing experiment followed an event-related (ER) design. In an ER design, each trial is presented as an event with longer inter-trial-interval as a recovery time is needed for the hemodynamic response to decline between trials. This was done with the goal of isolating fMRI response to each item separately. ER designs allow different trials to be presented in random sequences, eliminating potential confounds such as habituation, anticipation, set, or other strategy effects [51]. In this experiment, we had 39 trials (12 easy fake, 13 difficult fake and 14 real), out of which 3 trials (1 difficult fake and 2 real) presented at the beginning of the experiment, were considered as practice trials to familiarize the subjects with the task. The following instruction was given to the participants: "*In this experiment, you will see several websites. You have to respond whether the website is real or fake via the response page*".

The experiment also had a fixation baseline condition, each of which lasted for 10s. Fixations, in the context of an fMRI experiment, are short blocks of time when the participants are asked to look at a cross on the screen and relax. They are considered as windows of baseline brain activity. Each trial displayed a website snapshot for 6s followed by a gap of 6s. The experiment started with the set of instructions followed by a fixation for 10s, and after every 6 trials, a fixation of 10s was displayed on the screen. Thus, in total, there were 7 fixations and 39 trials and the experiment lasted for 553s. The trials were presented to each participant in a randomized order and the participants had to express whether the site depicted in the snapshot was "real" or "fake" by pressing the designated joystick button. We recorded the response given by users and the corresponding response time.

(2) Experiment Design (Phishing Control): The phishing control experiment was designed as a control for the stimuli presented in the phishing experiment. This experiment was identical to the phishing experiment, except that participants were instructed to just look at the images displayed on the screen, and not to engage in an active task. Thus, this experiment had all the visual demands of the phishing

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2016.2566265, IEEE Transactions on Information Forensics and Security

experiment except for the decision-making (real or fake website) aspect.

In this experiment, 20 snapshots of login pages of different websites, including: Citibank, USPS, Orkut, hi5, 6pm.com, Google, BankofAmerica, LinkedIn, Chase, Instagram, Coupons, Spotify, onlineshoes, Hotmail, BestBuy, Yahoo, Discover, AT&T, and Apple and a portal for our university, were shown to participants. We used different websites than those used in the phishing detection task, as we did not want to influence participants' decisions of real-fake identification based on the websites they had seen in this experiment. In total, we had 4 fixations (1 in the beginning of the trials and 3 after every 6 trials) and 20 trials, and the experiment lasted for 268s.

B. Malware Warnings

Malware is software created to obtain unauthorized access to computer resources and collect private information. We wanted to identify the neural patterns when people responded to warnings associated with malware. Modern browsers use these warning mechanisms to alert users in case they visit a likely suspicious website and rely upon users' input to proceed [11]. Our malware warnings experiment consisted of several snapshots of news samples and pop-ups of two types: *non-warnings* and *warnings*. A non-warning pop-up contained casual information or questions in it like, "CNN is a pretty popular news website. We have found that 65% of the people like reading news on CNN. We want to know how you feel about it. Do you like CNN?", and a warning pop-up that contained details about the malware threat. In this way, the non-warning pop-up served as a control condition for the warning pop-up. The article itself served the purpose of a primary task in which the user was engaged. The news samples were collected from popular news websites such as *CNN, BBC, LA Times, ABC News, and Slashdot.org*. We collected news items from major categories at the sites including entertainment, sports, politics, and general news. We recreated the webpages on our own as the fMRI video screen only supports a resolution of 640*480 formatted in Bitmap configuration. This task required that the subject read a series of articles. While reading the articles, they were randomly interrupted by a pop-up asking a specific question (non-warning), or by a pop-up warning (about a malicious threat).

Experiment Design (Malware Warnings): The experiment started with a set of instructions followed by a fixation trial of 10s. After the fixation, the abstract was presented for 10s, followed by a pop-up (warning or non-warning randomly presented) for 6s asking the user if he/she wanted to proceed. If the user chose not to proceed, a blank screen was displayed for 10s; otherwise, a full news article was shown for 10s. Fixation of 10s duration was displayed after each trial. This was an event-based design and the user gave his input of yes/no by pressing the appropriate button on a joystick. We incorporated the malware warnings of popular web browsers like Chrome, Internet Explorer, Opera, and Mozilla [11]. It was difficult to display all the details of warnings shown by these browsers but we kept, to the extent possible, the excerpts similar to the warnings of these browsers (see Figure 2). In total, there were 10 fixations, 20 trials, and the experiment lasted for 751s.

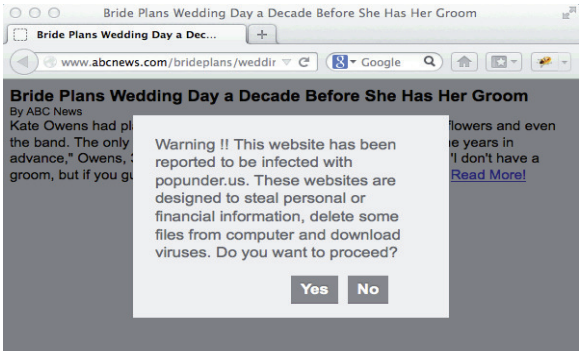


Fig 2: A Snapshot of Warning

C. Our Experimental Set-Up

Throughout the project, fMRI data were acquired using the 3T Siemens Allegra Scanner available to us at the University of Alabama at Birmingham. An MRI compatible IFIS-SA (*Invivo* Corp., Gainesville, FL) auditory and visual system was used for stimulus presentation. However, in our experiments only visual information was presented. This system consists of two computers: one for stimulus presentation and another for experimental control and analysis. A master control unit is used to interface the two computers. We used E-Prime [2] software run on the IFIS-SA system to present visual stimuli. The visual display in the magnet utilizes an IFIS-SA LCD video screen of size 640 * 480 located behind the head-coil that is viewed through a mirror attached to the radio frequency (RF) coil. MRI compatible response boxes (e.g., joysticks and button boxes) are used to receive user responses. The E-Prime IFIS-SA systems record reaction times as well as participant responses to each stimulus item presented and creates data files titled *e-dat* and *t-dat*.

All fMRI tasks followed the same data acquisition protocol, as follows. For structural imaging, initial high resolution T1-weighted scans were acquired using a 160-slice 3D MPRAGE (Magnetization Prepared Rapid Gradient Echo) volume scan with TR = 200 ms, TE = 3.34 ms, flip angle = 1210, FOV = 25.6 cm, 256 x 256 matrix size, and 1 mm slice thickness. For functional imaging, we used a single-shot gradient-recalled echo-planar pulse sequence that offers the advantage of rapid image acquisition (Repetition Time = 1000 ms, Echo Time = 30 ms, flip angle = 60 degrees, Field of View = 24 cm, matrix = 64 x 64). This sequence covers most of the cortex (seventeen 5-mm thick slices with a 1 mm gap) in a single cycle of scanning (1 TR) with an in-plane resolution of 3.75 x 3.75 x 5 mm^3.

IV. Study Procedures

Our fMRI study followed a *within-subjects* design, whereby each participant performed all the three tasks, phishing control, phishing detection, and malware warnings. All tasks were performed in one single fMRI scanning session.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2016.2566265, IEEE Transactions on Information Forensics and Security

In our experiments, only visual stimuli were presented. The study, including participant recruitment and MRI scanning, ran for a period of about 6 months.

A. Ethical and Safety Considerations

Our study was approved by the Institutional Review Board (IRB) at our university. Care was taken to maximize the safety of the participants while being scanned by following standard practices. Their participation in the study was strictly voluntary. They were given the option to withdraw from the study at any point in time. Best practices were followed to protect the confidentiality and privacy of participants' data acquired during the study by de-identifying the collected data.

B. Participant Recruitment & Demographics

Twenty five healthy university students (14 males and 11 females; mean age: 21.5 years) participated in our fMRI study. Participant demographic information is summarized in Table II. The participating students were enrolled in various educational programs, including Biology, Music, Athletics, Psychology, Physical Education, Biomedical Engineering, Mathematics, Medicine, and other programs, resulting in a diverse sample of majors.

TABLE II. PARTICIPANT DEMOGRAPHICS SUMMARY

	N=25
Gender	14 male; 11 female
Age Range	19 – 32 years
Handedness	24 right-handed; 1 left-handed
Race	13 Caucasian; 5 Hispanic; 6 Asian; 1 African American
Non-Native English Speakers	7

Participants were not included if they indicated having metal implanted in their bodies (either surgically or accidentally), indicated they were possibly pregnant or were currently breastfeeding, or indicated having a history of kidney disease, seizure disorder, diabetes, hypertension, anemia, or sickle cell disease. Individuals were also excluded if they were taking psychotropic medications, had claustrophobia, or had hearing problems. Participants were not recruited if they indicated a history of a developmental cognitive disorder, anxiety disorder, schizophrenia, or obsessive-compulsive disorder.

C. Pre-Scanning Phase

The scans were performed at the neuroimaging facility available to us at our university. Participants signed an informed consent form approved by our university's Institutional Review Board. In addition, participants filled out an Edinburgh Handedness form [54], an MRI safety questionnaire, and a Barratt's Impulsivity questionnaire [1]. The purpose of the Edinburgh form was to determine handedness because handedness may relate to the lateralization of hemispheric activity in the participants (right-handed individuals may be more left-lateralized). The purpose of the impulsivity questionnaire was to determine the trait impulsivity level of the participants (details in Section V.B).

Prior to the scan, each participant was shown sample images for both the tasks in the form of images on paper. We also explained that the participant was to use the button response system in the MRI scanner during the tasks. But we did not tell the participants before the fMRI scan as to what they are supposed to be doing in the experiments.

D. Scanning Phase

fMRI data was collected using a Siemens 3.0 T Allegra head-only scanner (as discussed in Section III.C). For each participant, we set the order of the phishing and malware warnings tasks randomly, but always left the phishing control as the first task as we did not want the decision making aspect of the phishing detection task and malware warnings task to affect the phishing control task. We gave appropriate instructions to the participants via an intercom before each experiment started. Instructions were also provided visually on the display screen in the MRI scanner at the beginning of each task. Each task was run through the *IFIS System Manager*.

After the scanning phase was over, we compensated the participant with either course credits or a $50 cash reward, depending on their status.

V. ANALYSIS AND STUDY RESULTS

A. Behavioral Data Analysis

Phishing Detection Experiment: During the phishing experiment, we recorded the response made by the participants and the corresponding response time.

TABLE III: ACCURACY(%) AND RESPONSE TIME (MILLISECOND)

Trials	μ_{acc} (σ_{acc})	μ_{time} (σ_{time})
Real	76.68 (18.84)	3323 (1066)
Fake	46.48 (20.58)	3276 (584)
Easy Fake	56.57 (23.29)	3077 (625)
Difficult Fake	33.98 (23.61)	3538 (645)
All	60.42 (13.99)	3347 (654)

Based on the recorded data, we collected statistics for participant accuracy (acc) and response time (time) for different types of trials (see Table III**).** Accuracy is defined as the fraction of times a particular trial was correctly identified out of the total number of occurrences for that trial.

We observed that, on average across all trials, participants took 3.35 seconds to make their decisions, but their accuracy was only about 60%, only slightly better than a random guess. Prior work by Dhamija et al. [10] reported very similar results based on their computer-based lab study. We used repeated measure ANOVA with Greenhouse-Geisser correction, and determined that the mean response times for real, easy fake and difficult fake trials were statistically significantly different ($F(1.91, 40.20) = 10.14$, $p<.001$). On further analysis using paired t-tests with Bonferroni correction, we found that users spent statistically significantly more time in real websites as compared to easy fake websites ($t(21) = 3.307$, $p=.003$), and in difficult fake websites as compared to easy fake websites ($(t)21) = 4.05$, $p=.001$). Similarly, we found that a statistically significant difference existed among accuracies for these trials ($F(1.92, 40.51) = 48.13$, $p <.001$). On further analysis using paired t-tests with Bonferroni correction, we found statistically

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2016.2566265, IEEE Transactions on Information Forensics and Security

significantly higher accuracy of real websites than fake websites ($t(21) = 7.5$, $p=.000$), easy fake websites ($t(21) = 4.86$, $p=.000$) and difficult fake websites ($t(21) = 9.098$, $p = .000$). We also found statistically significantly higher accuracy of easy fake websites compared to difficult fake websites ($t(21) = 5.44$, $p = .000$).

We did not find statistically significant correlation of phishing detection task performance with users' personality characteristics and gender.

Malware Warnings Experiment: Similar to the phishing experiment, we collected statistics for subjects' accuracy (acc) and response time (time) for the different malware warning conditions (see Table IV**).** Accuracy is defined as the fraction of times a participant pressed "No" for a warning or non-warning condition out of its total number of occurrences.

TABLE IV:ACCURACY(%) AND RESPONSE TIME (MS)

Condition	μ_{acc} (σ_{acc})	μ_{time} (σ_{time})
Non-Warnings	67.49 (26.57)	4228 (664)
Warnings	88.71 (28.62)	3715 (1141)

An important observation is that subjects' accuracy in heeding the warnings was quite high (about 89%), which means that participants paid attention to these warnings and chose not to "click-through" most times. This result is in line with the results from a recent large-scale field study of Akhawe and Felt [11]. It is also validated by the high brain activation in regions associated with language comprehension, visual attention and decision making as shown by our neuroimaging analysis (Section V. B.)

We did not find any statistically significant correlation of users' task performance in the phishing detection and malware warnings tasks.

B. Neuroimaging Data Analysis

All acquired fMRI images were converted from DICOM (Digital Imaging and Communications in Medicine) format to NIFTI (Neuroimaging Informatics Technology Initiative) format using the Free Surfer software (http://surfer.nmr.mgh.harvard.edu/). Data was preprocessed using SPM8 software (Wellcome Trust Centre for Neuroimaging, London, United Kingdom) within MATLAB and an in-house software. Functional data preprocessing started with slice time correction to account for the interleaved pattern of scan slice acquisition. All slices were realigned to the mean image in the scan. All images were then normalized to the EPI template provided by SPM8 using a $2mm^3$ resampling voxel. Head motion was examined in three translational directions x, y, and z, and three rotations: pitch, roll, and yaw. A cut off point of 1 mm in any direction was kept as the criteria for motion. After these quality control measures, data from three participants from the phishing experiment were discarded resulting in 22 usable datasets for that experiment and also for the phishing control experiments. All participants' datasets were used for the malware warnings. Finally, all normalized images were smoothed using a Gaussian filter of 8mm full width half maximum.

Statistical analyses were performed on individual and group data using the General Linear Model (GLM). In GLM analysis, each voxel in the brain will have a signal time-series for a given experiment based on how that voxel behaves in response to a specific task. The GLM formula is $Y = X*\beta+\varepsilon$, where Y is the fMRI signal at various time points at a single voxel, X is several components (the design matrix with different conditions, such as real, fake, or malware) that can explain the observed fMRI signal, β is the parameter that defines the contribution of each component of the design matrix to the value of Y, and ε is the difference between the observed data (Y) and that predicted by the model ($X*\beta$). Group analyses were performed using a random-effects model. Regions of interest (ROIs) with statistically significant activation were identified using a *t*-statistic on a voxel by voxel basis. Separate regressors were created for real, fake, and fixation stimuli in the phishing experiment, and abstract, warning, and no-warning for the malware experiment by convolving a boxcar function with the standard hemodynamic response function as specified in SPM. Statistical maps were superimposed on normalized T1-weighted images. All data were intensity-thresholded at $p=.001$, with a cluster size correction per region for a family wise error (FWE) rate of .05. To determine the voxel threshold for significance, a minimum cluster thresholding operation was performed using the AlphaSim software package in AFNI (Analysis of Functional Neuroimages) [56]. Ten thousand Monte Carlo simulations were generated to maintain the FWE rate at .05 for the whole brain. Thus, for a given region to be considered significantly active, it would need to have a minimum cluster size of $64mm^3$ [21].

TABLE V. ABBREVIATIONS FOR BRAIN REGIONS

Acronym	Brain Region
MPFC	Medial Prefrontal Cortex
RIFG/LIFG	Right/Left Inferior Frontal Gyrus
RMFG/LMFG	Right/Left Middle Frontal Gyrus
ROFC/LOFC	Right/Left Orbitofrontal Cortex
RMTG/LMTG	Right/Left Middle Temporal Gyrus
RSTG /LSTG	Right/Left Superior Temporal Gyrus
RIPL/ LIPL	Right/ Left Inferior Parietal Lobule
ROC/LOC	Right/Left Occipital Cortex
SMA	Supplementary Motor Area

(1) Phishing Control vs Phishing Detection Task

To examine the overlapping and unique activity associated with the phishing task and a visual control task, we compared the phishing with the phishing control experiment using a paired sample t-test. Both tasks elicited significantly increased activity in the visual cortex, perhaps in line with the visual demands of the stimuli ($p < .05$, FWE corr.). However, the phishing task showed significantly greater and unique activation in various brain regions, such as RMFG and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2016.2566265, IEEE Transactions on Information Forensics and Security

bilateral insula (see Figure 3), a pattern not seen in the phishing control experiment ($p < .05$, FWE corr.).

The anterior insula has been implicated in a variety of functions, such as affective and cognitive judgments. Activation in anterior insula, along with MFG, has been associated with making choices [44, 45]. The middle frontal gyrus also has been found to be playing a critical role in cognitive control especially in selecting an appropriate choice of action [46]. The activation of these important decision-making regions of the brain in the phishing experiment (vs. the control experiment) suggested that the participants were conscientiously making an effort as to differentiate "fake" websites from "real" websites.

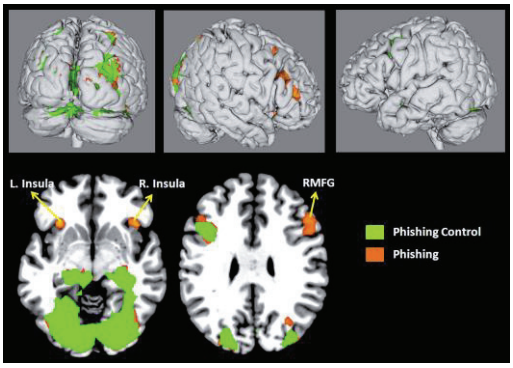


Fig 3: **Phishing vs. Phishing Control Activation**. Both tasks show significant activity in the visual cortex. Phishing shows greater and unique activation in the right middle frontal gyrus (RMFG) and bilateral insula. (The top right corner brain image only shows little activation).

(2) Phishing Detection Experiment Results: In the phishing task (Section IIIA), participants could be looking at the website address or the symbols or logos on the snapshot to make their decision of real or fake.

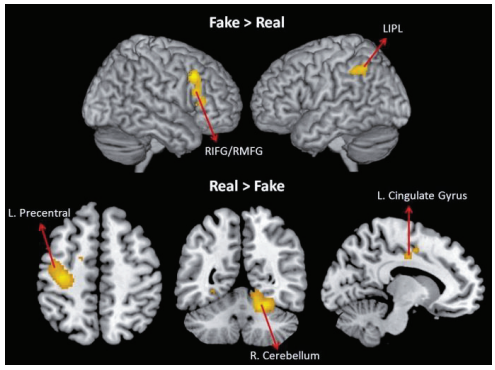


Fig 4: **Contrast of "Real" and "Fake" Activation**. Fake vs. Real activation regions include right middle, inferior, and orbital frontal gyri (RIFG/RMFG), and left inferior parietal lobule. Real vs. Fake activation regions include left precentral gyrus, right cerebellum, left cingulate gyrus, and occipital cortex.

Direct subtraction of real trials from fake trials, and fake trials from real trials revealed statistically significant activity in several areas of the brain that are critical in, and specific to, making "real" or "fake" judgments ($p < .05$, FWE corr.). For websites that the participants identified as "fake" (contrasted with "real"), participants activated the right middle, inferior, and orbital frontal gyri, and left inferior parietal lobule (see Figure 4) ($p < .05$, FWE corr.). On the other hand, when "real" websites were identified participants showed increased activity in several regions, including the left precentral gyrus, right cerebellum, left cingulate gyrus, and the occipital cortex ($p < .05$, FWE corr.).

All participants of this study also completed the Barratt's Impulsiveness Scale (BIS), a 30 item self-report instrument designed to assess the personality/behavioral construct of impulsiveness [1]. Studies have shown that BIS possesses reliability and criterion-related validity across samples [65]. Impulsive responding can result in behavioral errors, and such responses can be critical in computer security interactions where the consequences can be costly. Thus, our goal was to examine the impact of impulsive decisions on phishing task performance and identifying the neural circuitry underlying such behavior. A regression analysis involving BIS scores from participants as a covariate with whole brain activation during all trials revealed a statistically significant negative relationship in the MPFC ($p < .05$, FWE corr.) (See Figure 5).

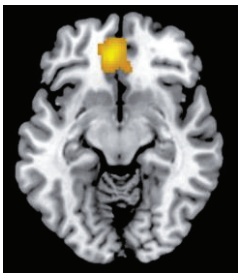


Fig 5 **Impulsivity vs. MPFC Activation:** There exists a negative relationship between impulsivity and brain activity in medial prefrontal cortex (MPFC).

Interpretation and Discussion (Phishing Detection): Increased activation was found in the right frontal and left parietal regions of participants while deciding that a given website was "fake" (Figure 4). At one level, this is evidence of a strategic and controlled approach to completing a more difficult task (identifying fake websites). These findings are, however, consistent with at least one previous fMRI study [24], where participants were asked to identify whether a series of Rembrandt paintings were real or fake. This study found increased activity in RMFG when participants identified fake paintings. Fake websites may pose more of a challenge to participants as they may have to spend more time thinking about different attributes, sometimes recalling from memory. Middle frontal, inferior frontal, and inferior parietal areas have also been implicated in working memory [25]. Identifying real websites activated precentral, cerebellum, cingulate and visual areas of the brain (Figure 4). In addition to their motor functions, the cerebellum and precentral gyrus have topographically organized feedforward and feedback projections [26]. This network may mediate the decision-making process of whether a given website is real.

Yet another finding from the present study pertains to a brain-behavior relationship. Personality traits, such as impulsivity, may prove vital in the way an individual

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2016.2566265, IEEE Transactions on Information Forensics and Security

approaches a cognitively demanding task. The present study found an inverse relationship between impulsivity and MPFC activity during phishing decisions (Figure 5). Evidence from previous studies suggests MPFC's executive/regulatory function mediates competing and conflicting cognitive operations and scenarios [27, 28, 29, 30, 31]. Studies involving animal models suggest a pivotal role of MPFC in impulsive decision-making [32]. Functional MRI studies of delay discounting have found inverse correlations between participants' impulsive choice of decisions and activity in regions like MPFC [33, 34]. Delay discounting refers to giving future consequences less weight relative to more immediate consequences (e.g., [35]). In other words, delay discounting can be construed as the tendency to choose a smaller, sooner occurring reward over a larger, later occurring reward. Similar finding of inverse correlations in the present study suggests the conflict and difficulty involved in making real or fake decisions during the phishing task for impulsive individuals.

(3) Malware Warnings Experiment Results: In this experiment (section IIIB), there were three experimental conditions: *abstract*, *warning*, and *non-warning*. Comprehending a warning, relative to comprehending the news abstracts, elicited a statistically significant increase in activation in several regions of the right hemisphere, such as the RIPL, RMTG/RSTG, and cuneus (see Figure 6). Processing non-warning pop-ups, relative to news item abstracts, also elicited similar general patterns of brain activation, albeit with some differences depending on the condition. There was bilateral activation in middle/superior temporal cortex in this contrast. In addition, the right parietal activation was relatively more anterior, in the postcentral gyrus.

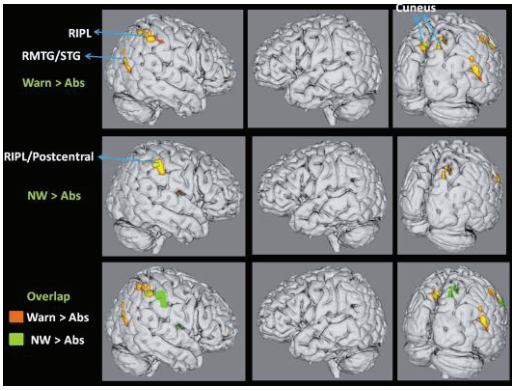


Fig 6: (**Warning or Non-Warning) vs. Abstract Activation**. Activation regions include right inferior parietal lobule (RIPL), right middle/superior temporal gyrus (RMTG/RSTG), and cuneus, as well as bilateral middle/superior temporal cortex, and right parietal in the postcentral gyrus. (The second column brain images do not show any activation; they are included for the sake of completeness)

One of the main goals of this study was to examine the brain areas that may mediate how people approach malware warnings. Our study participants showed significant increases in brain activity in several areas while processing warnings, compared to non-warnings. These regions included LIFG and LMTG, both primarily associated with processing language. There were also increases in activity in regions such as the MPFC, and in the bilateral occipital cortices ($p < .05$, FWE corr.) (see Figure 7). On the other hand, we did not find any increase in brain activity for the non-warning condition, compared to the warning condition.

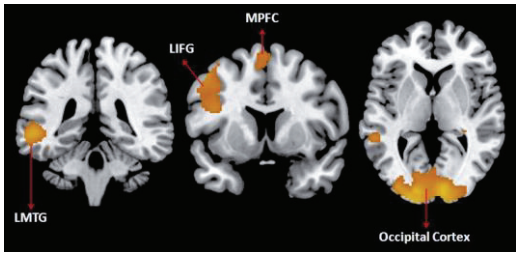


Fig 7: **Warning vs. Non-Warning Activation.** Activation regions include left middle temporal gyrus (LMTG), left inferior frontal gyrus (LIFG) as well as medial prefrontal cortex (MPFC), and bilateral occipital cortices.

To examine personality traits and their impact on computer security decisions, as in the phishing data analysis, we used impulsivity scores as a covariate in a regression analysis with brain activity while reading security warnings. This analysis revealed significant negative relationship between impulsivity and brain activity in MPFC and precuneus ($p < .05$, FWE corr.) (See Figure 8).

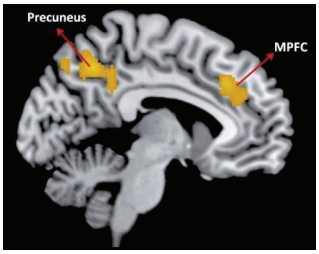


Fig 8: **Impulsivity vs. Activation:** There is a negative relationship between impulsivity and brain activity in medial prefrontal cortex and precuneus

Interpretation and Discussion (Malware Warnings):

In this study, reading warnings as contrasted to reading news abstracts generated significant brain activity in regions such as the RIPL and RMTG/RSTG (Figure 6). This activation pattern provides further evidence of the role of these regions in different aspects of language comprehension (see [36, 37, 38]). Activation in these areas suggests that the participants in the present study were progressing through the warnings to understand the conveyed message and make a decision.

There were also qualitative differences in activation between processing warning and non-warning pop-ups. Warnings generated statistically significant increase in activity in the language comprehension areas of the brain, such as LIFG and LMTG and in decision making areas like MPFC (Figure 7). In addition, there was a statistically significant activation in bilateral occipital cortices, which may provide evidence of how much visual attention and inspection participants were engaging in during warnings. On the other

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2016.2566265, IEEE Transactions on Information Forensics and Security

hand, non-warnings, which usually were not a threat, did not generate any extra activation when compared with the warning condition.

Impulsive decisions can affect user safety and security in a computer security interaction (as we demonstrated in the case of phishing). We found trait impulsivity in our participants, measured by the Barratt's Impulsivity Scale, to negatively predict brain activity in MPFC and the precuneus while paying attention to security warnings (Figure 8). Thus, more impulsive participants had less activity in these regions during the malware task. This finding is consistent with findings from several previous neuroimaging studies. For example, the precuneus was found to be negatively correlated with measures of impulsivity in a response inhibition task [42]. MPFC grey matter volume has also been found to be negatively correlated with impulsivity [43].

VI. Functional Connectivity Analysis

Functional connectivity (the synchronization of the time-course of activity across different brain areas) was also examined to understand coordination among different brain regions in accomplishing phishing decisions[2]. The regions of interest (ROIs) for all analyses consist of the bilateral inferior frontal gyrus (LIFG, RIFG), inferior parietal lobule (LIPL, RIPL), middle frontal gyrus (LMFG, RMFG), middle temporal gyrus (LMTG, RMTG), occipital cortex (LOC, ROC), orbitofrontal cortex (LOFC, ROFC), superior temporal gyrus (LSTG, RSTG), and medial prefrontal cortex (MPFC). These ROIs were chosen based on the group activation for the entire task vs. the fixation contrast. This was done to insure it represented the activation pattern in individual subjects, rather than resorting to anatomical ROIs. Seeds were created using spherical binary masks (6mm-radius) and residual time-series were extracted from each study condition (phishing and phishing control), thus enabling the comparison of functional connectivity between the two. To reduce the number of ROI pairwise comparisons and control for Type I error, the functional ROIs were grouped into 4 different anatomical networks based on the lobe to which they belong: Frontal (LIFG, RIFG, LMFG, RMFG, LOFC, ROFC, MPFC), Parietal (LIPL, RIPL), Temporal (LMTG, RMTG, LSTG, RSTG), and Occipital (LOC, ROC). This grouping allowed us to examine connectivity across these four networks.

Our first analysis consisted of examining functional connectivity of a specific ROI with the entire brain (whole-brain analysis). This analysis served to examine functional connectivity from one specific region with every other region in the brain as a measure of global connectivity. For this, we chose four ROIs: LMFG, RMFG, LIPL and RIPL, the regions activated when participants were involved in phishing decision-making. In addition, these regions have also been implicated in several cognitive tasks such as decision- making, attention-shift, and visual processing, including our previous study (See Section V.B).

Using the residual time courses, these were correlated with every other voxel in the brain for every participant. A Fisher's r to z transformation was applied to the correlation maps for each participant before averaging and computing statistical maps for each seed. We then statistically compared Phishing vs. Phishing Control using AFNI's 3dttest++ (paired-sample t-tests). To correct for multiple comparisons, 10,000 Monte Carlo simulations were computed to obtain a cluster-size-corrected threshold of $p < .05$ family wise error (FWE). We also examined the relationship between each functional connectivity map derived from whole-brain analysis with the measure of impulsivity from each participant, and correction for multiple comparisons was performed as described above. Our second analysis of functional connectivity consisted of examining connectivity among all ROIs and networks listed above, also known as ROI analysis, where correlation coefficients were calculated across the residual time courses and were subsequently z-transformed using an inverse hyperbolic tangent function, followed by direct comparison of the z-transformed correlations between Phishing and Phishing Control using paired-sample t-tests.

Whole-Brain Analysis Results: During the Phishing task, strong functional connectivity was detected in middle frontal, occipital, superior parietal, SMA, and superior temporal regions across all four seeds ($p < .05$, FWE corr.; LMFG, RMFG, LIPL, RIPL). On the other hand, during Phishing Control task, the same pattern was observed, although less robustly (Figure 9). This pattern of reduced connectivity during Phishing Control task was corroborated by the results of statistical comparison between the two tasks.

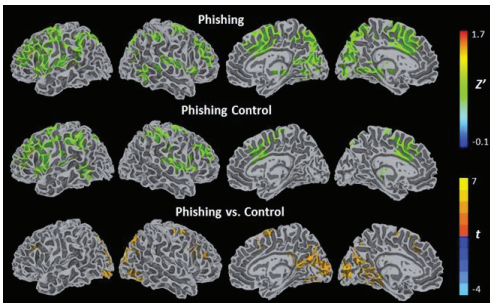


Fig 9.a: Connectivity of LMFG with rest of the brain

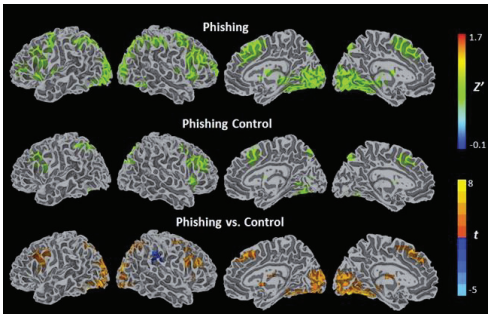


Fig 9.b: Connectivity of RMFG with rest of the brain

[2] A similar analysis was performed for the malware warning task. However, no statistically significant results were obtained and are thus not reported in this paper.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2016.2566265, IEEE Transactions on Information Forensics and Security

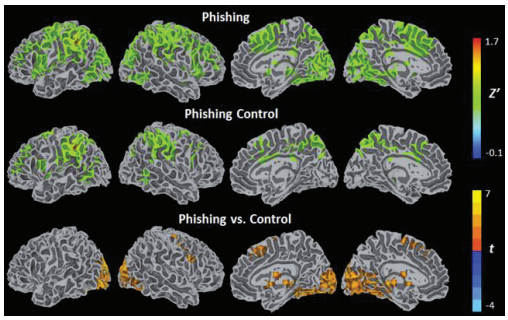


Fig 9.c: Connectivity of LIPL with rest of the brain

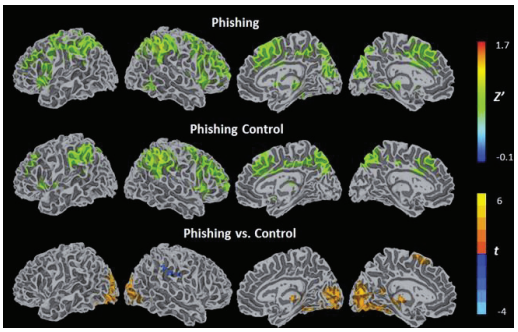


Fig 9.d: Connectivity of RIPL with rest of the brain

Fig 9: **Phishing vs Phishing Control**, Strong functional connectivity was detected in middle frontal, occipital, superior parietal, SMA, and superior temporal regions across all four seeds.

Using the LMFG seed, stronger functional connectivity was found with left calcarine sulcus, right angular gyrus, left middle occipital gyrus, right SMA, left and right IFG, RMFG, and right thalamus ($p < .05$, FWE corr.). No inverse effects were found (control > phishing). The RMFG seed showed stronger functional connectivity during phishing compared to phishing control with bilateral occipital gyrus, bilateral IFG, right thalamus, left superior medial gyrus, left hippocampus, right insula lobe, and RMFG ($p < .05$, FWE corr.); and stronger functional connectivity during phishing control compared to phishing with right supramarginal gyrus ($p < .05$, FWE corr.). The LIPL seed also showed stronger connectivity with left middle occipital gyrus, left caudate nucleus, right middle cingulate cortex, right precentral, and left SMA during Phishing Task ($p < .05$, FWE corr.). No inverse effects were found (control > phishing). Finally, the RIPL seed showed stronger connectivity with left occipital gyrus, left hippocampus, right thalamus, and left SMA ($p < .05$, FWE corr.); and stronger functional connectivity during phishing control compared to phishing with right supramarginal gyrus ($p < .05$, FWE corr.).

Region of Interest and Network Analysis Results: This analysis revealed stronger functional connectivity during phishing compared to phishing control in the following ROI pairs: LIFG: LIPL ($p = .03$), RIFG: LMTG ($p = .04$), RIFG: LOCC ($p = .03$), RIFG: ROCC ($p = .03$), RMFG: LOCC ($p = .0006$), RMFG: ROCC ($p = .004$), LMTG: LSTG ($p = .009$), LMTG: RSTG ($p = .008$), RMTG: LSTG ($p = .002$), RMTG: RSTG ($p = .0007$), RMTG: MPFC ($p = .03$), LOFC: ROFC ($p = .006$), LOFC: MPFC ($p = .02$), and LSTG: MPFC ($p = .02$) (see Figure 10). No inverse effects (phishing control > phishing) were detected. However, these results did not survive multiple comparisons correction; therefore caution is advised when interpreting these results. On the other hand, after grouping the ROIs into their respective anatomical networks (See Methods), stronger functional connectivity was found during phishing compared to phishing control task in Frontal: Parietal ($p = .02$), Temporal: Occipital ($p = .03$), and Parietal: Occipital ($p = .0004$). Parietal: Occipital functional connectivity was the only significant result that survived multiple comparisons correction (Bonferroni correction, $p < .05/6 = .008$). No inverse effects (Phishing Control > Phishing) were detected.

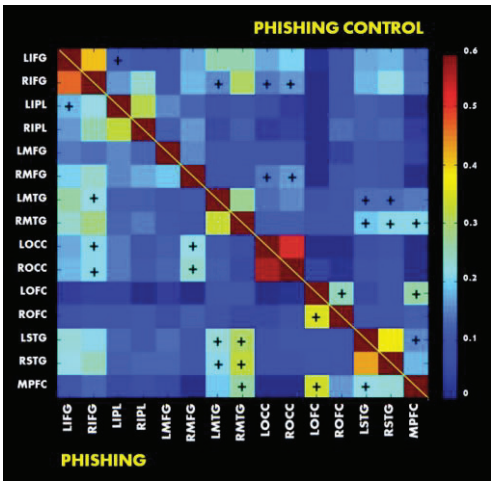


Fig 10: **Phishing vs. Phishing Control,** Functional connectivity detected in different pairs of Regions of Interests.

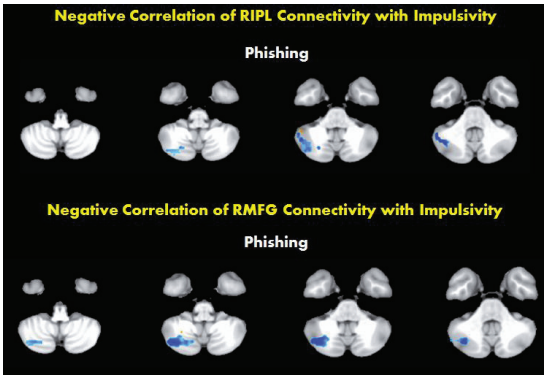


Fig 11: **Brain-behavior correlations,** Negative correlation between Impulsivity and functional connectivity across four ROIs.

Brain-Behavior Correlations: Using the functional connectivity map derived for each condition from all four ROIs, we found significant relationships (FWE corrected) with Impulsivity scores measured by the Barratt Impulsiveness Scale [1,4]. The RIPL seed showed negative correlations with left cerebellum ($r = -0.7$, $p = .0002$) and the RMFG seed also showed negative correlations with left cerebellum ($r = -0.77$, $p = .0002$) during phishing (Figure 11).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2016.2566265, IEEE Transactions on Information Forensics and Security

Interpretation and Discussion: Using two frontal and two parietal ROIs as seeds of interest, we found increased connectivity of this frontal-parietal network with the rest of the brain when engaged in phishing task. The phishing task is more complex and demanding, than the passive viewing involved in the phishing control task, thus eliciting stronger coordination among core regions of the executive network of the brain. The four seed regions in the present study are part of the frontal-parietal control system, which is usually engaged in tasks that require controlled processing, problem-solving, and decision-making [31,57,41]. The frontal-parietal control system is particularly engaged in tasks that elicit controlled processing related to the simultaneous consideration of multiple interdependent contingencies [58], conflicting stimulus-response mappings [40], and integrating working memory with attentional resource allocation [8].

During the phishing task, the RIPL and RMFG seeds showed negative correlations with left cerebellum on brain-behavior correlations (see Figure 11). Regions such as RIPL and RMFG have been shown to be functionally connected with cerebellum and have an important role in supramodal cognitive [64]. Therefore, it is possible that individuals who have higher impulsivity may have decreased functional connectivity between these regions during the phishing condition.

VII. Cross-Experiment Analysis

A. Phishing vs. Malware

Both phishing and malware tasks in our study involved decision-making, perhaps in slightly different ways. At the neural level, we examined the correlation between these two tasks in terms of the brain activity in two regions, LMFG and RMFG, which are associated with decision-making. We found a significant positive correlation in both LMFG and RMFG activity, particularly in the RMFG region (see Figure 12).

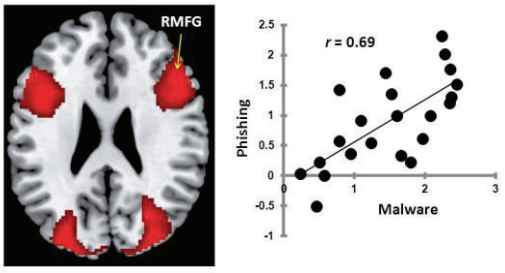


Fig 12: Correlation in Phishing and Malware in RMFG Activation

At the UI level the two tasks are different – warnings involve reading and comprehension, while phishing detection involves explicit decision making. Still, these results suggest that both phishing detection and malware warnings involve similar, higher level cognitive and neural processes. We may thus infer that participants' behavior in these two distinct yet related tasks may be well-aligned in that one's ability to heed malware warnings may be associated with his/her decisions about the legitimacy of websites and vice versa.

VIII. Discussion: Study Insights and Implications

Our neuroimaging data showed that users exhibited significant brain activation and connectivity in areas of the brain associated with decision making, problem solving, attention and visual search during the phishing detection task, while their accuracy in this task, as determined by behavioral data, was only slightly better than making a random guess (in line with a prior lab study [10]). This suggests that although the eventual decision made by the participants to differentiate between fake and real websites may be far from accurate, they expended considerable effort in making this decision as reflected by their brain activity in regions correlated with higher order cognitive processing. Perhaps this was because many of the participants did not know what markers (e.g., URL or logo) to look for on the sites to make their decisions. We note that a large fraction of our participants were majoring in a non-technical (non-computer) field. Overall, these findings further justify the need for specialized *education and training* for everyday users that focuses on phishing in particular (such as the efforts of [47, 48]) and security in general (such as [49, 50]). These training and awareness programs may help to improve users' phishing detection performance and reduce the chances of their susceptibility to other attacks. At the same time, the findings also demonstrate the need for continued research on designing phishing resistant software solutions and user interfaces.

Another important application of our work to cyber-security pertains to the automated (subconscious) detection of "real" and "fake" websites based on neural signatures. Our study revealed differences in brain areas activated during identification of "real" and "fake" websites. This means that users' may be subconsciously detecting differences between the two websites, although consciously they may fail to detect them. This result is in line with the study of real and fake Rembrandt paintings by Huang et al. [24]. These brain differences may be leveraged to build an automated real-fake detection engine in the future (e.g., in real-time using EEG measures).

The malware warnings task triggered significant brain activity in regions primarily associated with language comprehension and reading. Importantly, actual malware warnings, in contrast to casual pop-ups, generated significantly more activation in brain areas governing language comprehension, visual attention, and inspection. This suggests that participants were reading through the warnings carefully to understand the message conveyed and attempting to make an appropriate decision. Indeed, this was validated via our behavioral data which showed that participants heeded warnings about 90% of the time (also in line with the recent large-scale field study of [11]). We therefore believe that our study provides a *neurological basis* for users' ability to process and heed malware warnings, further validating the results of [11]. It should be noted that since our security warnings were simplified, our results may underestimate users' performance when faced with malware warnings, which could be improved with better warnings

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2016.2566265, IEEE Transactions on Information Forensics and Security

(such as those employed by modern browsers and variants thereof [11]).

Another key component of our study was to asses users' performance in user-centered security tasks based on one personality trait, impulsivity. Specifically, we studied the effect of impulsivity measured via a simple questionnaire. The study conducted by Pattison et al. [62] had found that less impulsive individuals were better at identifying and managing phishing emails. In our study, we did not find statistically significant relationship between impulsivity and task performance. However, we found that, in both phishing detection and malware warnings tasks, impulsive individuals showed significantly less brain activation and connectivity in regions governing decision-making and problem solving. This implies that impulsive behavior might be counter-productive to phishing detection and malware warnings task performance. A long-term impact of this finding can be in developing targeted security training programs. For example, an organization may concentrate their security training efforts on employees who are highly impulsive, as determined by their scores in the impulsivity questionnaire [1]. Similarly, school authorities may focus their online child safety efforts on children with high impulsivity levels. In such cases, for ethical and privacy reasons, we expect that users' personality scores and neural activation levels would be kept private in secure storage (just like other personal records). These scores would then be used for identifying clusters of personnel needing different types of training.

A unique advantage of our study was that it allowed for a direct comparison between phishing detection and malware warnings tasks. In this respect, we found significant correlation in participants' brain activity governing decision-making regions (bilateral middle frontal gyri). This suggests that both tasks involve, at a higher level, similar cognitive processes and that users' performance in the two tasks might be correlated with each other. Note that, although language comprehension is unique to the malware task, both tasks involved a crucial decision making aspect. Broadly, this seems to indicate that the cognitive mechanisms underlying these security tasks are related, which may translate into similarity in users' performance in the two tasks.

Although fMRI scans are usually expensive, we believe that our methodology could also serve the purpose of *security screening* of individuals. Impulsivity questionnaires alone might be helpful in predicting users' susceptibility to attacks in some cases. However, since those questionnaires are "self-reported," the users may, knowingly or unknowingly, not provide the accurate responses, although the BIS have been shown to possess high levels of reliability. By scanning users using neuroimaging techniques like fMRI, we can capture and analyze users' brain signals, which users will not be able to change or lie about, and predict their potential for attacks in real-life. Such neural signatures governing users' phishing detection capability (or lack thereof) – the primary subject of our study - has applications in organizations with high security requirements, such as national defense.

IX. Study Limitations

In line with any study involving human subjects, ours also has certain limitations. A primary limitation pertains to the constraints posed by the fMRI experimental set-up. Since participants were performing tasks inside the fMRI scanner, the set-up did not mimic "real-world" online browsing experiences. The discomfort associated with lying in a supine position and being stationary may have also impacted participants' brain activity. In addition, the fact participants were being scanned may have impacted their brain activation and behavioral responses. The constrained interface (image-based display, binary input and no internet connectivity, unlike a modern computer) available during the scans may have limited participants' interactions with the system. For example, the participants were presented with *images* of websites rather than with the websites themselves in the phishing task. Similarly, the malware warning images were very simplistic and rudimentary due to equipment constraints. We believe this may have negatively affected participants' performance in the underlying security tasks. Furthermore, although we corrected for participants' head motion in the MRI scanner, it may have impacted fMRI data quality. During our neural analysis of the phishing detection task, we investigated only real and fake conditions, irrespective of the correct or incorrect responses given to them. The primary reason for not directly comparing correct and incorrect conditions for the phishing task was not having a large enough number of trials in each condition to have the necessary statistical power to detect a significant effect. We suggest that future studies consider the users' judgments in neural analysis. Finally, the lab-based environment of the study may have impacted participants' behavior, as they may not have felt actual security risks were occurring during the experiments.

The effective sample size used in our study ranged from 22 (phishing detection task and phishing control task) to 25 (malware warnings task) participants (see Section V.B), which previous power analysis studies have found to be optimal. 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 [55].

X. Conclusions and Future Work

In this paper, we presented an fMRI study to bring insights into user-centered security by focusing on phishing detection and responding to malware warnings. Our results provide a largely positive perspective towards users' capability and performance vis-à-vis these crucial security tasks. We found that users showed significant brain activity in key regions known to govern decision-making, attention, and problem-solving ability (phishing and malware warnings) as well as language comprehension and reading (malware warnings). Apart from that, we saw strong functional connectivity in several regions of the brain while performing the phishing task. This level of activation and connectivity indicates that users were actively engaged in the tasks and were not ignoring or bypassing them (as prior lab studies have concluded [12,

13, 14, 15, 16, 17]). In the case of the malware warnings task, brain activity and behavioral performance (accuracy) were complementing each other validating that users heed malware warnings with a high likelihood (as also shown by a recent field study [11]). For the phishing task, however, task performance was poor despite significant brain activity associated with decision making. This divergent result demands future investigation. It could be attributed to users' lack of knowledge as to the markers for "fake" vs. "real" website decisions (e.g., URLs), which may be overcome by user education and training. We also demonstrated that individuals with higher impulsivity may not utilize brain areas (MPFC) associated with making decisions of a conflicting nature as efficiently as non-impulsive individuals and may result in poorer cognitive and behavioral outcomes. This suggests it would be valuable to study whether individual trait characteristics should factor into user-centered security design. Finally, we discovered a high degree of correlation in brain activity with respect to decision-making regions across phishing detection and malware warnings tasks. This correlation suggests users' behavior in one task may be predicted by their behavior in the other.

We see a clear path-forward for subsequent research using neuroimaging techniques (e.g., fMRI, EEG or fNIRS) to inform the design of user-centered security systems. In the long-run such studies may provide a neural signature for poor and good security decisions which can be used for predicting - as well as correcting - users' security behavior. Future research may conduct subsequent evaluation with diverse participant samples, study the effect of warning fatigue or habituation, consider user-centered security domains other than phishing detection and malware warnings (e.g., password memorization and recall), and evaluate the effect of security training and education on users' performance.

XI. Acknowledgments

We are grateful to Keya Kuruvilla and Michael Georgescu for their help with various aspects of the study. We also thank N. Asokan, Cali Fidopiastis, Lauren Libero, Ivan Martinovic, Paul Van Oorschot, and John Sloan for their feedback on a previous version of this paper, and Rishi Deshpande for initial help with the experimental set-up. We also thank David Wagner and NDSS'14 anonymous reviewers for their constructive input and guidance.

References

[1] Barratt, E.S. (1994). Impulsiveness and Aggression. In Monahan, J. and H. J. Steadman (Eds.), Violence and Mental Disorder: Developments in Risk Assessment (pp. 61-79). University of Chicago Press, Chicago, IL.

[2] An introduction to E-Prime, Laurence Richard, Miami University, Dominic Charbonneau, Université de Montréal, Tutorials in Quantitative Methods for Psychology 2009, vol 5(2), p. 68-76.

[3] http://step.psy.cmu.edu/materials/manuals/users.pdf

[4] Patton, J.H., Stanford, M.S., and Barratt, E.S. (1995). Journal of Clinical Psychology, 51, 768-774.

[5] S. Ogawa, TM Lee, AR Kay, and DW Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proceedings of the National Academy of Sciences, 87(24):9868, 1990.

[6] R.S. Menon, J.S. Gati, B.G. Goodyear, D.C. Luknowsky, and C.G. Thomas. Spatial and temporal resolution of functional magnetic resonance imaging. *Biochemistry and cell biology*, 1998.

[7] Neupane, A., Saxena, N., Kuruvilla, K., Georgescu, M., & Kana, R. (2014). 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)* (pp. 1-16).

[8] Koechlin, E., Basso, G., Pietrini, P., Panzer, S., & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature*, *399*(6732), 148-151.

[9] Cranor, L. F. A framework for reasoning about the human in the loop. In Proceedings of the 1st Conference on Usability, Psychology, and Security, UPSEC'08, pages 1:1–1:15, 2008.

[10] Dhamija R., J. D. Tygar, and Marti Hearst. 2006. Why phishing works. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '06), 581-590, 2006.

[11] Akhawe, D., & Felt, A. P. (2013, August). Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness. In *Usenix Security* (pp. 257-272).

[12] 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* (CHI '06)

[13] Sunshine, J., Egelman, S., Almuhimedi, H., Atri, N., & Cranor, L. F. (2009, August). Crying Wolf: An Empirical Study of SSL Warning Effectiveness. In*USENIX Security Symposium* (pp. 399-416).

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

[15] Friedman B., Hurley D, Howe D.C., Felten E., and Nissenbaum H. Users' conceptions of web security: a comparative study. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems*.

[16] Egelman, S., Cranor, L. F., & Hong, J. (2008, April). You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1065-1074). ACM.

[17] Bravo-Lillo, C., Cranor, L. F., Downs, J. S., & Komanduri, S. (2011). Bridging the gap in computer security warnings: A mental model approach. *IEEE Security & Privacy*, *9*(2), 0018-26.

[18] Adam W. Craig, Yuliya Komarova Loureiro, Stacy Wood, Jennifer M.C. Vendemia (*2012*) Suspicious Minds: Exploring Neural Processes During Exposure to Deceptive Advertising. Journal of Marketing Research: June 2012, Vol. 49, No. 3, pp. 361-372.

[19] Hristo Bojinov. Daniel Sanchez, Paul Reber, Dan Boneh, Patrick Lincoln. Neuroscience Meets Cryptography: Designing Crypto Primitives Secure Against Rubber Hose Attacks. In *21st USENIX Security Symposium*. USENIX Association. August, 2012.

[20] Martinovic, I., Davies, D., Frank, M., Perito, D., Ros, T., & Song, D. (2012, August). On the Feasibility of Side-Channel Attacks with Brain-Computer Interfaces. In *USENIX Security Symposium* (pp. 143-158).

[21] Ward, B.D. (2000). Simultaneous inference for fMRI data. Available at: http:// homepage.usask.ca/~ges125/fMRI/AFNIdoc/AlphaSim.pdf

[22] Shomstein,Sarah, Cognitive Functions of the Posterior Parietal Cortex: Top-down and bottom-up attentional control, Frontiers in Integrative Neuroscience, Volume 6, July 2012 , ISSN 1662-5145.

[23] DiQuattro, N. E., Geng, J.J.,Contextual knowledge configures attentional control networks, The Journal of neuroscience : the official journal of the Society for Neuroscience, Volume 31, Issue 49, December 2011, Pages 18026-1803.

[24] Huang Mengfei, Bridge Holly, Kemp Martin J, Parker Andrew J, Human cortical activity evoked by the assignment of authenticity when viewing works of art, Frontiers in Human Neuroscience, Volume 5, 2011, Number 00134, ISSN 1662-5161.

[25] Juliana V. Baldo, Nina F. Dronkers, The Role of Inferior Parietal and Inferior Frontal Cortex in Working Memory, Neuropsychology, Volume 20, Issue 5, 2006, Pages 529-538.

[26] Nigel Blackwood, Dominic ffytche, Andrew Simmons, Richard Bentall, Robin Murray, Robert Howard, The cerebellum and decision making

under uncertainty, Cognitive Brain Research, Volume 20, Issue 1, June 2004, Pages 46-53, ISSN 0926-6410.

[27] Buckner, Randy L., and Daniel C. Carroll, Self-projection and the brain, Trends in Cognitive Sciences Volume 11 Issue 2, 2007, Pages 49-57.

[28] Burgess GM, Mullaney I, McNeill M, Dunn PM, Rang HP. Second messengers involved in the mechanism of action of bradykinin in sensory neurons in culture. J Neurosci. 1989;9:3314–3325.

[29] Gallagher, HL and Frith, CD (2003) Functional imaging of 'theory of mind'. Trends Cogn Sci 7:77–83.

[30] Gilbert DL, Wang Z, Sallee FR, Ridel KR, Merhar S, Zhang J, et al. Dopamine transporter genotype influences the physiological response to medication in ADHD. Brain.Volume 129, Issue 8, 2006, 2038–2046.

[31] Ramnani, Narender., owen,Adrian M., Anterior prefrontal cortex: insights into function from anatomy and neuroimaging, Nat Rev Neurosci, Volume 5, Issue 3, march 2004, Pages 184-194

[32] Loos, M., Pattij, T., Janssen, M. C., Counotte, D. S., Schoffelmeer, A. N., Smit, A. B., ... & van Gaalen, M. M. (2010). Dopamine receptor D1/D5 gene expression in the medial prefrontal cortex predicts impulsive choice in rats.*Cerebral Cortex*, *20*(5), 1064-1070.

[33] Luhmann, J. G.,Curtis, D. W.,Schroeder, P.,McCauley, J.,Lin, R. P.,Larson, D. E.,Bale, S. D.,Sauvaud, J. A.,Aoustin, C.,Mewaldt, R. A.,Cummings, A. C.,Stone, E. C.,Davis, A. J.,Cook, W. R.,Kecman, B.,Wiedenbeck, M. E.,Rosenvinge, T.,Acuna, M. H.,Reichenthal, L. S.,Shuman, S.,Wortman, K. A.,Reames, D. V.,Mueller-Mellin, R.,Kunow, H.,Mason, G. M.,Walpole, P.,Korth, A.,Sanderson, T. R.,Russell, C. T.,Gosling, J. T., STEREO IMPACT Investigation Goals, Measurements, and Data Products Overview, Space Science Reviews, Volume 136, Issue 1-4, April 2008, Pages 117-184.

[34] Sripada, C. S., Gonzalez, R., Phan, K. L., Liberzon, I., The neural correlates of intertemporal decision-making: contributions of subjective value, stimulus type, and trait impulsivity, Human brain mapping, Volume 32, Issue 10, October 2011, Pages 1637-1648.

[35] Frederick, S. Valuing future life and future lives: A framework for understanding discounting. *Journal of Economic Psychology*, 2006.

[36] Price, C.J. The anatomy of language: contributions from functional neuroimaging. *Journal of Anatomy,* 197, 335–359, 2000.

[37] Bookheimer, S., Functional MRI of language: New approaches to understanding the cortical organization of semantic processing, Annual Review of Neuroscience, Volume 25, 2002, Pages 151-188.

[38] Friederici, A.D., Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences,* 6, 78–84, 2002.

[39] Hagoort, P. On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences,* 9, 416-423, 2005.

[40] Crone EA, Wendelken C, Donohue SE, Bunge SA. Neural evidence for dissociable components of task-switching. Cereb Cortex 16: 475–486, 2006.

[41] Botvinick, M.M, Cohen, J.D., & Carter, C.S. (2004). Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn Sci., 8:539–546.

[42] Mortensen M, Ebert B, Wafford K & Smart TG, Distinct activities of GABA agonists at synaptic- and extrasynaptic-type GABAA receptors, The Journal of Physiology, Volume 588, 2010, Pages 1251-1268.

[43] Moreno-López L, Catena A, Fernández-Serrano MJ, Delgado-Rico E, Stamatakis EA, Pérez-García M, Verdejo-García A. (2012). Trait impulsivity and prefrontal gray matter reductions in cocaine dependent individuals. Drug Alcohol Depend. 2012 Oct 1;125 (3):208-14.

[44] Ernst, M. & Paulus, M.P. Neurobiology of decision making: a selective review from a neurocognitive and clinical perspective. *Biological Psychiatry*, 58, 597–604, 2005.

[45] Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E. & Cohen, J.D. The neural basis of economic decision-making in the Ultimatum Game. Science, 300, 1755–1758, 2003.

[46] Miller, E. K., & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167-202, 2001.

[47] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong. Teaching johnny not to fall for phish. ACM Trans. Internet Technol., 10(2):1{31, 2010.

[48] Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007, July). Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In *Proceedings of the 3rd symposium on Usable privacy and security* (pp. 88-99). ACM.

[49] Sukamol Srikwan and Markus Jakobsson. Using cartoons to teach internet security. Cryptologia, 32(2):137{154, 2008.

[50] Alain Forget, Sonia Chiasson, and Robert Biddle. Lessons from brain age on password memorability (poster). In Future Play '08: Proceedings of the 2008 Conference on Future Play, pages 262{263, 2008.

[51] Rosen BR, Buckner RL, Dale AM. Event-related functional MRI: past, present, and future. Proc Natl Acad Sci USA 95:773– 780, 1998.

[52] Julie Thorpe, P. C. van Oorschot, and Anil Somayaji. Pass-thoughts: authenticating with our minds. In the workshop on New security paradigms (NSPW '05), 2005.

[53] John Chuang, Hamilton Nguyen, Charles Wang, and Benjamin Johnson. I Think, Therefore I Am: Usability and Security of Authentication Using Brainwaves. In Workshop on Usable Security (USEC), 2013.

[54] Oldfield, R.C. The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia, 9, 97-113, 1971.

[55] Murphy, K., & Garavan, H. An Empirical Investigation into the number of subjects required for an event-related fMRI study. Neuroimage, 22, 879-885, 2004.

[56] Cox, R.W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29, 162–73,1996.

[57] Dosenbach, N.U., Fair, D.A., Miezin, F.M., Cohen, A.L., Wenger, K.K., Dosenbach, R.A., Fox, M.D., Snyder, A.Z., Vincent, J.L., Raichle, M.E., Schlaggar, B.L., Petersen, S.E. (2007). Distinct brain networks for adaptive and stable task control in humans. Proc Natl Acad Sci USA 104: 11073–11078.

[58] Kroger JK, Sabb FW, Fales CL, Bookheimer SY, Cohen MS, Holyoak KJ. Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: a parametric study of relational complexity. Cereb Cortex 12: 477–485, 2002.

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

[60] Wogalter, M. S. (2006). Communication-human information processing (C-HIP) model. Handbook of warnings, 51-61.

[61] Mayhorn, C. B., Welk, A. K., Zielinska, O. A., & Murphy-Hill, E. (2015, August). Assessing Individual Differences in a Phishing Detection Task. In*Proceedings 19th Triennial Congress of the IEA* (Vol. 9, p. 14).

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

[63] Wogalter, M. S., and C. B. Mayhorn. "The future of risk communication: Technology-based warning systems." Handbook of warnings (2006): 783-793.

[64] O'Reilly JX, Beckmann CF, Tomassini V, Ramnani N, Johansen-Berg H, "Distinct and overlapping functional zones in the cerebellum defined by resting state functional connectivity". Cereb Cortex 20: 953–965.

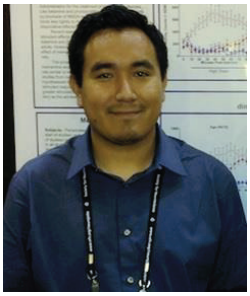
[65] Vasconcelos, A. G., Malloy-Diniz, L., & Correa, H. (2012). Systematic review of psychometric proprieties of Barratt Impulsiveness Scale–version 11 (BIS-11). *Clinical Neuropsychiatry–Journal of Treatment Evaluation*, *9*, 61-74.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2016.2566265, IEEE Transactions on Information Forensics and Security

Ajaya Neupane is a PhD candidate and a Graduate Research Assistant in Department of Computer and Information Sciences at University of Alabama at Birmingham. He has been working on research projects involving the study of human behavior and neuro-physiological underpinnings pertaining to the user-centered security tasks. His research interests include information security and human computer interaction. He has a Bachelors degree in Computer Engineering from the National Institute of Technology, Surat, India. He is the receipent of the "Distinguised Paper Award" at Network and Distributed System Security Conference, 2014.

Nitesh Saxena is an Associate Professor of Computer and Information Sciences at the University of Alabama at Birmingham (UAB), and the founding director of the Security and Privacy in Emerging Systems (SPIES) group/lab. He works in the broad areas of computer and network security, and applied cryptography, with a keen interest in wireless security and the emerging field of usable security.

Saxenas current research has been externally supported by multiple grants from NSF, and by gifts/awards/donations from the industry, including Google (2 Google Faculty Research awards), Cisco, Intel, Nokia and Research in Motion. He has published over 70 journal, conference and workshop papers, many at top-tier venues in Computer Science, including: IEEE Transactions, ACM CCS, ACM WiSec, ACM CHI, ACM Ubicomp, IEEE Percom, and IEEE S&P. On the educational/service front, Saxena is a co-director for UABs MS program in Computer Forensics and Security Management. He was also the principal architect and a co-director of the M.S. Program in Cyber-Security at the Polytechnic Institute of New York University (NYU-Poly). Saxena has instructed over a dozen core fundamental courses in Computer Science, including Computer Security, Network Security, Modern Cryptography and Discrete Structures. Saxena has also advised and graduated numerous graduate (Ph.D. and M.S.) and undergraduate students as well as a few high school students. He is serving as an Associate Editor for flagship securityjournals, IEEE Transactions on Information Forensics and Security (TIFS), and Springers International Journal of Information Security (IJIS). Saxenas work has received extensive media coverage, for example, at NBC, MSN, Fox, Discovery, ABC, Bloomberg, ZDNet, ACM TechNews, Yahoo News, Slashdot and Computer World.



Omar Maximo is a third year graduate student in the Lifespan Developmental Psychology program at UAB. His current interest is functional brain integration in autism spectrum disorders (ASD) using various neuroimaging approaches to better understand its neural complexity. His long-term goal is to determine the significance of brain functioning as a neurobiological signature of ASD.

Dr. Rajesh Kana is an Associate Professor in the Department of Psychology at the University of Alabama at Birmingham, AL, USA. He is the director of the *Cognition, Brain, and Autism Research Laboratory* as well as the Co-Director of the *UAB Undergraduate Neuroscience Program*. Dr. Kana is a neuroscientist who used several brain imaging techniques to understand the brain organization in healthy people and in individuals with neurodevelopmental disorders.