Subliminal Warnings – A New Approach to Change Users' Behavior

Mini Zeng
mzeng@ju.edu
Computer Science Department
Jacksonville University
Jacksonville, USA

Feng Zhu
fzhu@uah.edu
Computer Science Department
University of Alabama in Huntsville
Huntsville, USA

Sandra Carpenter
carpens@uah.edu
Psychology Department
University of Alabama in Huntsville
Huntsville, USA

Abstract— **People use web applications and services every day. They encounter many cyber-attacks in their daily lives. Cybersecurity warnings remind users to be cautious but often they are not effective because users ignore them. Existing warnings often compete for the limited resource – conscious attention. Cognitive psychology indicates that human behavior can be affected by non-conscious processing. The goal of this paper is to explore the effectiveness of subliminal warnings. We propose a subliminal warning design that directly suggests users behave cautiously. We used a pilot study to guide our design. We conducted a lab experiment to evaluate the subliminal warning strategy in the context of privacy protection. We implemented an integrated hardware and software environment to evaluate our warning design, study users' behavior, and to validate and verify the proper display of subliminal warning messages. Our experimental results showed that subliminal warnings effectively reduced identity disclosure. Warning designs based on cognitive psychology and human factors may better protect people's privacy.**

Keywords - Subliminal Messages; Non-conscious Processing; Warnings.

I. INTRODUCTION

People use web applications and services every day. They encounter many cyber-attacks in their daily lives (e.g., phishing and Domain Name System spoofing attacks). Websites or apps may send numerous spam emails. Many users have connected to legitimate websites with erroneous or self-signed certificates [1]. They often unnecessarily disclose identity information, thus putting their privacy at risk. The identity information may be collected, sold, and even maliciously used. Computer warnings protect users from exploits, but many people still fall into attackers' traps. Humans are not very good at mitigating cyber-attacks. They are considered the weakest component in security systems [2].

Warnings are used as one of the last lines of defense in computer security and are fundamental to users' security interactions with technology. When people see computer warnings, however, most of the time, they ignore them [3]. Moreover, some people do not even pay attention to security warnings [4]. On the other hand, when individuals were asked about their concerns about their private information, they claimed that they had been careful about their personal information [5]. This contradiction may be explained by the ineffectiveness of warnings [8]. The effectiveness of warnings depends on the fact that they attract users' attention, communicate clearly about the risks, and provide straightforward instructions for avoiding the hazards[6].

Researchers believe one of the fundamental problems is the lack of attention paid to warnings [7] So, recent warning designs try to grab users' attention by discontinuing their primary tasks [9]. Other researchers claimed that too much exposure to warnings and interruptions would make users quickly habituate to warnings [7] and feel "warning fatigue" [8].

Current warning theories and practices occur only at the conscious level and require attention [9]. However, researchers found that a great deal of information processing happens at noncouscious level. In addition, human behaviors are affected by non-conscious stimuli [10].

We utilize the non-conscious processing ability for warnings instead of competing for conscious attention. The last 50 years of cognitive psychology research indicate that non-conscious processing can be nudged by human perception and behavior[11]. We want to evaluate the impact of cybersecurity warnings via non-conscious processing, specifically using subliminal messages. Our previous work showed that subliminal warnings can remind users about their security and privacy attitudes [12].

In this paper, we further evaluate subliminal warnings that directly suggest cautious behavior. Our subliminal warning designs require no active attention and do not interfere with a user's primary task. The warning messages pop up just in time when the user clicks the text entry for identity information. These messages are only displayed for dozens of milliseconds. The nature of our very brief warning stimuli in non-conscious processing may even reduce habituation (warning fatigue) [13].

We use our warnings to help people avoid unnecessary identity exposure and more cautiously share their personal information online. This is an important area because there are 17 million identity theft victims every year in the U.S [17]. Unfortunately, while people are very concerned about their privacy, they do not protect their personal information well and unnecessarily expose their information online [15].

The experiment for the subliminal warning is conducted in a context of a restaurant table reservation app. We developed an experiment environment that includes eye-tracking and a scene camera. The eye tracker was used to study participants' gaze behavior, whereas the scene camera verified the proper display of the warning messages. Our experiment results indicate that our subliminal warning messages can effectively reduce disclosure.

 ISBN: 978-1-61208-982-9

II. BACKGROUND AND RELATED WORK

A. Computer Warning Framework

Warnings encourage safe behaviors. Researchers have been working on warning designs for decades. They find that users frequently ignore warnings and focus, instead, on their tasks [6]. To understand how people process warning, several models were developed, such as the Communication-Human Information Processing (C-HIP) Model [6][19], the sequential model [17], and the levels of performance model [18]. The C-HIP model separates the warning process into two aspects of the warning (i.e., source, channel) and six stages of human information processing: delivery, attention switch, comprehension, attitudes and beliefs, motivation, and behavior [6]. To increase the effectiveness of warnings, researchers and warning designers have developed many approaches to attract users' attention, to communicate clearly about the risks, and to provide straightforward instructions for avoiding the hazard [22][23].

B. Cybersecurity Warnings

Due to the reality that the human is the weakest link of the cybersecurity chain, researchers have been trying to enhance the security of this weakest link by developing and improving the security warnings for decades [24]–[26]. Researchers found that there are two common issues of ineffective security warnings. One is the lack of attention [24]. The other one is habituation [25]. Researchers have used eye tracking, mouse tracking, functional magnetic resonance imaging [29], and galvanic skin response [27] to study user behavior.

For the lack of attention, a critical approach is the warning designs with required actions, called active warnings. Such as forcing users to click the URL in the SSL warning [28], making users click through several buttons or sub-pages with multi-level warnings [29], and a pop-up warning with required buttons or links to continue current workflow [30]. These successful warning designs make warnings hard to ignore [24][34], or totally interrupt users' workflows [25][35]. But other researchers [19] argue that it is rational for users to ignore warnings because of the amount of conscious attention needed. Warnings can interrupt users' workflow, leading to "warning fatigue."

More effective approaches were therefore developed. One of the common approaches includes appropriate icons, symbols, and physical security metaphors. For example, police offer symbols were widely used [33]. Firewall warnings were designed with a physical security metaphor (using a figure dressed in a prisoner's uniform, carrying a knife and a thief's bag) [33]. Also, an SSL warning with a red background and a "Stop sign officer" security metaphor [29] has been effective.

To make the transition of attention switch and maintenance smoothly, passive warnings without forcing users to take actions were designed. For example, a windows action center sends notification massage in a "pop-up balloon" shows up and fades out after a few minutes [34]. A red open lock designed as a SSL warning indicator [31][33]. An eye-gaze controlled security warning is displayed when eye gaze moves to the hazard area and fades out when users do not need it [35].

C. Subliminal Stimulation and Applications

Contemporary research in cognitive psychology reveals that part of our information processing is at the conscious level, whereas other perception efforts may take place beneath a threshold that thus we are not consciously aware of (known as subliminal) [36]. Non-conscious level priming has not yet been considered an approach to improving computer warnings. But it has been used in other areas for years, such as advertising and education [40][41]. Many research has been done on the effect of non-conscious level priming on human behavior [39][42].

To what extent can non-conscious perception affect our behaviors? This has been one of the most controversial issues in psychology for decades. Researchers addressed this issue through experiments that use subliminal stimulation methods [40]. A subliminal stimulus is presented below a subjective threshold for conscious perception. Subliminal perception is inferred when a stimulus is demonstrated not to be consciously perceived while still influencing thoughts, feelings, actions, learning, or memory [41]. Three models could be applied for subliminal priming: subliminal priming mapping with response [42], priming by spatial attention [43] and priming by strategies [39].

Studies showed that arbitrary stimulus-response mappings could apply to subliminal stimuli [39][47]. The stimulus-response mapping is a strategy to generate a connection between a response and a stimulus. The goal of stimulus-response mapping is to encourage a response when a stimulus is displayed. Stimuli presented below the threshold of awareness can systematically influence choice responses determined by the instructed stimulus–response (S–R) mapping. Researchers also investigated whether such stimuli influence a free choice between two response alternatives under conditions in which the choice subjectively appears to be internally generated and free. This is relevant for our research because our goal is for people to choose cautious rather than risky disclosure behavior. For example, the primes were left- and right-pointing double arrows (<< and >>). The choices provided were either left- and right-pointing double arrows or outward-pointing double arrows (< >). The experimental results demonstrated that apparently "free" choices are not immune to not consciously triggered biases [44]. In our research, we apply S-R mapping in subliminal messages and icon design.

Subliminal messages have been used in advertising for decades to influence purchasing behavior. Subliminal advertising became notorious in 1957 through the publicity. James Vicary, a private market researcher tried to increase sales of Coca Cola and popcorn in a movie theatre by secretly

 ISBN: 978-1-61208-982-9

TABLE 1 PILOT STUDY CONDITIONS

Condition	Subliminal Prime	Duration for each display	Number of Displays	Display
1	Text message “Fake it”	50 ms	Five Times	At the top of Address textbox
2	Text Message “Fake it”	50 ms	Once	At the top of Address textbox
3	Icon: Yellow Triangle	50 ms	Five Times	At the right of “Next” button
4	Icon: Yellow Triangle	50 ms	Five Times	At the right of Address textbox

and subliminally flashing the message "Drink Coca Cola" and "Eat popcorn" [37]. After that, other researchers demonstrated that subliminal priming of a brand name of drink positively affected participants' choice and their intention, the primed brand, but only for participants who were thirsty [40][47]. This led some people to claim that subliminal advertising was unethical [48]. It should be used only for non-profit and beneficial purposes, such as stopping smoking or improving academic performance. For example, subliminal words were randomly displayed in different locations on slides presented to students [38]and enhanced their later performance.

In our previous work, we designed a subliminal warning message to remind users of their security and privacy attitudes based on C-HIP model. The warning was using a yellow background message "Privacy" to strengthen user's access to their own security attitudes and beliefs [12]. The approach emphasizes the effect of warning on attitudes and beliefs level of C-HIP model. In this paper, we bypass the other stages of C-HIP model and directly use the stimulus-response mapping to trigger the safe behavior.

D. Privacy and Identity Exposure

Personal privacy has been emphasized for decades. People often claim that they care about their privacy when discussing their privacy attitudes and concerns. However, they still provide sensitive information such as their income, investments, and home addresses on the Internet without a good reason [46]. People's identity disclosure behavior often does not match their privacy attitudes, privacy concerns, and actions that they claimed to take to protect their identities [47]. They often lack adequate information about protecting their identity information and tend to have a sense of personal immunity to common hazards [48].

III. SUBLIMINAL WARNING STRATEGY AND PILOT STUDY

Subliminal stimuli are defined as occurring below the threshold of conscious awareness [49]. The threshold is defined by whether less than half of people can consciously perceive the stimuli. We conducted a set of pilot studies to guide us on the duration of the subliminal stimuli in our restaurant reservation context and how many times we should display the subliminal prime. The pilot study helped us choose the key design factors to maximize the success rate of presenting a warning message. We tested two types of subliminal stimuli - text messages and icons. The subliminal warnings were presented just in time. For example, after a user clicks the input textbox and right before disclosing information, a subliminal stimulus is shown. In Table 1, we represent the conditions tested in the pilot studies.

Condition one tested subliminal mitigation - the message "Fake it" was used as a subliminal warning message. The message is displayed between a pre-mask image and a post-mask image. The pre-mask was a food image (from the restaurant being viewed) displayed 2000ms before the subliminal prime. The post-mask was a food image displayed 2000ms after the subliminal prime. The subliminal message is displayed five times before the identity input page. The message was shown again on the identity input page when a user clicked the address input box. Subliminal messages were shown for 50ms.

Condition two tested the message "Fake it" only once for 50ms.

Condition three tested the effectiveness of an icon - a yellow triangle (pointed to the left). We used the same pre-mask and post-mask strategy. It is displayed 5 times before an identity input page. The duration was set to 50ms. The icon appeared on the right side of the "Next" button on the identity input page.

Condition four tested the same icon as condition three but the icon displayed at the right side of the address input textbox.

We recruited 22 participants for our pilot study from computing science students and psychology students. Participants were randomly assigned to one of the four conditions. We tested 7 participants in condition one, and 5 participants for each other conditions. Institutional Review Board (IRB) approved the pilot study.

Participants navigated through a restaurant reservation application that requests identity information. The subliminal stimuli automatically appeared in pages or when identity elements were requested.

The percentage of participants who disclosed their identity information is shown in Table 2. From the results of the pilot studies, we can see that a warning design with the subliminal message that shows up once was more effective than other subliminal warning designs. This finding aligns with those of other research projects that focused on semantic processing – that short messages are one of the most effective stimuli [50]among all effective subliminal message forms (i.e., photos, words, signs, and shapes/polygons).

Because the subliminal warnings were displayed during a very short period of time, a concise message is needed to suggest safe behavior. When an app or a website collects a large amount of identity information, we can suggest users to

 ISBN: 978-1-61208-982-9

TABLE 2 PILOT STUDY RESULT: PERCENTAGE OF PARTICIPANTS DISCLOSING PERSONAL INFORMATION

Subliminal Prime	Condition	First name	Last name	Email	Phone Number	Address	Zip code
Message	1	71%	57%	43%	29%	43%	57%
	2	80%	80%	20%	20%	25%	20%
Icon	3	100%	100%	80%	80%	80%	80%
	4	100%	100%	100%	80%	100%	100%

"fake it" or "skip it." Warning words need to be congruent with users' attitudes, beliefs, motivations, and behavior. Research has shown that congruity between a subliminally presented word and a user's goal has a significant influence on behavior [39]. Most people want to keep their information private [51], so there should be congruency between the suggested cautious behavior and attitudes.

More than one strategy may be used to warn users. For example we may suggest the user to "fake it" or "skip it." Considering some of the web applications set most of the personal information fields as required, users can not skip them. In this paper, we decided to present the subliminal warning message "fake it" with a 50 ms display duration, one time, above the requested identity information field.

IV. EXPERIMENTAL DESIGN

We conducted an experiment to evaluate the effectiveness of the subliminal warnings. We used an eye-gaze based verification system with an eye tracker and a scene camera and a post-experiment questionnaire for evaluation.

The design of this experiment has two conditions: the control condition (with no warning) and the subliminal warning condition (with the subliminal message to "fake it", display 50ms on top of the address line one). We target the street address for not disclosure based on an assumption that the street address would be more sensitive than another field (zip code, state, etc) for a table reservation app. We conducted our study to achieve the following goals.

- Ascertain that subliminal warnings are effective.
- Verify and validate that subliminal warnings can be displayed at the right time, right place, and for the proper duration.
- Gain an understanding of participants' identity disclosure behavior under the subliminal warning condition. We wanted to document participants' behavior (e.g., eye gaze, personal information disclosure) when the warning was shown for milliseconds.

A. Participants

We recruited 58 participants on campus for the experiment. 40 were women, and 18 were men. Their ages ranged from forty-eight to seventeen, with a median of twenty one. We list this experiment on a psychology department's website. We prepared a cover story for the experiment as a user evaluation for a restaurant reservation app. Students received activity points as an option toward a course assignment.

B. The Hardware and Software

Because of the short display duration of the subliminal message, we require high-frequency sampling rate for the eye tracking system. The sampling frequency of the scene camera was 120Hz. The sampling frequency of the eye tracker was 250Hz. We ran all software (Eyetracking, scene camera recording, the restaurant scheduling application, and eyetracking analysis) on one machine in the first demo. The high rates of data sampling caused performance degradation (i.e., the sampling intervals were prolonged). Thus, we moved to two PCs that were connected via a wired LAN that was dedicated to the experiment. We executed eye tracking and the restaurant scheduling application on the SMI Server and the scene camera on another laptop computer to guarantee the performance of eye tracking sampling frequency. We also developed new software, which collected and analyzed eye tracking data and synchronized events and eye tracking. Figure 1 shows an integrated software environment that synchronized the eye tracking system, the ReserveME app, the scene camera, and the respective event logs.

Based on this setting, the eye tracking system consistently sampled at 4ms, but the scene camera could not reach the speed of its specification. We sampled 1000 frames each for 6 participants (3 in the control condition and 3 in the subliminal condition) in our pilot study. We generated timestamps for the scene camera videos. Theoretically, the scene camera samples at every 8.3ms (120Hz). The actual sampling rate was on average at 11.24ms (89Hz). Thus, we used the 11.24ms as our frame duration for our analysis of the scene camera videos.

C. The Warning Design

In the restaurant scheduling application, ReserveME, we implemented a subliminal warning that suggested participants

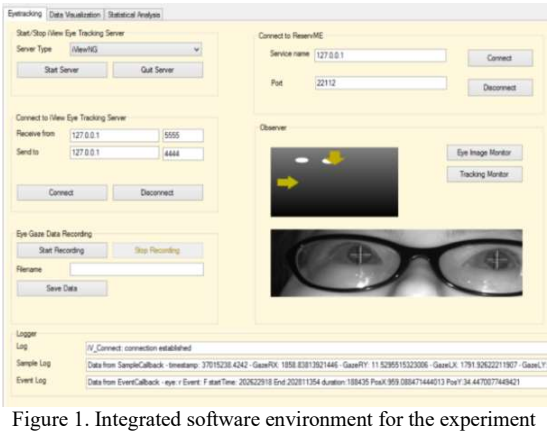

Figure 1. Integrated software environment for the experiment

provide a piece of fake identity information. In our experiment, we used "FAKE IT" as shown in Figure 2. We wanted to display the message just-in-time. That is, when a user clicked on the street address textbox and right before they input their information, a subliminal text message was shown. The subliminal message was set to display for 50ms right above the input textbox, where participants' eye gazes were most likely to be located.

We were specifically interested in six pieces of information: street address, city, state, zip code, email address, and phone number. When zip code, phone number, and email address were requested, the warning message was not shown. (Phone number and email address were requested on the following page of the app.) We wanted to find the impact of subliminal warnings for the targeted behavior – disclosure of street address – as well as for disclosure of the other information.

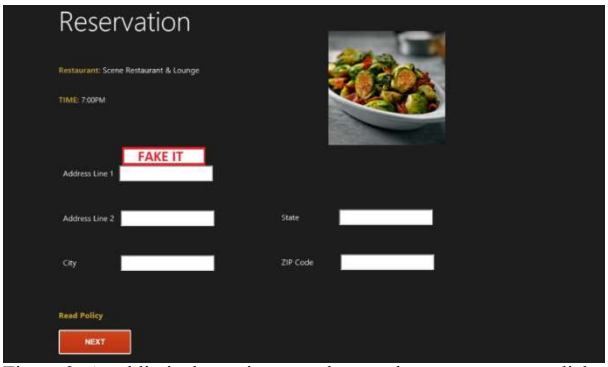

Figure 2. A subliminal warning was shown when a user mouse-clicked the input textbox.

D. Procedure

The participant signed the consent form and disclaimer first. The disclaimer stated that the restaurant reservation app was developed by a third-party software company and the purpose of this study is to evaluate the useability of the software. In the disclaimer, we informed that participants' information would be shared with the software company. This mock disclaimer was used to reduce the university environment bias so that the university researchers would be cautious with participants' information. We did not share their information with any third-party company. A short introduction to the experimental procedure was provided to participants. Participants were told to use the ReserveME app without communication with the researcher unless they had technical questions. A post-experiment survey was given after they reached the last page of the ReserveME app. The entire procedure for each participant took about 30 minutes. Durations varied somewhat for different participants because of eye-tracking calibration or other technical issues. Our university's Institutional Review Board (IRB) approved the experiment procedure.

In the experiment, we did not directly record their identity information. We recorded whether they provided a certain piece of information. In our database of the restaurant reservation app, we coded "1" for participants who provided accurate information and "0" for those not provided accurate information. (The accuracy of the information was obtained from responses participants gave on the post-experiment survey; see below.) Their input of the identity information was discarded as they were typing the information, but participants were not aware of this during the experiment. The eye tracker collects the X, Y coordinates of eye fixation and movement. The scene camera record the change of pages and the pop-up and fade out of the subliminal warning. In our after-experiment debriefing, we told participants that none of their actual information had been saved in the database or would be shared with a third party. We also told them the subliminal warning message was displayed on the address page and the real purpose of the experiment.

E. Post-Experiment Survey

The first part of the post-experiment survey asked questions about participants' experience with the ReserveME app.

In the second part of the survey, we asked participants whether or not they provided accurate information on the identity entry questions in the ReserveME app. We also asked why they provided accurate or inaccurate information. In the third part of the survey, we asked whether they saw anything on the page on a specific spot. We provided a screenshot of the application page to remind them. If they reported yes, we asked them to describe the message they saw.

Then, the focus of the questionnaire shifted to privacy attitudes in the fourth part. We asked participants to rate four pieces of identity elements and whether it was important to keep the identity elements private. We also asked about their concerns related to spam, identify theft, and the like.

Last, we asked participants about their demographic information.

V. Experimental Results

A total of fifty-eight participants were in our two experimental conditions – thirty in the control condition and twenty-eight in the subliminal warning condition.

A. The Display of the Subliminal Warnings

Verified by the data collected by scene camera, the subliminal warning message was displayed in all cases in the subliminal condition. The average display duration of the warning message was 84ms, although we set the display duration for 50ms (see Section 4.2 for how we measured the durations). For twenty-seven participants, the warning message lasted 56ms (5 frames) to 101ms (9 frames). For one participant, the message lasted much longer 135ms (12 frames). This was an unusually long display of the message. The long display was consistent with the event logs of ReserveME and eye tracking data. (The participant did not report seeing the message.) Therefore, the display durations of subliminal warning message may have variations in different runs.

B. Were Participants Consciously Aware of the Warning Message?

Six out of twenty-eight participants in the subliminal warning condition could report the subliminal warning

 ISBN: 978-1-61208-982-9

TABLE 3. SUBLIMINAL WARNING EXPERIMENT RESULTS

	Did not input information				Faked information				Exposed real information			
	Con-trol	Sub-liminal	P-value	Odds ratio	Con-trol	Sub-liminal	P-value	Odds ratio	Con-trol	Sub-liminal	P-value	Odds ratio
Address	1	5	**0.035**	6.30	5 (29)	10 (23)	**0.048**	2.78	24	13	**0.004**	0.22 (4.62)
City	0	2	0.068	∞	3 (30)	7 (26)	**0.065**	3.00	27	19	**0.017**	0.23 (4.26)
State	0	2	0.068	∞	3 (30)	5 (26)	0.192	1.95	27	21	0.065	0.33 (3.00)
Zip code	0	2	0.068	∞	3 (30)	9 (26)	**0.019**	4.26	27	17	**0.004**	0.17 (5.82)
Email	0	4	**0.016**	∞	2 (30)	10 (24)	**0.003**	7.78	28	14	**<0.001**	0.07 (12.1)
Phone	2	5	0.095	3.04	3 (28)	9 (23)	**0.019**	4.26	25	14	**0.003**	0.20 (5.00)

message in one way or another. Three of them recalled the exact action suggested, "FAKE IT;" two remembered the message "falsify" which was semantically correct, and the other one ("red boarder") seemed not to have processed the warning message at the semantic level. Unlike the other five, this participant provided accurate information for all identity elements.

The average display duration of the subliminal warning message for this small group who recalled the warning message was 90ms (8 frames). Next, we ran a t-test to evaluate whether the mean of the number of frames displayed for this group was different from the rest of the participants in the subliminal condition. Those participants who recalled the subliminal warning messages were shown 8.0 frames on average, and those who did not recall the message were shown 7.36 frames on average. Having the p-value of 0.333, we believe there was no difference between the two groups. In addition, we ran six one-way analysis of variance (ANOVA) tests to determine whether the number of frames of the prime impacted its effectiveness on identity exposure of the six elements. None of the tests turned out to be significant.

C. Identity Disclosure Behavior

We analyze participants' behavior between the control and subliminal conditions from three aspects: (a) the number of participants who did not input their information; (b) the number of participants who faked their identity elements when they input; and (c) the number of participants who provided accurate identity information. The sums of the numbers in all three aspects equal the total number of participants. We ran the two proportion Z-tests (left-tailed) to determine whether participants in the subliminal warning condition group behaved statistically differently from the control group in all three aspects and for all six identity elements. Table 3 shows subliminal warning experiment results for the two experimental conditions. Probability values (p-values) less than 5% (bold) indicate statistically significant differences between the control and subliminal message conditions. The numbers in parentheses indicate the number of participants who input information. The numbers in parentheses indicate the odds ratios when we compare the control condition to the subliminal condition (i.e., reciprocals of the odds ratios that compared the subliminal condition to the control condition).

More participants in the subliminal warning condition did not input their information for all six identity elements than those in the control condition. Note that the subliminal warning message was triggered by the mouse click event on the street address field. This indicated that participants in the subliminal condition wanted to input information for the first identity element, but they did not. Two proportion Z-tests show that for street address and email, participants in the subliminal condition behaved statistically different from the control group. For the address field, for example, participants in the subliminal condition were about 6.3 times (i.e., odds ratio) more likely to skip their input in the field.

For participants who input their information, those in the subliminal condition were much more likely to fake their identity elements than those in the control condition. Except for the state, all other five identity elements were statistically different for the subliminal and control conditions. Participants were two to eight times (1.95 – 7.78) more likely to fake their identity information for the subliminal group than the control group.

When we combine the first two aspects (a and b), we have the total effect of the subliminal warning message – the reduction of the real identity disclosure. For instance, thirteen participants provided real street address information in the subliminal condition compared to twenty-four of them in the control condition. The two proportion Z-tests show that the difference was statistically significant. Participants in the subliminal condition had a 22% chance (i.e., odds ratio) to provide their street address when compared to those in the control group. That is, those in the control group were 4.6 times (the reciprocal of 0.22) more likely to provide their real street addresses than those in the subliminal condition. The results (Table 3) show that participants were much less likely to provide their real information in the subliminal condition for all elements except for their state information.

The scale of identity exposure and behavior was very different between the two groups, as shown in Table 3. For example, all except one participant did not provide the street address; all others typed in information in the control group. For the subliminal group, two persons did not type in anything in all four fields and went to the next page. Another three skipped the street address part and moved on to the following fields (city, state, and zip code).

Emails and phone numbers were requested on a subsequent page of the ReserveME app. Participants in the subliminal condition, however, were still much more likely to not provide or to fake the information than those in the control condition. Thus, we have an interesting and important observation: even though "FAKE IT" was suggested once for the street address, it may have impacted the disclosure of identity elements after that.

 ISBN: 978-1-61208-982-9

D. What the Eye Tracking Data Show

We were interested in the eye gaze fixation and the time period when the subliminal warnings were shown. We acquired quality eye tracking data in twenty-four cases in the subliminal condition and did not have quality data for the other four due to the eye tracker's loss of the participants' eye gazes during the display of the subliminal message and by an automated system update which negatively impacted the eye-tracking sampling rates.

In twenty out of twenty-four cases, participants' eye gazes were near the subliminal warning message as shown in the white rectangle area (up to 50 pixels away in X or Y coordinates from the "FAKE IT" message) as shown in Figure 3. This was the most reasonable place for participants' gazes to be located because when a participant needed to click on the textbox, they would look at the textbox. Therefore, we displayed the subliminal warning message at that location and at the moment of input as one of the best bets to be successful. Eye-tracking data confirmed that most participants were looking at or near that location. Figure 3 shows a participant's eye gaze during the display of the subliminal warning message. The colored dots represent eye-gaze locations with 4ms intervals that overlay on top of each (the lighter the color, the more recent the eye gaze location).

Participants behaved differently from the rest in the subliminal condition in three cases. First, their eye gaze locations were not on or near the street address textbox. It seemed that they looked at the textbox and moved their mice over the textbox a few seconds prior. They looked elsewhere and then clicked on the field. Thus, they missed the warning message. Two of them provided real identity elements for all the six pieces requested. One of the participants perhaps noticed a message; about 80ms after the subliminal message, his/her eye gaze moved right to the location where the subliminal message was displayed and looked at the area for about 300ms.

In the other case, there seemed to be a significant delay (about 120ms) between the time when the participant clicked the street address textbox and the display of the subliminal message. When the subliminal message was displayed, the participant had already been looking at the keyboard and typing. This participant provided real information for all identity elements.

Learning from these cases, we may display the warning when the mouse is over the street address textbox. Such an approach needs to solve the issues that a user is not looking at the field but the mouse happens to move over or stop on the specific location. Further research and experiments are needed to improve these mechanisms. In addition, we will evaluate the display of the subliminal message within the input textbox in future experiments.

E. How Participants Explained their Disclosure Behavior

In the questionnaire, we asked participants why they provided truthful or false identity information during their interaction with the app. Their 72 responses were classified into one or more of nine categories of explanations shown in Table 4. We use thematic analysis for the classification of this qualitative data. We took the bodies of the data and then groups them according to similarities.

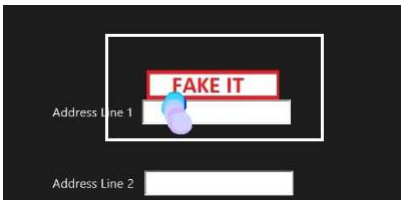

Figure 3. A participant's eye gaze information when a subliminal warning shows up

TABLE 4. PARTICIPANTS' EXPLANATIONS OF WHY THEY DID OR DID NOT PROVIDE THEIR PERSONAL INFORMATION.

Reasons for Providing Truthful Information		
	Control	**Subliminal**
Information is not sensitive or it is not risky to disclose.	7	11
Trust in the app or experimental context.	12	6
Information is needed by the app or experiment.	4	2
Information is already known.	2	2
Habit.	4	0
Reasons for Falsifying Personal Information		
	Control	**Subliminal**
Information is sensitive or risky to disclose.	4	11
Distrust the app/experimental context.	1	1
Information is not needed by the app/experiment.	1	1
Told to "fake it."	0	3

Some participants gave explanations fitting into more than one category. The most frequently occurring categories (n = 33 responses; 7 + 11 giving being truthful; 4 + 11 falsifying) used to explain disclosure (or non-disclosure) were related to the sensitivity (i.e., riskiness) or lack of sensitivity of the information. For example, "I feel uncomfortable sharing this information" or "this information is safe to give." The second most frequent reason for disclosure was related to trust/distrust of the app or the experimental context (n = 20 responses), such as "I trust apps like this." Note that only half as many participants in the warning condition, compared to the control condition, indicated that they trusted the app/ experiment. Eight participants indicated that they thought the information was (or was not) needed by the app or experimenter. Four participants indicated that they disclosed out of habit. Another four explained that they thought the information was already (or readily) available to the researcher or app. Three participants said they falsified information because the app told them to "fake it."

F. The Relationships among Behavior, Privacy Attitudes, Concerns, and the Subliminal Warning

People's privacy attitudes, concerns, and other unknown factors may all affect their privacy disclosure behavior. Having the data of participants' privacy attitudes and concerns from the questionnaire and our experimental data, we

analyzed the relationships between these factors and participants' behavior.

We created indices from the data in the following way: a) a disclosure index that combined participants' disclosure behavior by giving weights of 1 to a street address, email, and phone number, and weights of 0.5, 0.3, and 0.2 to zip code, city, and state, respectively; The weights were assigned based on the sensitivity of different field. (state is much less risky to be disclose than street address) b) an attitude index that combined participants' ratings of four attitude questions; c) a concern index that combined participants' ratings of six related questions; and d) dummy coded 1 for eye gaze at or near the subliminal message, and 2 for other cases including for participants in the control condition. These indices were used in multiple regression models for the relations between the predictors and the disclosure behavior.

In the full linear regression model that included all variables, the eye gaze location had very strong correlations with experiment conditions (-0.861) and the number of frames of the warning message (-0.886). Due to this high collinearity, we ran a reduced model with just eye gaze location, experiment condition, or the number of frames. The models

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.581[a]	.338	.298	1.33455

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	45.393	3	15.131	8.496	.000[b]
	Residual	89.051	50	1.781		
	Total	134.444	53			

Coefficients

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	2.406	1.295		1.858	.069
	DistanceToSubliminalMsg	1.543	.386	.475	4.003	.000
	Attitudes	-.091	.065	-.175	-1.403	.167
	Concerns	-.042	.039	-.132	-1.086	.283

Figure 4. Liner regression results. The relationships between behavior, privacy attitudes, concerns, and eye gaze locations.

respective R squares were 0.338, 0.286, and 0.263. Thus, we used eye gaze location for the reduced linear regression model as shown below.

$$Disclosure = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_x x_3$$

$$where\ x1 = \text{“Eye gaze location”}$$

$$x2 = \text{“Attitudes”}$$

$$x3 = \text{“Concerns”}$$

The linear regression results are shown in Figure 4. Compared to the full model, the adjusted R squares for this reduced model increased from 0.275 to 0.298. The eye gaze location was a statistically significant factor in predicting participants' behavior, but attitudes and concerns were not significant factors in predicting disclosure in this context.

VI. LIMITATIONS

Like any other experimental study, our experiment has its limitations. First, we used a lab environment to study whether subliminal warnings reduce identity disclosure. We are keenly aware of the questions raised regarding the realism of the method and the generalizability of experiments. Experiments, however, provide the best control over factors that might influence disclosure. They are the most effective way to test our hypotheses by studying participants' behavior directly when warning messages are shown for dozens of milliseconds. They allow us to measure the duration of the subliminal messages and participants' eye gazes.

Second, our ReserveME app collects identity information and also displays warnings. This does not seem logical in a real-world application, but the approach avoids further complication of our software and does not affect the proof of concept of subliminal warnings. In a commercial implementation, the functionality of subliminal warnings may be realized by operating systems or as a web browser plug-in (similar to the autofill function in browsers to identify form fields).

Third, our experiment was run in a university lab setting. Often, participants trust researchers in this type of setting, thus challenging privacy research. People may believe that researchers will not put them at risk. It was the case for our experiment, as shown in Table 2. Given the trusting nature of our research participants, however, warnings that do decrease disclosure of private information will likely be effective in a lower trust real-world environment.

Last, our participants were college students (most of them were between eighteen and twenty-two). Thus, our results might have limited generalizability. This age group, however, may be one of the most representative age groups who use apps to make reservations and purchases online. In our future work, we plan to study participants with more diverse backgrounds and age groups.

VII. CONCLUSION AND FUTURE WORK

Instead of reminding users of their privacy and security attitudes as we did in a previous study using the word “privacy” [15], our primary goal and contribution of this research were to design a subliminal warning by suggesting safe behavior using stimulus–response mapping model. When an app or a website collects a large amount of identity information, we can suggest users to “fake it.” This design bypassed the upper stages and emphasized the last level of C-HIP model. Cognitively more straightforward to trigger the safe behavior (not provide personal information). We conducted a pilot study and tested two categories of warnings (message and icon) to guide the design of the subliminal warning. We used eye tracking and scene camera recording to verify the display duration of the subliminal warning and users' attention during the experiment. The result of the experiment showed that the subliminal warning with the suggested response could effectively reduced disclosure of identity information.

This paper proposed the basic idea by focusing on scientific theory and evaluations. The subliminal warnings

 ISBN: 978-1-61208-982-9

could be implemented as an application solution to motivate safety behavior. Part of our future work is to implement the application solution of the subliminal warning. We envision that an application implementation of subliminal warning could be developed as a third-party application or as a web browser plug-in.

We learned several lessons and limitations of our study. The application of the strategy that uses mouse click events to trigger the warning display may not work for all types of user behavior. System delays might also have an impact on the warning effectiveness. In this experiment, we evaluated only one display duration. We will study the warning effectiveness with regard to different display durations of the subliminal messages. Moreover, future experiments may examine different warning words and their locations.

It is also quite interesting that, in the pilot study, condition two obtained better results than condition one. The warning was presented five times in condition one, while the warning was only displayed once in condition two. It could be because of the pre-mask and post-mask in condition one. We added them to make sure the warning satisfies the subliminal threshold. In our future work, we plan to extend the experimental study with multiple times displays of subliminal warnings, different warning words, duration, colors, and background of the message. We will also apply other statistical analyses such as the omnibus test to test different parameters and Bonferroni correction to mitigate family-wise errors.

Our ongoing research is to develop a framework for non-conscious security warnings. We want to discover other effective subliminal warning strategies. We want to investigate how these strategies facilitate users' memory access, remind them of their security and privacy attitudes, and motivate them to take safe actions. Another goal is to compare the effectiveness of the different strategies and to find their limitations.

REFERENCES

[1] J. Sobey, R. Biddle, P. C. van Oorschot, and A. S. Patrick, "Exploring user reactions to new browser cues for extended validation certificates," in *European Symposium on Research in Computer Security*, 2008, pp. 411–427.

[2] A. Wiesmann, A. Stock, M. Curphey, and R. Stirbei, "A guide to building secure web applications and web services," *The Open Web Application Security Project*, 2005. http://www.i-pi.com/Training/BSS/OWASPGuide2.0.1.pdf (accessed May 19, 2022).

[3] K. Krol, M. Moroz, and M. A. Sasse, "Don't Work. Can't Work? Why It's Time to Rethink Security Warnings," *Risk and security of internet and systems (CRiSIS), 2012 7th International conference*, pp. 1–8, 2012.

[4] A. Vance, D. Eargle, J. L. Jenkins, C. B. Kirwan, and B. B. Anderson, "The fog of warnings: how non-essential notifications blur with security warnings," in *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*, 2019, pp. 407–420.

[5] S. Carpenter, F. Zhu, M. Zeng, and M. Shreeves, "Expert sources in warnings may reduce the extent of identity disclosure in cyber contexts," *International Journal of Human--Computer Interaction*, vol. 33, no. 3, pp. 215–228, 2017.

[6] M. S. Wogalter, "Communication-Human Information Processing(C-HIP)Model," in *Handbook of Warnings*, M. Wogalter, Ed. Mahwah,NJ, USA, 2006, pp. 51–62.

[7] B. Anderson, T. Vance, B. Kirwan, D. Eargle, and S. Howard, "Users aren't (necessarily) lazy: using NeuroIS to explain habituation to security warnings," *Thirty Fifth International Conference on Information Systems*, pp. 1–15, 2014.

[8] D. Akhawe and A. P. Felt, "Alice in warningland: a large-scale field study of browser security warning effectiveness," *Proceedings of the 22nd USENIX Security Symposium*, pp. 257–272, 2013.

[9] C. Bravo-Lillo, L. F. Cranor, J. Downs, and S. Komanduri, "Bridging the gap in computer security warnings: A mental model approach," *IEEE Security and Privacy*, vol. 9, no. 2, pp. 18–26, 2011, doi: 10.1109/MSP.2010.198.

[10] S. Dehaene, J. P. Changeux, L. Naccache, J. Sackur, and C. Sergent, "Conscious, preconscious, and subliminal processing: a testable taxonomy," *Trends in Cognitive Sciences*, vol. 10, no. 5, pp. 204–211, 2006, doi: 10.1016/j.tics.2006.03.007.

[11] J. Kihlstrom, "The Congnitive Unconscious," *Science (1979)*, vol. 237, pp. 1445–1452, 1987.

[12] F. Zhu, S. Carpenter, and M. Zeng, "Subliminal Warnings: Utilizing the High Bandwidth of Nonconscious Visual Perception," in *17th International Conference PERSUASIVE,* , Mar. 2022, pp. 255–271.

[13] R. M. Montoya, R. S. Horton, J. L. Vevea, M. Citkowicz, and E. A. Lauber, "A re-examination of the mere exposure effect: The influence of repeated exposure on recognition, familiarity, and liking.," *Psychological Bulletin*, vol. 143, no. 5, pp. 459–498, 2017, doi: 10.1037/bul0000085.

[14] E. Harrell and L. Langton, "Victims of identity theft, 2014," *U.S. Department of Justice*, no. December, p. 27, 2015.

[15] S. Spiekermann, J. Grossklags, and B. Berendt, "E-privacy in 2nd generation E-commerce: privacy preferences versus actual behavior," in *Proceedings of the 3rd ACM conference on Electronic Commerce*, 2001, pp. 38–47.

[16] V. C. Conzola and M. S. Wogalter, "A Communication – Human Information Processing (C – HIP) approach to warning effectiveness in the workplace," *Journal of Risk Research*, vol. 4, no. 4, pp. 309–322, 2001, doi: 10.1080/1366987011006271.

[17] M. Lehto and J. Miller, "Warning volume I: Fundamentals, design, and evaluation methodologies," *Ann Arbor MI: Fuller Technical Publications*, vol. I, no. Warning, 1986.

[18] M. R. Lehto, "A proposed conceptual model of human behavior and its implications for design of warnings.," *Percept Mot Skills*, vol. 73, pp. 595–611, 1991.

[19] C. Bravo-Lillo, L. F. Cranor, J. Downs, and S. Komanduri, "Bridging the gap in computer security warnings: A mental model approach," *IEEE Security and Privacy*, vol. 9, no. April, pp. 18–26, 2011, doi: 10.1109/MSP.2010.198.

[20] M. S. Wogalter, R. J. Sojourner, and J. W. Brelsford, "Comprehension and retention of safety pictorials," *Ergonomics*, vol. 40, no. 5, pp. 531–542, 1997.

[21] D. Holender, "Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal," *Behavioral and brain Sciences*, vol. 9, no. 01, pp. 1–23, 1986.

[22] D. H. Nguyen, A. Kobsa, and G. R. Hayes, "An empirical investigation of concerns of everyday tracking and recording technologies," in *Proceedings of the 10th International*

 ISBN: 978-1-61208-982-9

Conference on Ubiquitous Computing, 2008, pp. 182–191. doi: 10.1145/1409635.1409661.
[23] B. Schneier, *Secrets and lies: digital security in a networked world*. John Wiley & Sons, 2011.
[24] P. Sumner, P.-C. Tsai, K. Yu, and P. Nachev, "Attentional modulation of sensorimotor processes in the absence of perceptual awareness.," *Proc Natl Acad Sci U S A*, vol. 103, no. 27, pp. 10520–10525, 2006, doi: 10.1073/pnas.0601974103.
[25] A. Amran, Z. F. Zaaba, and M. K. Mahinderjit Singh, "Habituation effects in computer security warning," *Information Security Journal: A Global Perspective*, vol. 27, no. 2, pp. 119–131, 2018.
[26] C. Bravo-lillo and et al., "Your attention please: Designing security-decision UIs to make genuine risks harder to ignore," in *SOUPS '13 Proceedings of the Ninth Symposium on Usable Privacy and Security*, 2013, pp. 1–18. doi: 10.1145/2501604.2501610.
[27] M. R. Lehto, "Human Factors Models," in *Handbook of Warnings*, M. S. Wogalter, Ed. Mahwah, NJ, USA: Lawrence Erlbaum Associates, 2006, pp. 63–88.
[28] A. P. Felt, R. W. Reeder, H. Almuhimedi, and S. Consolvo, "Experimenting at scale with google chrome's SSL warning," *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, pp. 2667–2670, 2014, doi: 10.1145/2556288.2557292.
[29] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F. Cranor, "Crying wolf: An empirical study of SSL warning effectiveness," in *18th USENIX Security Symposium*, 2009, pp. 399–432. doi: 10.1016/S1353-4858(01)00916-3.
[30] A. P. Felt *et al.*, "Improving SSL warnings: Comprehension and adherence," in *Conference on Human Factors in Computing Systems - Proceedings*, Apr. 2015, vol. 2015-April, pp. 2893–2902. doi: 10.1145/2702123.2702442.
[31] D. Akhawe and A. P. Felt, "Alice in warningland: a large-scale field study of browser security warning effectiveness," in *Proceedings of the 22nd USENIX Security Symposium*, 2013, pp. 257–272.
[32] S. Egelman and S. Schechter, "The importance of being earnest [in security warnings]," in *International Conference on Financial Cryptography and Data Security*, 2013, pp. 52–59.
[33] F. Raja, K. Hawkey, S. Hsu, K.-L. Wang, and K. Beznosov, "A brick wall, a locked door, and a bandit: a physical security metaphor for firewall warnings," 2011. doi: http://doi.acm.org/10.1145/2078827.2078829.
[34] P. Domingues, L. Andrade, and M. Frade, "A Digital Forensic View of Windows 10 Notifications," *Forensic Sciences*, vol. 2, no. 1, pp. 88–106, Jan. 2022, doi: 10.3390/forensicsci2010007.
[35] M. Zeng, F. Zhu, and S. Carpenter, "Eye Gaze Based Dynamic Warnings," *9th International Conference on Advances in Computer-Human Interactions (ACHI)*, vol. i, no. c, pp. 204–211, 2016.
[36] E. K. Cressman, M. Y. Lam, I. M. Franks, J. T. Enns, and R. Chua, "Unconscious and out of control: Subliminal priming is insensitive to observer expectations," *Consciousness and Cognition*, vol. 22, no. 3, pp. 716–728, 2013, doi: 10.1016/j.concog.2013.04.011.
[37] L. Smarandescu and T. A. Shimp, "Drink coca-cola, eat popcorn, and choose powerade: testing the limits of subliminal persuasion," *Marketing Letters*, vol. 26, no. 4, pp. 715–726, 2015, doi: 10.1007/s11002-014-9294-1.
[38] R. Radel, P. Sarrazin, P. Legrain, and L. Gobancé, "Subliminal priming of motivational orientation in educational settings: Effect on academic performance moderated by mindfulness," *Journal of Research in Personality*, vol. 43, no. 4, pp. 695–698, 2009.
[39] E. J. Strahan, S. J. Spencer, and M. P. Zanna, "Subliminal priming and persuasion: Striking while the iron is hot," *Journal of Experimental Social Psychology*, vol. 38, no. 6, pp. 556–568, 2002, doi: 10.1016/S0022-1031(02)00502-4.
[40] P. M. Merikle and M. Daneman, "Psychological investigations of unconscious perception," *Journal of consciousness studies*, vol. 5, no. 1, pp. 5–18, 1998.
[41] S. Kouider and S. Dehaene, "Levels of processing during non-conscious perception: a critical review of visual masking.," *Philos Trans R Soc Lond B Biol Sci*, vol. 362, no. 1481, pp. 857–75, 2007, doi: 10.1098/rstb.2007.2093.
[42] H. F. Sperdin, L. Spierer, R. Becker, C. M. Michel, and T. Landis, "Submillisecond unmasked subliminal visual stimuli evoke electrical brain responses," *Human Brain Mapping*, vol. 36, no. 4, pp. 1470–1483, 2015, doi: 10.1002/hbm.22716.
[43] H. Reuss, A. Kiesel, W. Kunde, and P. Wühr, "A cue from the unconscious - masked symbols prompt: Spatial anticipation," *Frontiers in Psychology*, vol. 3, no. OCT, pp. 1–10, 2012, doi: 10.3389/fpsyg.2012.00397.
[44] F. Schlaghecken and M. Eimer, "Masked prime stimuli can bias 'free' choices between response alternatives.," *Psychon Bull Rev*, vol. 11, no. 3, pp. 463–468, 2004, doi: 10.3758/BF03196596.
[45] M. Rogers and K. H. Smith, "Public perceptions of subliminal advertising: Why practitioners shouldn't ignore this issue," *Journal of Advertising Research*, vol. 33, no. 2, pp. 10–19, 1993.
[46] S. Spiekermann, J. Grossklags, and B. Berendt, "E-privacy in 2nd Generation E-Commerce: Privacy Preferences versus actual Behavior," 2001.
[47] F. Zhu, S. Carpenter, and A. Kulkarni, "Understanding Identity Exposure in Pervasive Computing Environments," *Pervasive and Mobile Computing*, vol. 8, no. 5, pp. 777–794, 2012.
[48] J. R. Bettman, J. W. Payne, and R. Staelin, "Cognitive considerations in presenting risk information," *Learning about Risk–consumer and worker responses to hazard information*, pp. 13–41, 1987.
[49] M. Bar and I. Biederman, "Subliminal Visual Priming," *Psychological Science*, vol. 9, no. 6, pp. 464–468, 1998, doi: 10.1111/1467-9280.00086.
[50] R. M. Montoya, R. S. Horton, J. L. Vevea, and E. A. Lauber, "A Re-Examination of the Mere Exposure Effect : The Influence of Repeated Exposure on Recognition , Familiarity , and Liking," vol. 143, no. 5, pp. 459–498, 2017.
[51] M. S. Ackerman, L. F. Cranor, and J. Reagle, "Privacy in e-commerce: examining user scenarios and privacy preferences," in *Proceedings of the 1st ACM conference on Electronic commerce*, 1999, pp. 1–8.

 ISBN: 978-1-61208-982-9