

The Emperor's New Security Indicators

An evaluation of website authentication and the effect of role playing on usability studies^{*†}

Stuart E. Schechter Rachna Dhamija

MIT Lincoln Laboratory

Harvard University &
CommerceNet

Andy Ozment

MIT Lincoln Laboratory &
University of Cambridge

Ian Fischer

Harvard University

To be presented May 20-23, 2007 in Oakland, California at:
The 2007 IEEE Symposium on Security and Privacy.

©2007 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Abstract

We evaluate website authentication measures that are designed to protect users from man-in-the-middle, 'phishing', and other site forgery attacks. We asked 67 bank customers to conduct common online banking tasks. Each time they logged in, we presented increasingly alarming clues that their connection was insecure. First, we removed HTTPS indicators. Next, we removed the participant's site-authentication image—the customer-selected image that many websites now expect their users to verify before entering their passwords. Finally, we replaced the bank's password-entry page with a warning page. After each clue, we determined whether participants entered their passwords or withheld them.

We also investigate how a study's design affects participant behavior: we asked some participants to play a role and others to use their own accounts and passwords. We also presented some participants with security-focused instructions.

We confirm prior findings that users ignore HTTPS indicators: no participants withheld their passwords when these indicators were removed. We present the first empirical investigation of site-authentication images, and we find them to be ineffective: even when we removed them, 23 of the 25 (92%) participants who used their own accounts entered their passwords. We also contribute the first empirical evidence that role playing affects participants' security behavior: role-playing participants behaved significantly less securely than those using their own passwords.

^{*}This research was supported in part by the Center for Research on Computation and Society at Harvard University.

[†]This work is sponsored by the I3P under Air Force Contract FA8721-05-0002. Opinions, interpretations, conclusions and recommendations are those of the author(s) and are not necessarily endorsed by the United States Government. This work was produced under the auspices of the Institute for Information Infrastructure Protection (I3P) research program. The I3P is managed by Dartmouth College, and supported under Award number 2003-TK-TX-0003 from the U.S. Department of Homeland Security, Science and Technology Directorate. Points of view in this document are those of the authors and do not necessarily represent the official position of the U.S. Department of Homeland Security, the Science and Technology Directorate, the I3P, or Dartmouth College.

1. Introduction

We conducted a study with two parallel research goals. First, we wanted to evaluate how effectively website-authentication indicators protect users from fraudulent sites. More fundamentally, we wanted to investigate whether participants behave differently in security usability studies than they would in real life. We evaluated two artificial conditions common to security usability studies: the use of role playing and the presentation of security instructions.

1.1. Evaluating website authentication indicators

Growing incidents of online fraud have spurred new ideas for strengthening web authentication. In response, web browsers are introducing new user interfaces to display security warnings [7]. Browser developers are also working to standardize web security interfaces [9, 14, 19].

The FFIEC, a US financial services regulatory body, has even issued guidance that requires financial institutions to strengthen their authentication mechanisms [15]. In response, many financial institutions are rushing to deploy supplementary authentication mechanisms.

In one approach, customers select an image to be displayed when they login to the site. The login process is then separated onto two distinct pages. On the first page the website presents a form through which the customer enters and submits her username. If the customer is using a client from which she has not previously logged into the website, she may then be asked to answer a challenge question. On the second page, the website presents the customer's chosen site-authentication image above the password-entry field. Customers are instructed to verify the presence of their chosen image before entering their password. If customers fail to verify the site-authentication image, it cannot provide any security benefit. (We illustrate login processes that employ site-authentication images in Appendix A.)

Among the institutions adopting this approach are ING Direct [10], Bank of America [3], Vanguard [18] and Ya-

hoo [24]. These institutions hope that new security features will assuage the fears of regulators and customers alike and bring more customers online. However, a critical question remains to be answered: do website-authentication indicators actually help customers to detect attacks and protect their passwords?

We address the following questions of password security in online banking:

- Will customers of an online bank enter their passwords even if their browsers' HTTPS indicators are missing?
- Will customers of an online bank enter their passwords even if their site-authentication images are missing?
- Will customers of an online bank enter their passwords even if they are presented with an IE7 warning page?

1.2. Designing realistic security usability studies

In designing our study of website authentication indicators, we tried to realistically simulate the conditions that a user would experience during an attack. A study will not be *ecologically valid* unless participants behave as they would in the real-world [4].

One real-world condition that is difficult to replicate in an experimental environment is the experience of risk. Many studies ask participants to assume the role of someone else to avoid exposing participants to real risks. In these *role-playing* scenarios, the consequences of behaving insecurely are borne by the fictional role, not by the participants themselves. Until now, no studies have tested whether participants playing roles behave as securely as they do when they are personally at risk.

To create a realistic experience, study designers need to create realistic scenarios and goals: in real life, security is rarely a user's primary goal. Participants in a usability study may not behave realistically if they are told, or can infer, that security is the focus of the study. However, it is often difficult to conceal the focus of the study. Researchers may need to provide participants with the training or knowledge required to behave securely, especially if new security features are being tested.

Security usability researchers frequently debate whether it is possible to replicate the real-world experience of being attacked in an ethical way [1, 13]. Researchers are obligated to minimize risks to participants: they must take great care to protect sensitive information, such as usernames and passwords, that was used or collected during the study. Researchers may also be obligated to inform participants of risks and obtain their informed consent prior to the study. Thus, there is a tension between the requirement to obtain informed consent from participants and the desire for participants to perceive realistic risk.

We constructed our study to address the following questions of study design:

- Do participants behave less securely when playing a role than when the risk is their own?
- Do participants behave more securely when they are informed that security is a focus of the study?
- Can we ethically replicate the experience of a real attack in a usability study?

Roadmap

We begin with a discussion of our study design in Section 2 and participant demographics in Section 3. Next, we describe the study infrastructure and attack clues in Section 4. We present the study results in Section 5 and a discussion of the results in Section 6. Finally, we describe related work in Section 7 and summarize our conclusions in Section 8.

2. Study Design

2.1. Goals and overview

We asked 67 customers of a single bank to conduct common online banking tasks. As they logged in, we presented them with increasingly conspicuous visual clues that indicate a site-forgery attack.

One goal of the study was to investigate whether participants' security behavior is affected by the type of instructions and the type of risk in a security study. We used a between-subjects design, where the participants were divided into three groups; 19 participants were instructed to play a role and to login using the credentials of that role. 20 participants used the same role-playing scenario and were also given additional instructions to behave securely. 28 participants were required to complete tasks by logging into their own bank accounts.

2.2. Ethical guidelines

Ethical guidelines are of particular concern in this study, because we ask participants to perform tasks using their own account information. Our study protocol was jointly reviewed and approved by the institutional review boards of Harvard University and MIT [16, 17]. One strict rule was at the core of our study design: participants must only be deceived in ways that cause them to believe they are *less* secure than they actually are. In addition, we took the following steps to ensure that participants were aware of risks and that these risks could be minimized.

- Our consent form notified participants that we would be observing their actions. (To obscure the purpose of

Group	Name	Key characteristics
1	Role playing	Played a role, given no indication that security is focus of study
2	Security primed	Played a role, told that their role was concerned about security
3	Personal account	Used their own account, given no indication that security is focus of study
1U2	All role playing	The union of groups 1 & 2: all of the participants who played a role

Table 1. Participants were assigned to one of three groups.

the study, we did not detail that we were specifically observing password behavior.)

- Our observation system did not record user IDs, passcodes, or other private information.
- We did not introduce risks to participants beyond those inherent to accessing their bank from a university-managed computer. We took additional technical precautions to protect sensitive information revealed by participants during study tasks, as discussed in Section 2.4.3.
- At the end of the study, we provided participants with a debriefing that explained the purpose of the study, the attack clues that we had presented, the precautions we had taken, and how participants could protect themselves from real site-forgery attacks in the future.

2.3. Participant recruitment

We recruited participants by posting and circulating flyers on and around the Harvard University campus. The flyers offered participants the opportunity to “earn \$25 and help make online banking better” and listed the requirements for participating in the study: we selected a single bank and only accepted participants who banked online regularly at this bank.

Participants were also required to be familiar with both the Windows operating system and the Internet Explorer browser. We chose a single operating system (Windows XP) and browser (Internet Explorer version 6) because different browsers provide different security warnings and indicators.

2.4. Study procedure

The study was conducted in a Harvard classroom building. After verifying that participants met our requirements, we asked them to sign a consent form and to complete a demographic survey.

Each participant was placed alone in a private classroom. We seated participants in front of a laptop computer where nobody, including the study facilitators, could observe the details on the screen while participants completed study tasks. Each laptop was equipped with a mouse, wireless network card, and identical installations of Windows XP SP2 and Internet Explorer version 6.

2.4.1 Group assignment and scenarios

We assigned participants to groups using a weighted round-robin algorithm. If two or more acquaintances arrived at once, round-robin assignment ensured they would be assigned to different groups. Table 1 summarizes the differences between the three groups.

Participants in Group 1, the role playing group, were instructed to assume the role of a named individual and to conduct tasks on behalf of that role. These participants used test accounts that we created for the study. They received the following instructions, with no indication that security was a focus of our study.

Imagine that you are [role name], a medical doctor who has an account at [bank name]. You often use this account to transfer money to your retirement plan. You are at home on a Sunday afternoon and decide to tackle a number of banking errands. All of your bank branches are closed, so you decide to access [bank name]’s online banking web site.

Participants in Group 2, the security primed group, were also instructed to play a role. The instructions provided to this group were identical to those in the first group, with one exception: an extra paragraph indicated that they were playing the role of someone who was concerned about the security of his password.

As [role name], you chose [bank name] because it advertises additional security features. Control over your account is protected by a passcode (also known as a password), and you want to ensure that this passcode doesn’t fall into the wrong hands.

Participants in the first two groups were also given a second sheet containing all of the information they needed to login and to complete tasks. This information included the role’s name, online ID (username), password, site-authentication image, three challenge questions, and answers to those challenge questions.

Participants assigned to Group 3, the personal risk group, were asked to perform tasks using their own bank account. These participants, like those in the first group, were given no indication that security was a factor in the study. Instead,

<i>Task</i>	<i>Information to look up</i>	<i>Answer derived from</i>	<i>Attack clue</i>
1	Number of ATMs in zipcode		N/A (no login required)
2	Account balance	Number of pennies	N/A
3	Time of last login	Minutes	HTTPS indicators removed
4	Date of last posted transaction	Day of month	Site-authentication image removed
5	Date of last statement	Day of month	IE7 warning page

Table 2. In the last four tasks, we asked participants to report non-sensitive account information. We presented attack clues during the last three tasks.

their incentive to behave securely was that their own account information, including their username and password, would appear to be at risk. Their instructions simply began:

We will now ask you to perform five online banking tasks at [your bank's] web site.

Upon receiving their instructions, participants in all groups were allowed to ask us questions. However, we instructed participants that after this point, facilitators would not be able to provide assistance or to answer questions about the study tasks.

2.4.2 Study tasks

We asked participants to complete five online banking tasks. We presented instructions for each task its own sheet of paper. Participants had to complete each task before receiving instructions for the next task, and they were not permitted to return to previous tasks. Each task sheet asked participants to report the information they retrieved, how difficult the task was to complete, and any difficulties that were encountered. These questions were designed to focus participants on the tasks, rather than on the process of logging in.

The first task was designed to re-familiarize participants with the bank website: participants were asked to identify the number of ATMs within a given zipcode.

For the second task, we asked participants to look up their account balance. This task was designed to re-familiarize participants with the process of securely logging into online banking: no attack clues were presented, and no security indicators were modified or removed. Because participants had not previously logged into their account from this computer, the bank's website presented to them with a single challenge question.

To avoid collecting sensitive account data, we asked participants to report non-sensitive information *derived from* account data; we identified the least sensitive numerical component of account data and asked participants to report whether the value was odd or even. For example, in the second task we asked participants to report whether the number of pennies in their account balance was even or odd. Table 2 lists the information requested for each task and the value from which the even/odd answer was derived.

The purpose of the three remaining tasks was to determine how participants would respond to increasingly conspicuous attack clues; these clues indicated the connection from their browser to the bank's password-entry page was insecure and potentially compromised.

We presented these clues to participants via existing security indicators and warnings that are supposed to protect users from site-forgery attacks, such as man-in-the-middle attacks or 'phishing' sites. These attack clues, described in detail in Section 4, were injected by an HTTP proxy that resided on the same laptop as the browser used by the participant.

We presented the tasks and clues to participants in the same order. This static ordering allows us to measure the effect of group assignment on security behavior. However, order-induced biases prevent us from comparing the efficacy of the clues themselves.

2.4.3 Study infrastructure

We configured Internet Explorer to route all HTTP (but not HTTPS) traffic through a proxy that ran locally, on the same computer. The proxy recorded the domain name and path of all pages loaded via HTTP. The proxy also recorded whether the participant submitted a passcode from the bank's password-entry page. The proxy did not record the passcode itself.

The proxy was also used to present attack clues. During all three tasks in which clues were presented, the proxy prevented the activation of HTTPS. Thus, the participant's browser appeared to be connected to the password-entry page insecurely. Preventing the activation of HTTPS also made it possible for the proxy to rewrite pages, which enabled us to present the final two clues.

Our ethical guidelines mandated that the connection from the computer to the bank's web site only *appear* to be insecure. While subjects were shown evidence that their connection to the password-entry page might be insecure and even under attack, we ensured that all available security measures were actually in place. Though hidden from our participants, our proxy always connected to the online banking site using HTTPS and verified the authenticity of the site's certificate. The only link in the connection that

took place over HTTP was the connection from the browser to the proxy, and this internal connection was not exposed to the network.¹

2.4.4 Post-task questionnaire and debriefing

We provided a post-task questionnaire to verify that the behaviors observed by our proxy were consistent with the behaviors recalled by the participants. Finally, we debriefed participants and offered to answer any questions they might have. After the debriefing, participants were paid the \$25 participation fee.

3. Participant Demographics

3.1. Demographics

67 people participated in our study. 40 participants (60%) were male, and 27 participants (40%) female. 35 participants (52%) were age 18-21, 11 participants (16%) were age 22-25, 14 participants (21%) were 26-30, 6 participants (9%) were 31-40, and 1 participant (1%) was 61 or over. No participants were colorblind.

61 participants (91%) were part-time or full time university students. None of the students were pursuing degrees in computer science or engineering. Of the students, 39 (64%) were undergraduates and 22 (36%) were graduate or professional students. All 6 of the non-students worked in non-technical occupations, ranging from lawyers to clergy members.

54 participants (81%) reported using Windows XP as their primary operating system, 12 participants (18%) use Mac OS X, and 1 participant (2%) uses Windows 2000.² Of the 13 participants who did not use Windows XP as their primary operating system, 6 (46%) used Windows XP as their secondary operating system.

28 participants (42%) reported using Microsoft Internet Explorer as their primary browser, 30 participants (45%) use Mozilla Firefox, 7 participants (10%) use Apple Safari, 1 participant (2%) uses Opera, and 1 participant (2%) uses an unspecified browser. Of the 39 participants who did not use Internet Explorer as their primary browser, 28 (72%) use Internet Explorer as their secondary browser.

23 participants (34%) were online banking customers of the bank for less than six months, 13 participants (19%) for six months to a year, 18 participants (27%) for one to two years, and 13 participants (19%) for more than two years.

¹In order to compromise the connection between the browser and the local proxy, an attacker would need administrative access to the computer: he or she could then just as easily install a key-logger or other spyware.

²Totals may be larger than 100% due to rounding.

3.2. Excluded participants

67 participants met our recruitment criteria and were included in the above demographics, but 21 people did not and were excluded from the study. Potential participants were excluded for a number of reasons. Three refused to sign the consent form and specifically mentioned security concerns about revealing their private banking data. One was not a customer of the bank we studied. Eleven were customers of the bank we studied but did not know about its site-authentication image feature or had never used it. Five were bank customers but could not remember their passwords or challenge questions and therefore could not complete the tasks. One participant completed only the first task, which did not require him to login.³

The three potential participants (3%) who refused to sign the consent form cited privacy concerns and the terms under which they would be observed. It is possible that other excluded participants lied to protect their privacy: we have no way of knowing if participants truly forgot their passwords or if they were simply uncomfortable providing their passwords during the study.

4. Presenting Attack Clues

We now describe three clues that the bank's password-entry page may have been forged by an attacker. These clues are indistinguishable from those that a user might encounter during a real attack. We activated each attack clue during one of three separate login tasks.

4.1. Removing HTTPS indicators

The first attack clue was the absence of browser HTTPS indicators when the bank's password-entry page was displayed. We prevented IE6 from displaying either the `https` in the browser address bar or the lock icon at the bottom right of the browser frame. We wanted to determine what fraction of participants treat these indicators as prerequisites to entering their password.

We instructed our proxy to replace `https` with `http` in all content that linked to the domain name that hosted the password-entry page. We also disabled javascript code in the bank's unauthenticated (HTTP) home page that would otherwise redirect the browser to the authenticated (HTTPS) version of the site.

An attacker able to intercept and modify traffic (a man-in-the-middle attack) could use the same approach to prevent the activation of HTTPS on the bank's password-entry page. The attacker could then continue to impersonate the

³We do not know why he stopped participating: perhaps he did not have an account, could not remember his password, or did not want to provide his password.

bank's servers and forge pages that are supposed to be authenticated via HTTPS. To detect such attacks, users must verify the presence of indicators that show HTTPS is active. In IE6, users can either look for the `https` in the address bar or for a lock icon at the bottom right of the browser frame.

4.2. Removing site-authentication images

Banks and other websites have introduced site-authentication images to help users distinguish a bank's real website from impostor sites. These sites repeatedly instruct their customers to verify their site-authentication images before entering their passwords. The absence of a site-authentication image is a clue that might reveal a man-in-the-middle attack against the bank's web address. Alternatively, this clue might also alert participants that the page has been loaded from the address of a 'phishing' web site. We wanted to measure how many participants would disregard the absence of their site-authentication images.

In this task, we continued to disable HTTPS. We used our HTTP proxy to rewrite the bank's password-entry page—removing the site-authentication image (and any accompanying text phrase) and replacing it with an upgrade message.

In the study, we used the real brand name of the bank's site-authentication image feature in the message. For this paper, we've replaced that brand name with the acronym "SAI".

SAI Maintenance [sic] Notice:
[bank name] is currently upgrading our award winning SAI feature. Please contact customer service if your SAI does not reappear within the next 24 hours.

4.3. Presenting a warning page

The final attack clue was hard not to notice: we replaced the password-entry page with a warning page copied from Internet Explorer 7 Beta 3.⁴ We wanted to see if participants would disregard such a conspicuous warning message.

This warning page, illustrated in Figure 1, strongly discourages participants from continuing to the website. It offers two options: to close the website (recommended) and the option to continue (not recommended). The warning

⁴This page is the same as that in the final release of Internet Explorer 7. We modified the IE7 warning page so that images and scripts would be loaded from a fictional address that could be intercepted and served by our proxy: `http://browser.security/`. Therefore, we were able to simulate a warning page in Internet Explorer 6 that looked identical to that of Internet Explorer 7. There was one difference: when a user clicked to close the page, they would see a dialog from Internet Explorer 6 asking them to confirm that they wanted the window closed.

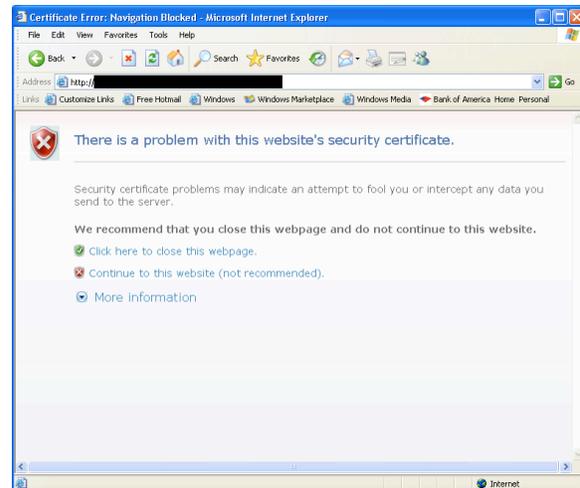


Figure 1. An image of the warning page sent in place of the password-entry page. The black rectangle in the address bar was not present in the study; we have added it here to hide the identity of the bank.

explains that "Security certificate problems may indicate an attempt to fool you or intercept any data you send to the server."

5. Results

5.1. How we report results

In this study, we collected password-entry data from two sources: data observed by our proxy and data self-reported by participants.

Our proxy recorded a binary outcome for each task: either the participant disregarded attack clues and entered their password, or they withheld their password and did not login. Our proxy reported that a participant entered his or her password only if it observed the participant submit a form field named 'passcode' to the password-entry page. We used data self-reported by participants to corroborate our observed response.⁵

Of course, experimental realities prevent us from perfect observation: we identified participants whose self-reported responses cannot be corroborated with our observed responses. A discrepancy may occur because the participant did not enter the bank's correct web address, because the experiment facilitator made an error in engaging the proxy,

⁵In addition to responses reported on the task sheet, we used a post-task questionnaire to ask participants if they entered their passcode for each task (participants were allowed to refer to their task sheets to help them recall their behavior).

or because the participant did not correctly recall his or her behavior. When we cannot determine conclusively that participants were exposed to an attack clue or whether they entered passwords, we exclude those participants from the results for that task. Furthermore, we exclude those participants from the results of subsequent tasks, because they may not have been exposed to the same attack clues as other participants.

Participants who withhold their passwords during one task for security reasons may be more likely to be aware of attack clues during subsequent tasks. For this reason, when participants withhold their passwords in one task, we note this fact when reporting their responses to subsequent tasks. We observed that all participants who withheld their passwords in one task did so on all future tasks.

5.2. Removing HTTPS indicators

During the third task, we removed HTTPS indicators from the password-entry page and then asked participants for information that could only be retrieved by logging in.

We were able to collect and verify responses for 63 out of 67 participants. All 63 participants entered their passwords and completed this task, despite the absence of HTTPS indicators on the password-entry page.

This included 18 participants from the role playing group (Group 1), 18 participants from the security primed group (Group 2), and 27 participants from the personal account group (Group 3). No participant mentioned the absence of HTTPS indicators when asked if they had any difficulties performing this task.

We do not report responses for 4 out of 67 participants because we were not able to record or verify their response to the attack clues. Two participants failed to follow instructions and never saw the password-entry page. One participant, who was using her own account, appears to have reset her online ID and passcode during the task—this process required her to enter her username and passcode or to answer a series of challenge questions. Finally, there was one participant who reported entering a password, but our proxy did not record that a password was entered. We do not include the responses of these participants in future tasks.

5.3. Removing site-authentication images

Of the 63 participants whose responses to prior tasks had been verified, we were able to corroborate 60 participants' responses to the removal of their site-authentication images. 58 of the 60 participants (97%) entered their passwords, despite the removal of the site-authentication image. Only 2 participants (3%) chose not to login, citing security concerns.

Table 3 shows how participants in each group responded to the removal of the site-authentication image from the password-entry page. All 18 participants in the role playing group (Group 1) entered their passwords, despite the absence of their site-authentication images from the password-entry page. All 17 participants in the security primed group (Group 2) also entered their passwords. Even 23 of 25 participants (92%) in the personal account group (Group 3), who were using their own accounts and whose own site-authentication images had been removed, chose to enter their passwords.

The two participants who withheld their passwords both cited the absence of their site-authentication images as the reason for their decision.

We could not corroborate three participants' responses: we inspected our proxy's logs and found that none of these had even reached the password-entry page from which we would have removed their site-authentication images. Because we know that these participants were not exposed to the attack clue, we do not report their data in Table 3. They are also excluded from our analysis of future tasks because they experienced one fewer attack clue than other participants.

5.4. Presenting warning pages

Of the 60 participants whose responses to prior tasks had been verified, we were able to corroborate 57 participants' responses to the warning page. Despite the overttness of the warning page and its strong wording, 30 of 57 participants (53%) entered their passwords. 27 participants (47%) did not login.

Table 4 shows how participants in each group responded to the warning page. 10 of 18 participants (56%) in the role playing group (Group 1), 12 of 17 participants (71%) in the security primed group (Group 2) and 8 of 22 participants (36%) in the personal account group (Group 3) entered their passwords despite the warning page.

Two of the 27 participants who did not enter their password during this task had also not done so during the previous task (during which their site-authentication image was removed). Both of these participants were from the personal account group (Group 3).

We could not corroborate three participants' responses: all were members of the personal account group (Group 3) and all reported entering their passwords. Our proxy reported that they had not. It is possible that these participants' response to the warning page was to withhold their passwords. By excluding these subjects, it is possible that our results under-represent the level of security vigilance exhibited by participants assigned to the personal account group (Group 3).

	Group 1		Group 2		Group 3		Groups 1 ∪ 2		Total	
	Role playing		Sec. primed		Pers. acct.					
Sent password	18	100%	17	100%	23	92%	35	100%	58	97%
Didn't login	0	0%	0	0%	2	8%	0	0%	2	3%
Total	18		17		25		35		60	

Table 3. Participant responses to the removal of site-authentication images.

	Group 1		Group 2		Group 3		Groups 1 ∪ 2		Total	
	Role playing		Sec. primed		Pers. acct.					
Sent password	10	56%	12	71%	8	36%	22	63%	30	53%
Didn't login	8	44%	5	29%	14 [†]	64%	13	37%	27	47%
Total	18		17		22		35		57	

[†]We include the two members of Group 3 who also withheld their passwords on the previous task. If we exclude those two participants, 12 members (60%) of Group 3 withheld their passwords on this task.

Table 4. Participant responses to the warning page.

5.5. Comparing participant scores between groups

One can not accurately compare groups using responses to the second or third attack clues alone: responses to these clues may have been influenced by the presence of prior clues. Instead, we assigned a score to each participant that reflects the *insecurity* of that participant's behavior. The score denotes the number of times that a participant behaves insecurely (entering his or her password after being presented with an attack clue) before first behaving securely (withholding his or her password). The highest score, 3, was reserved for participants who entered their passwords on all attacks.

Thus, a score of 3 is more insecure than 2, which in turn is more insecure than 1. The score is ordinal, so the difference in the level of security vigilance between scores of 1 and 2 is not necessarily equal to the difference between 2 and 3.

Out of 67 participants who entered our study, we were able to establish the response of 57 to all attack clues. We found that no participants withheld their passwords in response to the first attack clue, 2 participants (4%) first withheld their passwords in response to the second attack clue, and 25 participants (44%) first withheld their passwords in response to the third attack clue. 30 participants (53%) disregarded all three of the attack clues and entered their passwords on all tasks. Table 5 reports the total number of participants in each group who received each score. Figure 2 shows the percentage of each group that received each score.

We used the Mann Whitney U test to test the hypothesis that group assignment increases or decreases a partic-

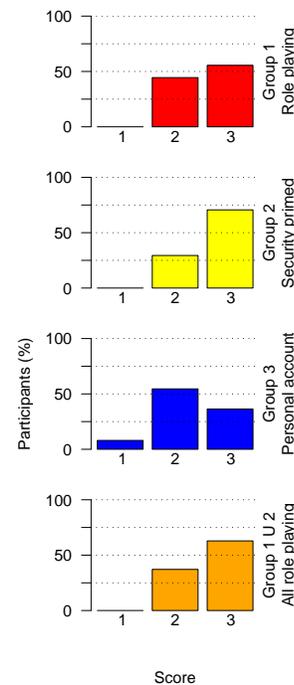


Figure 2. The distribution of scores in each group. Scores distributed further to the right indicate less secure behavior. No participant received a perfect (0) score.

Score	First chose not to enter password...	Group				Total
		1	2	3	1 ∪ 2	
0	upon noticing HTTPS absent	0 0%	0 0%	0 0%	0 0%	0 0%
1	after site-authentication image removed	0 0%	0 0%	2 9%	0 0%	2 4%
2	after warning page	8 47%	5 29%	12 55%	13 37%	25 44%
3	never (always logged in)	10 53%	12 71%	8 36%	22 63%	30 53%
<i>Total</i>		18	17	22	35	57

Table 5. The number of participants in each group (column) who received each score (row). (Some totals are more than 100% due to rounding.)

ipant’s score.⁶ This test allows us to compare the median scores between two groups and suggests whether those two groups come from the same population or not. Our null hypothesis is that the two populations represented by the two groups have the same distribution of scores.

We found that participants assigned to the security primed group (Group 2) did not behave more securely than those in the role playing group (Group 1). In fact, a greater fraction of the participants assigned to the role playing group (Group 1) exhibited more secure behavior (had lower scores) than those assigned to the security primed group (Group 2). The difference between the two groups was not statistically significant (U=130, exact two-tailed P=0.489).

When comparing the role playing group (Group 1) to the personal account group (Group 3), we did not find a significant difference in scores. While a greater fraction of the participants assigned to the personal account group (Group 3) chose the secure behavior (had a lower score) than those in the role playing group (Group 1), the difference between the two groups was not statistically significant (U=154, exact two-tailed P=0.184).

We found that participants assigned to the personal account group (Group 3) were significantly more likely to behave securely than those in security primed group (Group 2). Roughly two thirds of the participants in the security primed group (Group 2) entered their passwords on all tasks, compared to slightly more than a third of participants in the personal account group (Group 3). The difference between the two groups was statistically significant (U=118, exact two-tailed P=0.038).

Finally, we compared the personal account group (Group 3) to all participants who played a role, regardless of the level of security priming (the union of the two role playing groups). Again, we observed that those who used their own accounts in the study had significantly lower scores and thus behaved more securely than those playing roles (U=270, exact two-tailed P=0.037).

⁶The Mann Whitney U test, equivalent to the Wilcoxon Rank-Sum test, is a non-parametric statistical significance test for comparing the medians of two independent samples of ordinal ranks. It is used as an alternative to the t-test when the data is not normally distributed.

6. Discussion

6.1. The efficacy of HTTPS indicators

Prior studies have reported that few users notice the *presence* of HTTPS indicators such as the browser lock icon [6, 20]. Our results corroborate these findings and extend them by showing that even participants whose passwords are at risk fail to react as recommended when HTTPS indicators are *absent*.

The failure of all of our participants to respond to the removal of HTTPS indicators cannot be entirely attributed to ignorance. In the post-task questionnaire, three participants mentioned HTTPS indicators, though only in the context of explaining another attack clue.

We should caution that these results cannot be automatically applied to websites that don’t employ site-authentication images: customers may be less likely to pay attention to HTTPS indicators when instructed to focus on their site-authentication images.

6.2. The efficacy of site-authentication images

Even though the bank repeatedly instructed customers not to login if their site-authentication images are absent, the vast majority of participants *using their own bank accounts* did not comply—23 of 25 (92%) entered their own account passwords even though their site-authentication images were absent.

One explanation for low compliance is that more security-conscious customers did not participate in our study. A small fraction (3 potential participants) excluded themselves from the study after reading the consent form. Even if we assumed that they would have withheld their passwords, they would still be among a small minority who behaved securely. Therefore, we believe that attacks that remove site-authentication images will fool most online banking customers.

It is also important to note that the presence of a site-authentication image does not guarantee that a connec-

tion is secure or that it is safe to enter a password: site-authentication images have been shown to be vulnerable to man-in-the-middle attacks that capture and display the user's site-authentication image [25]. To prevent man-in-the-middle attacks, users must still verify the site's address and the activation of HTTPS. Furthermore, the use of site-authentication images will not protect the passwords of customers using computers infected with spyware.

Despite these facts, sites that deploy site-authentication images often instruct their customers that they need only verify their site-authentication image to ensure the security of their password. This reinforces the message that site-authentication images are not only necessary, but sufficient to ensure that it is safe to enter a password. For example, consider the following instructions from ING Direct, Bank of America, Vanguard, and Yahoo!:

“...your image and phrase will be displayed so you'll know immediately that it's safe to enter your Login PIN”—ING Direct [10]

“If you recognize your SiteKey, you'll know for sure that you are at the valid Bank of America site. Confirming your SiteKey is also how you'll know that it's safe to enter your Passcode and click the Sign In button.”—Bank of America [2]

“When you see your image, you can be confident that you're on Vanguard.com® and not an impostor site and can safely enter your password.”—Vanguard [18]

When you create a sign-in seal for your computer, you can be sure you're on a legitimate Yahoo! site each time you use this computer to sign in to Yahoo!—Yahoo! [23]

In fact, at least one participant attributed his decision to disregard the final attack clue to the presence of a site-authentication image. “Eventually, I ignored the IE warning and, on seeing the proper [site-authentication image], entered the...password.”

6.3. The efficacy of warning pages

Eight of 22 participants (36%) *who were using their own account* chose to login after seeing the warning page.

We did not attempt to emulate all of the interface features that Internet Explorer 7 (IE7) presents when it encounters a page with a certificate error. For example, if a user disregards a warning page, IE7 presents insecurely loaded pages with a red address bar and a drop-down warning message. Because we do not emulate this behavior, we cannot measure the efficacy of the IE7 browser's security warnings.

Our choice to emulate only one component of IE7's security warnings may have increased the likelihood that participants would enter their passwords. However, other factors may have decreased the likelihood of password entry.

First, because we maintained the same task and attack order throughout the study, participants had already been exposed to two attack clues. Although the great majority of participants had entered their passwords despite the previous clues, the combination may have made them more suspicious that they were under attack or that we were studying their security behavior.

Next, we conducted the study before the final release of IE7, so very few participants had seen warning pages before. Now that IE7 is widely available, users may see warning pages often enough to become complacent about heeding them.

Finally, one artifact of our implementation was that when a participant clicked through the warning, the form data posted to the password-entry page was lost. Participants had to re-enter their usernames in order to proceed.

The compliance rates for our warning pages may vary from the rates that would be obtained if we had used IE7 in our study. However, it is important to note that this does not affect our hypothesis that participants who use their own accounts respond differently to warnings than those who play a role.

6.4. The effect of role playing

Participants who used their own accounts in our study behaved more securely than those who were assigned to play roles. While we did not see a statistical difference when we compared the personal account group (Group 3) to the role playing group (Group 1), we did find a significant difference when comparing the personal account group (Group 3) to the security primed group (Group 2) and to the union of all role-playing groups.

Our results should give pause to researchers designing studies that rely on role playing. Participants who may be vigilant in securing themselves from real-life risks may be less motivated to behave securely when playing a role—especially if the risks are perceived as fictional.

It is possible that better study designs, with more compelling scenarios, could increase the security-vigilance of role playing participants to the levels exhibited by those exposed to more realistic risk. However, even if a scenario successfully approximates real-world conditions in one study, it may not be equally effective when experimental conditions change (*e.g.*, when a different system is being tested, a different population is used, or when the context of use differs). As with any experimental condition, it is not possible to isolate the effects of role playing on a given study without replicating the study without role playing.

Our results do not discount the usefulness of role-playing scenarios. In some cases, artificial scenarios may be the only way to simulate attack responses in an ethical manner. For example, role playing may be a useful device in qualitative studies where researchers want to closely observe participants without compromising their privacy. Role playing may also be useful in studies of the *comparative* efficacy of security mechanisms, in which each group uses an identical role-playing scenario but a different mechanism.

We note that role playing does not always protect participants from risk. Although participants in the role-playing groups were asked to use the credentials of the role assigned to them, a few participants occasionally disregarded or forgot the instructions and logged in to their own bank accounts. If participants asked us whether they should be role-playing, we pointed them to the role-playing instructions and role credential sheets; we allowed them to continue and treated them as if they had been role-playing the whole time. Role-playing failures have also been observed in a previous study in which participants logged into ‘phishing’ websites using their own account credentials, even though they knew the websites might be illegitimate and they had been asked to play a role [6].

6.5. The effect of security priming

Though the result was not statistically significant, we were surprised to find that participants assigned to the security primed group behaved less securely than those in the role playing group, who had no security-priming. Because the difference is not significant, it may be due to chance.

One alternative explanation is that the security-priming instructions were too subtle: we wanted to test if simply mentioning security would affect behavior. Participants may have behaved more securely if we had been more specific about how they should protect their password. It is also possible that the role-playing effect was actually stronger in the security primed group than the role playing group (Group 1): we informed participants that security was important in the *context of their role*, which significantly increased the length of the instructions devoted to role playing.

While our methodology and sample sizes did not produce a measurable effect of security-priming on security behavior, such an effect may still exist. Measuring the conditions under which priming affects security behavior is an area for future work. For example, future studies might focus specifically on the effects of security training or the effects of providing monetary incentives to behave securely.

6.6. Limitations of our study

While we took great efforts to maintain our study’s validity, there are limits to what can be achieved in a laboratory study.

Some design aspects of our study may have caused participants to behave less securely than they would in the real world. Because the experiments were conducted in a university setting, participants may have felt more safe than they otherwise would feel (*e.g.*, than if they were at an Internet cafe). The consent form informed participants that we would not record sensitive personal information, which may have also increased their perception of safety. Finally, we provided participants with a financial incentive of \$25. Although participants were informed that they could stop at any time and still receive the participation fee, they may have felt obligated to complete the tasks in order to claim the fee.

Other aspects of our design may have caused participants to behave more securely than they would in the real world. Even participants who received no security instructions may have been able to infer that security was the focus of the study, or they may have behaved more cautiously because they knew they were being observed. Furthermore, users conducting their own real-world banking tasks may be more motivated than those in our study to login in order to complete their task (*e.g.*, they may need to pay a bill before it is due).

7. Related Work

To avoid capturing sensitive personal information, the majority of security usability studies ask users to engage in role-playing scenarios.

In many studies, the researchers do not tell participants to behave securely or that they will be attacked. Instead they attempt to give participants a secret to protect that is “comparable to the value that users place on their own secrets in the real world” [21]. For example, in one of the earliest security usability studies, Whitten and Tygar asked participants to use email encryption software [21]. Participants were asked to assume the role of a political campaign coordinator who was tasked with sending sensitive email messages to other campaign volunteers. The study concluded that a majority of participants could not successfully sign and encrypt a message using PGP 5.0. Garfinkel and Miller used the same scenario to test a Key Continuity Management (KCM) email encryption interface [8]. However, they also simulated an escalating series of attacks (spoofed messages that appeared to come from other campaign members) and found that the interface was successful in preventing some attacks and not others.

Many studies instruct participants to behave securely. For example, Whalen and Inkpen [20] conducted a study in which they asked participants to perform common online browsing tasks, some of which required participants to login to an account and make purchases. Participants used login and credit card information created for the study; they were asked “to treat the data as their own” and to keep it confidential. An eyetracker was used to reveal whether security indicators were checked by participants. In the second half of the study, participants were specifically instructed to behave securely. Whalen and Inkpen found that unless instructed to behave securely, many participants did not check whether a page was secure because “it was not their own data and thus they took no care to protect the information”.

Wu *et al.* conducted a usability study to analyze how participants use anti-phishing toolbars to detect fake phishing websites [22]. Rather than asking participants to login using their own accounts, they created dummy accounts in the name of “John Smith” at various e-commerce websites. The participants were asked to play the role of John Smith’s personal assistant, to process email messages on his behalf, and to protect his passwords. Wu’s study design also featured a tutorial, where users were trained on how to use the anti-phishing toolbar. The results found that participants were fooled 34% of the time. Even when asked to focus on the toolbars, many participants ignored them when webpages looked convincing enough.

Other studies notify participants that attacks are part of the study. For example, Dhamija *et al.* conducted a usability study in which participants were asked to distinguish legitimate websites from spoofed phishing websites [6]. Participants were asked to assume the role of someone who had clicked on a link in email and arrived at the website in question. Despite the heightened security awareness, the study found that some phishing websites were able to fool a large fraction of participants.

Other studies have observed users as they provide their own credentials to real systems. Jagatic *et al.* conducted a study in which a social network was used for extracting information about social relationships [11]. The researchers used this information to send phishing email to students on a university campus that appeared to come from a close friend. 72% of users responded to phishing email that was from a friend’s address, while only 16% of users responded in the control group to phishing email from an unknown address. To preserve ecological validity, the researchers did not ask the students for their prior consent to participate in the study. Many who received the phishing email reported feeling angered and violated [5]. In contrast, we observed only participants who had consented to observation.

Jakobsson and Ratkiewicz used the features of an online auction website to send simulated phishing emails to that site’s members [12]. The phishing email only *appeared* to

be a phishing attempt; to respond to the message the recipient had to provide their credentials to the real auction site. Researchers could learn that users logged into the auction site if they received a response to their message, without having to collect user credentials. Using a variety of techniques, their experiments revealed that, on average, 11% of users logged into the auction site to respond to the illegitimate messages. In this study, participants logged into a legitimate website: their credentials were not at risk. A participant’s decision to respond to the forged email provides insight into how he or she authenticates messages; however, the researchers do not specifically study whether participants can authenticate websites or protect their credentials.

8. Conclusion

We contribute a study design for observing participants as they log into security-critical web sites with their own authentication credentials. Using this design, we measured the efficacy of security indicators (HTTPS indicators and site-authentication images) and the effect of role-playing on the study. We find that:

Users will enter their passwords even when HTTPS indicators are absent.

All participants entered their passwords after HTTPS indicators were removed, including all 27 who were using their own account credentials.

Users will enter their passwords even if their site-authentication images are absent.

23 of 25 (92%) participants using their own account, and all other participants, entered their passwords when the site-authentication image was replaced by an upgrade message. Thus, it is not clear that the deployment of site-authentication images increases customers’ ability to detect fraudulent websites.

Site-authentication images may cause users to disregard other important security indicators.

Many sites that have deployed site-authentication images instruct customers that the presence of their these images is a sufficient condition for security, when it is only one of many necessary conditions.

Role playing has a significant negative effect on the security vigilance of study participants.

Participants who played roles disregarded more attack clues before withholding their passwords than participants whose own passwords were at risk.

Acknowledgments

Harvard's Center for Research on Computation and Society (CRCS) provided the catalyst for this collaboration. Scott Briere and John O'Conner of MIT Lincoln Laboratory obtained and assembled the bulk of the hardware used in the study. Pat Tomaino (Harvard) assisted with subject recruitment. David Purdy (UC Berkeley) helped us choose statistical tests and provided scripts to reproduce scores. Ross Anderson (U Cambridge), Rob Cunningham (MIT Lincoln Laboratory), Ben Edelman (Harvard), Allan Friedman (Harvard), Jens Grossklags (UC Berkeley), Marti Hearst (UC Berkeley), Glenn Holloway (Harvard), Greg Morrisett (Harvard), Galen Pickard (MIT Lincoln Laboratory), Allan Schiffman (CommerceNet), Stuart Shieber (Harvard), Michael D. Smith (Harvard), Doug Tygar (UC Berkeley), and the anonymous reviewers added insight and provided much needed feedback on earlier drafts.

References

- [1] F. Arshad and R. Reeder. Symposium On Usable Privacy and Security Conference Report—When User Studies Attack: Evaluating Security by Intentionally Attacking Users. http://cups.cs.cmu.edu/soups/2005/SOUPS_2005_Conference_Report.html, 2005.
- [2] Bank of America. SiteKey Frequently Asked Questions. http://www.bankofamerica.com/onlinebanking/index.cfm?template=site_key. Downloaded and archived on November 1, 2006.
- [3] Bank of America. SiteKey: Online Banking Security. <http://www.bankofamerica.com/privacy/sitekey/>. Downloaded and archived on November 1, 2006.
- [4] M. B. Brewer. Research Design and Issues of Validity. In H. T. Reis and C. M. Judd, editors, *Handbook of Research Methods in Social and Personality Psychology*, pages 3–16. Cambridge University Press, Mar. 28 2000.
- [5] C. Corley. 'Phishing' Experiment Attracts National Debate about Ethics of Study. <http://www.idsnews.com/news/story.php?id=29791>, Apr. 28, 2005.
- [6] R. Dhamija, J. Tygar, and M. Hearst. Why Phishing Works. In *Human Factors in Computing Systems (CHI 2006)*, Quebec, Canada, Apr. 22–27, 2006.
- [7] R. Franco. Better Website Identification and Extended Validation Certificates in IE7 and Other Browsers. <http://blogs.msdn.com/ie/archive/2005/11/21/495507.aspx>, Nov. 21, 2005.
- [8] S. L. Garfinkel and R. C. Miller. Johnny 2: A User Test of Key Continuity Management with S/MIME and Outlook Express. In *Proceedings of the Symposium On Usable Privacy and Security (SOUPS)*, Pittsburgh, PA, USA, July 6–8 2005.
- [9] S. Hartman. IETF Internet-Draft: Requirements for Web Authentication Resistant to Phishing. <http://www.ietf.org/internet-drafts/draft-hartman-webauth-phishing-02.txt>, Oct. 21, 2006.
- [10] ING DIRECT USA. Privacy FAQs. https://home.ingdirect.com/privacy/privacy_security.asp?s=newsecurityfeature. Downloaded and archived on November 1, 2006.
- [11] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer. Social Phishing. *Communications of the ACM*, To Appear.
- [12] Markus Jakobsson and Jacob Ratkiewicz. Designing ethical phishing experiments: a study of (ROT13) rOnl query features. In *Proceedings of the 15th international conference on World Wide Web (WWW '06)*, pages 513–522, New York, NY, USA, 2006. ACM Press.
- [13] Proposed User Studies, Security User Studies Workshop, Symposium On Usable Privacy and Security. http://cups.cs.cmu.edu/soups/2006/user_study_proposals.pdf, July 2006.
- [14] G. Staikos. Web Browser Developers Work Together on Security. <http://dot.kde.org/1132619164/>, Nov. 2005.
- [15] United States Federal Financial Institutions Examination Council (FFIEC). Authentication in an Internet Banking Environment. http://www.ffiec.gov/pdf/authentication_guidance.pdf, Oct. 2005.
- [16] Harvard University. Efficacy of Web Browser Warnings and Notices, Protocol Number F13264-102, Sept. 2006.
- [17] Massachusetts Institute of Technology. Efficacy of Web Browser Warnings and Notices, Protocol Number 0512001551, Sept. 2006.
- [18] T. Vanguard Group, Inc. Enhanced Logon FAQs. <https://flagship.vanguard.com/VGApp/hnw/content/UtilityBar/SiteHelp/SiteHelp/SecurityLogonFAQsContent.jsp>. Downloaded and archived on November 1, 2006.
- [19] W3C. Web Security Context - Working Group Charter. <http://www.w3.org/2005/Security/wsc-charter>, Sept. 7, 2006.
- [20] T. Whalen and K. Inkpen. Gathering Evidence: Use of Visual Security Cues in Web Browsing. In *Graphics Interface 2005*, 2005.
- [21] A. Whitten and J. D. Tygar. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. In *Proceedings of the 8th USENIX Security Symposium*, pages 169–184, Washington, DC, Aug. 23–26, 1999.
- [22] M. Wu, R. C. Miller, and S. Garfinkel. Do Security Toolbars Actually Prevent Phishing Attacks? In *Human Factors in Computing Systems (CHI 2006)*, Quebec, Canada, Apr. 22–27, 2006.
- [23] Yahoo! Inc. What is a sign-in seal? - Yahoo! Account Security. <http://help.yahoo.com/l/us/yahoo/security/phishing/phishing-110140.html>, 2006. Downloaded and archived on November 1, 2006.
- [24] Yahoo! Inc. Yahoo! Personalized Sign-In Seal. <https://protect.login.yahoo.com/>, 2006.
- [25] J. Youll. Fraud Vulnerabilities in SiteKey Security at Bank of America. Technical report, Challenge/Response, LLC, July 18, 2006. <http://cr-labs.com/publications/SiteKey-20060718.pdf>.

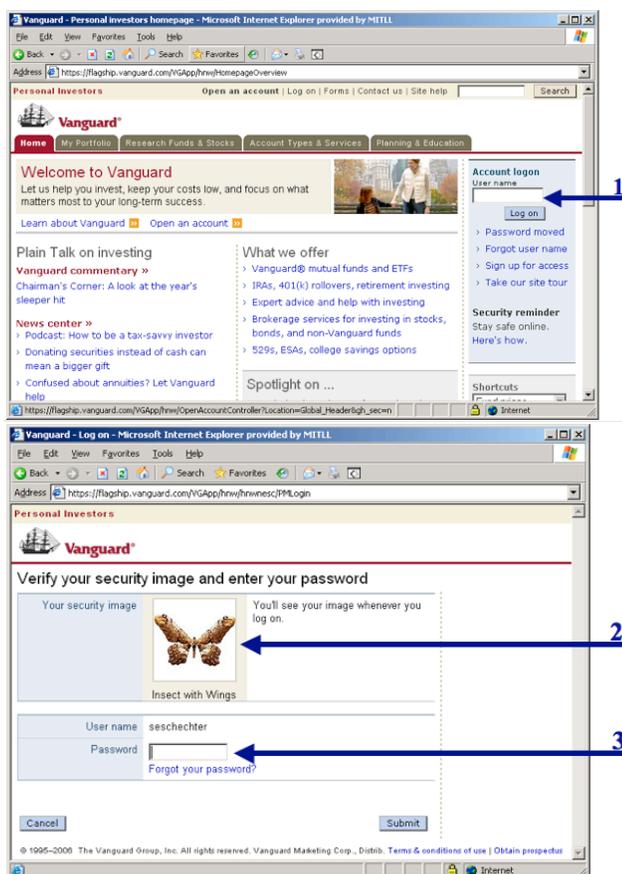
A. How site-authentication images work: example login processes

A.1. Vanguard

Vanguard is one example of a financial site that implements site-authentication images. The login process is divided into two pages: one for entering the username and one for entering the password. The login steps are:

1. Enter username and press the “Log on” button.
2. Verify the security image (site-authentication image).
3. Enter the password and press the “Submit” button.

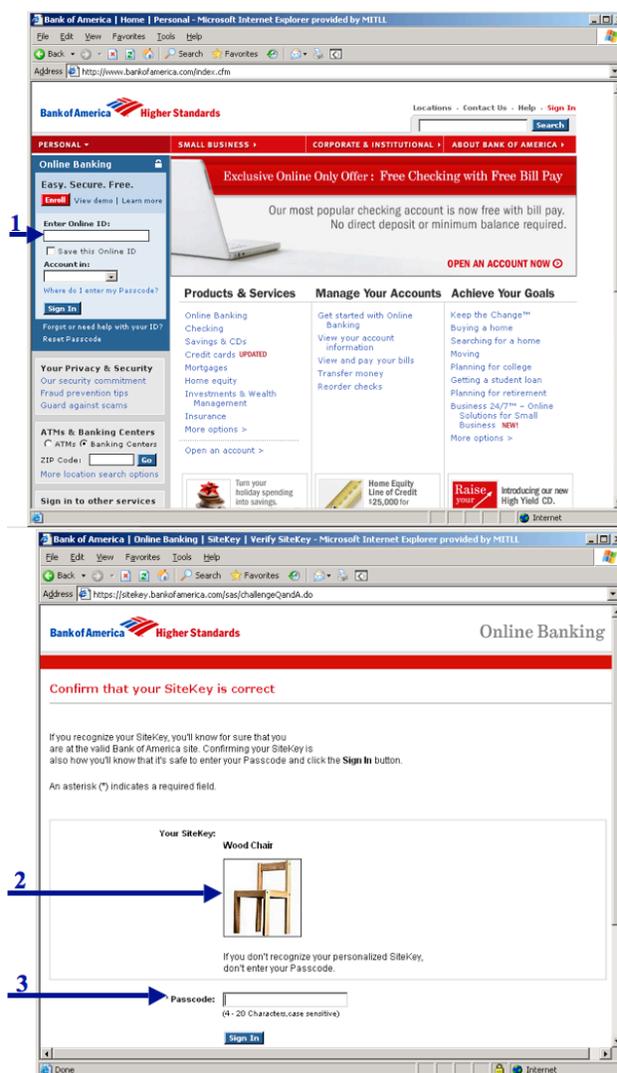
After the customer submits their username, the website checks to see if the user has previously logged in from the same client. It does this by looking for a HTTP cookie sent by the user’s browser or by using a Macromedia Flash object. If the website cannot confirm that the user has previously logged in using this client, it will ask the user to answer a challenge question. If the user answers the challenge question correctly, the website will then present the password-entry form.



A.2. Bank of America

Bank of America’s process is similar to that of Vanguard. One exception is that customers may need to identify their state along with their online ID. The login steps are the same as those for Vanguard:

1. Enter the Online ID (username) and press the “Sign In” button.
2. Verify the SiteKey (site-authentication image).
3. Enter the passcode (password) and press the “Sign In” button.



A.3. Yahoo

Yahoo associates site-authentication images with computers, rather than individual user accounts, by placing cookies and/or Macromedia Flash objects on the computer. All users of a computer share the same site-authentication image. Yahoo can thus show the site-authentication image, username entry box, and password entry box on a single login form. The login steps are re-arranged so that they begin with the verification of the site-authentication image.

1. Verify the sign-in seal.
2. Enter the “Yahoo! ID” (username).
3. Enter the password and press the “Sign In” button.

