Your Reputation Precedes You: History, Reputation, and the Chrome Malware Warning

Hazim Almuhimedi
Carnegie Mellon University
hazim@cs.cmu.edu

Adrienne Porter Felt
Robert W. Reeder
Sunny Consolvo
Google, Inc.
felt, rreeder, sconsolvo@google.com

ABSTRACT

Several web browsers, including Google Chrome and Mozilla Firefox, use malware warnings to stop people from visiting infectious websites. However, users can choose to click through (i.e., ignore) these malware warnings. In Google Chrome, users click through a fifth of malware warnings on average. We investigate factors that may contribute to why people ignore such warnings. First, we examine field data to see how browsing history affects click-through rates. We find that users consistently heed warnings about websites that they have not visited before. However, users respond unpredictably to warnings about websites that they have previously visited. On some days, users ignore more than half of warnings about websites they've visited in the past. Next, we present results of an online, survey-based experiment that we ran to gain more insight into the effects of reputation on warning adherence. Participants said that they trusted high-reputation websites more than the warnings; however, their responses suggest that a notable minority of people could be swayed by providing more information. We provide recommendations for warning designers and pose open questions about the design of malware warnings.

1. INTRODUCTION

Modern browsers such as Google Chrome and Mozilla Firefox try to stop users from visiting websites that contain malware. Simply visiting an infectious website can be enough to harm a user's computer, via a drive-by download attack. Instead of loading infectious websites, browsers present users with full-page warnings that explain the threat (Figure 1). Because the malware warning's false positive rate is very low [30], our goal is for no one to ignore the warning. Yet, people click through (i.e., ignore) 7% and 23% of Firefox and Chrome malware warnings respectively [5].

As part of an effort to improve the design of Chrome's malware warning, we investigate factors that may contribute to why people ignore such warnings. One hypothesis is that some users trust familiar websites enough to not believe warnings about the familiar websites, leading them to click through the warnings. In this paper, we test this familiarity hypothesis through (a) an analysis of nearly four million actual Google Chrome warning impressions, and (b) a survey-based controlled experiment conducted with 1,397

Symposium on Usable Privacy and Security (SOUPS) 2014, July 9–11, 2014, Menlo Park, CA.

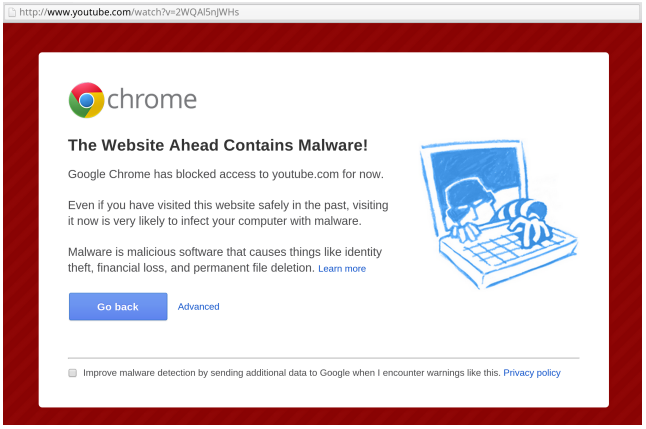


Figure 1: Malware warning in Google Chrome 32

Amazon Mechanical Turk workers. We investigate the impact of people's familiarity with the website they are attempting to visit, as well as how they found out about the website. We also tested minor variations of the instrument used in our survey-based experiment to determine how small wording changes affected responses (e.g., whether or not participants were primed with the word "warning").

Our field data and Mechanical Turk experiment both support our familiarity hypothesis. In our analysis of 3,875,758 malware warning impressions, users were twice as likely to click through the Chrome malware warning if the blocked website was in their browsing history. To further explore why users seem to ignore such warnings, we asked participants in our survey-based experiment about hypothetical warning scenarios. Participants said that it was unlikely that a well-known website would contain malware, so the warning was probably a mistake. They did not appear to realize that even reputable websites can be compromised to temporarily distribute or redirect to malware. However, when participants weren't familiar with the website, they said they would be more likely to play it safe and trust the browser's recommendation.

Contributions. We make the following contributions:

- Through field data and results of an online survey-based experiment, we demonstrate that a person's familiarity with a blocked website has a strong influence on her response to malware warnings.
- We are the first to investigate *why* participants might heed or ignore browser malware warnings.

- Based on the misconceptions and pain points revealed by participants in our survey-based experiment, we provide recommendations for the design of browser malware warnings.
- Through minor variations of our survey instrument, we explore how role-playing, priming, and interactivity affect results of online survey-based warning studies.

1.1 Why Show Malware Warnings?

Ignoring a malware warning carries substantial risk because the false positive rate is very low [30]. This naturally raises a question: why does Chrome let users click through the warning? We could achieve a 0% click-through rate for the warning simply by taking away the ability to proceed.

We don't fully block malicious websites because of the following concerns:

- A determined user might disable the Safe Browsing service to get to the desired content. This would leave the user without protection in the future.
- An unconvinced user could simply open the website in another browser that does not block the website. The user would likely be exposed to the same risk in another browser, but possibly without realizing it.

Thus, our goal is to convince users to heed the warning.

2. BACKGROUND

We explain when and why Google Chrome shows malware warnings. We then cover prior literature on browser warnings, which has primarily focused on SSL warnings.

2.1 Malware Warnings

Google Safe Browsing [3] scans websites for signs of malware or phishing. The service maintains a list of known malware and phishing sites. Google Chrome checks every page load against this list, looking for two things:

1. Is the destination URL on the list?
2. Does the page load resources (e.g., scripts) from third parties that are on the list?

For both conditions, Google Chrome halts the page load and shows a malware or a phishing warning. Users can click on "Advanced" (Figure 1) and then "Proceed at your own risk" (Figure 5) to dismiss the warning and load the page.

The Safe Browsing list includes many websites that primarily function as attack sites. However, legitimate websites can also temporarily end up on the list if they are compromised [30]. Attackers can subvert legitimate websites via vulnerabilities, user-contributed content, advertisements, or third-party widgets [31]. Websites are removed from the list when they no longer pose a risk.

2.2 Related Work

Malware warnings. Microsoft reported that the CTR for Internet Explorer's SmartScreen malware warning was under 5% in 2011 [21]. Akhawe and Felt reported telemetry data from Google Chrome and Mozilla Firefox for malware, phishing, and SSL warnings [5]. Based on their analysis, malware warning CTRs fluctuate in Google Chrome but not in Mozilla Firefox. They did not investigate the degree of fluctuation or its causes. In this paper, we delve further into the fluctuation issue with additional field data and an online, survey-based experiment.

Others have studied users' perceptions of malware in general, without focusing on warnings. Solic and Ilakovac asked electrical engineers and medical personnel about their security habits; all but one participant were concerned enough about malware to use security software [34]. Asgharpour et al. ran a card-sorting exercise to see whether expert and non-expert computer users had similar mental models of malware-related terms [6]. They found that physical world (e.g., locks) and criminal mental models were the best security metaphors for communicating risk to non-experts.

Phishing warnings. Egelman et al. studied phishing warnings and published several recommendations for warning design, including using interruptive (active) warnings and preventing habituation [10]. Egelman and Schechter [11] found that warnings that explain specific threats may reduce clickthroughs compared with warnings that have vague messaging such as "this website has been reported to be unsafe."

SSL warnings. SSL warnings serve a similar purpose: the browser stops a page load, warns the user of risk, and asks the user to make a decision. However, the threat model differs. With an SSL warning, the attacker is on the network; with a malware warning, the attacker is on the destination website. Furthermore, SSL warnings are commonly false positives whereas malware warnings are rarely unwarranted. Thus, it is not clear whether all of the lessons learned from SSL warnings also apply to malware warnings.

Dhamija et al. exposed laboratory study participants to Mozilla Firefox's SSL warnings during simulated phishing attacks [9]. Of their twenty-two participants, only one was able to correctly describe the contents of the warning to researchers. This study demonstrated that people may not pay attention to or understand SSL warnings.

Schechter et al. studied Internet Explorer 7's SSL warning [32]. In their experiment, participants saw SSL warnings while trying to perform tasks on a banking website. The researchers created three versions of the task in which participants used their own credentials, played a role with fake credentials, or played a role with fake credentials and priming. They found a statistically significant difference between the role-playing participants and the non-role-playing participants, but priming had little effect. We follow their lead and similarly test multiple variants of the instrument used in our online survey-based experiment.

Sunshine et al. tested several SSL warnings in an online survey and laboratory study [37]. In their experiment, participants saw warnings on either a banking website or a university library website. Their participants clicked through the SSL warnings at a slightly higher rate for the university library website than for the banking website. We similarly explore the relationship between the website blocked by a warning and participants' willingness to ignore the warning. However, trust plays different roles in SSL and malware warnings. With an SSL warning, the user must evaluate (1) how much she trusts the network connection, and (2) how sensitive the information on the destination website is. With a malware warning, the user must evaluate whether she thinks a website is going to infect her computer.

Sotirakopoulos et al. replicated Sunshine's prior labora-

tory study [35, 37]. Their primary finding was that the laboratory environment had influenced some participants' decisions. For this reason, we do not believe that participants' CTRs in our online survey-based experiment are indicative of their real world CTRs. When intepreting our survey-based experiment's results, we instead focus on differences between scenarios and understanding users' mental models.

Akhawe and Felt showed that Mozilla Firefox's SSL warning has a lower CTR than Google Chrome's SSL warning [5]. In follow-up work, Felt et al. ran a field study to test factors that could explain the difference between the two browsers' warnings [13]. When they ran Firefox's SSL warning in Chrome, it yielded a lower CTR than the default Chrome SSL warning. They found that the imagery, number of clicks, and styling were not responsible for the difference. However, the Firefox-UI-in-Chrome CTR was still higher than the Firefox-UI-in-Firefox CTR. They concluded that demographic factors or other unknown variables besides the warning UI must be influencing user behavior across browsers.

Credibility and trust online. As warning designers, we need users to trust our malware warning more than the infectious target website. To understand users' behavior and trust decisions, we turn to credibility and trust literature.

Fogg et al. identified seven factors that increase or decrease the credibility of websites [15]. Of those factors, five boost website credibility: real-world feel, ease of use, expertise, trustworthiness, and tailoring. Two hurt the credibility of websites: commercial implication and amateurism. In a follow-up study, Fogg et al. asked participants to comment on different aspects of credibility. The most frequently mentioned factor was the "look and feel" of websites. The second most mentioned factor was how well the website was structured. The authors proposed that "Prominence-Interpretation Theory" explains how users evaluate the credibility of a website. First, a user needs to notice an element of the website that increases or decreases its cerdibility. Second, the user needs to decide whether the element increases or decreases the website's credibility [17].

Briggs et al. introduced a "two-process" model of trust: a first impression, followed by careful analysis [7]. They conducted two studies to explore these processes. In the first study, they recruited fifteen participants to participate in sessions about house-purchasing advice. A qualitative analysis of these sessions suggested that the "look and feel" of the website influences the first impression. However, other factors played an important role when participants turned to a more detailed evaluation. To explore these factors, the authors conducted an online survey with more than 2500 participants who sought advised online. The authors identified three factors that influence the detailed evaluation of online advice: source credibility, personalization, and predictability. Further analysis showed that source credibility was the most important factor when users turn to detailed evaluation about online advices.

Kim and Moon conducted four consecutive studies to explore how to trigger a feeling of trust in cyber-banking systems (text-based, videotex, and online interfaces) [23]. They found correlations between design factors and four emotional factors: symmetry, trustworthiness, awkwardness, and elegance. Trustworthiness, in particular, was determined by the main clipart and the color of the interface.

Date	CTR	N	Date	CTR	N
Tu Oct 01	15%	97,585	Tu Oct 15	16%	73,370
We Oct 02	15%	96,076	We Oct 16	18%	85,266
Th Oct 03	15%	104,075	Th Oct 17	15%	68,947
Fr Oct 04	16%	84,165	Fr Oct 18	11%	132,410
Sa Oct 05	15%	80,433	Sa Oct 19	10%	99,778
Su Oct 06	15%	77,931	Su Oct 20	12%	95,163
Mo Oct 07	16%	80,640	Mo Oct 21	14%	91,651
Tu Oct 08	17%	90,356	Tu Oct 22	21%	131,700
We Oct 09	21%	145,893	We Oct 23	18%	121,944
Th Oct 10	21%	96,159	Th Oct 24	24%	151,387
Fr Oct 11	23%	93,059	Fr Oct 25	27%	117,002
Sa Oct 12	15%	79,295	Sa Oct 26	14%	64,740
Su Oct 13	15%	79,134	Su Oct 27	14%	70,713
Mo Oct 14	18%	89,180	Mo Oct 28	15%	59,567

Table 1: Chrome malware warning click-through rates (CTRs) and sample sizes for October 2013. Darker shaded values indicate higher CTRs. Note the wide variance in daily CTRs.

3. FIELD DATA: BROWSING HISTORY

Google Chrome's opt-in statistical reporting allows us to measure how end users respond to malware warnings in the field. This data allows us to see trends in how Chrome users react to malware warnings. We focus on the role of browsing history in users' malware warning decisions.

3.1 Motivation

Users respond very differently to malware warnings depending on the date. Within the last year (2013-2014), we have observed days where the CTR is as low as 7% or as high as 37%. This is a sizable range, and the degree of fluctuation is unique in Chrome: Chrome's SSL and phishing warning CTRs are stable over time [5].

To illustrate this phenomenon, Table 1 depicts the daily variation of the malware warning CTR in October 2013. Although the average is 17%, the daily CTR ranges from 10% to 27% within the month. "High" and "low" days tend to clump together in a series of similar days. The variation is not due to the day of the week.

As shown in Table 1, the CTR noticeably increased during October 22-25, 2013. We looked for changes in the Safe Browsing list that match these dates. Several high- and medium- reputation sites were added to and removed from the Safe Browsing malware list over a few days: `desitvforum.net` (3229 on the Alexa global ranking, 669 on the Alexa India ranking [1]), `php.net` (228 on the Alexa global ranking [2]), and `warriorforum.com` (117 on the Alexa global ranking [4]). This suggested to us that users might react differently to warnings on popular websites.

However, we are also aware of a counterexample on February 9, 2013. The compromise of an advertising network led to malware warnings on popular websites such as ABC News and YouTube. A few news outlets reported the incident, said that the cause was unclear, and recommended that users heed the warning [25, 36]. Social media posts rose to the top of search results, confirming that many people were seeing the warnings.[1] During these events, the daily CTR dropped

[1]For example:
`http://www.zyngaplayerforums.com/showthread.php?1748942-us-bernerverein-ch-malware-warning`,

from 15% to 8%. When the warnings were removed from the popular websites, the CTR returned to 15%. This indicates that the issue might be more complex than popularity alone – news media, word of mouth, and other factors might influence user behavior.

3.2 Hypotheses

The malware warning CTR varies daily, but so does the Safe Browsing list. Could changes in the Safe Browsing list be responsible for how people are reacting to the warning? We hypothesized that:

- H_1: People are more likely to ignore warnings about websites that they have visited before.
- H_2: When popular websites are on the Safe Browsing list, the CTR will be higher. That is, we expect to see a positive correlation between the CTR and the fraction of blocked websites that were previously visited.

3.3 Methodology

We leveraged Google Chrome's opt-in metrics to test our hypotheses. Google Chrome generates statistical reports on how many people click through the malware warning. We extended these reports for Google Chrome 32.

Implementation. We modified the malware warning to query the history manager. The history manager responds with the number of times that the user has previously visited the website that the warning is for. The malware warning then records two separate metrics: the overall CTR, and the CTR specifically for websites that the user has never visited. Only the history status and decision are recorded; the URL itself is *not* included in the statistics.

Sample. We analyzed metrics from the Google Chrome 32 stable channel from January 28, 2014 to February 24, 2014. Our overall sample size is 3,875,758 warning impressions.

Participation. During installation, Chrome users are asked whether they would like to send "crash reports and statistics" to Google. For users who choose to participate, Google Chrome sends statistical reports to Google. The reports are pseudonymous and cannot be traced back to the sending client once they are stored. We added new histogram entries to these reports. The reports do not contain any personal information (e.g., URLs are not allowed in the reports).

Limitations. Browsing history is an imperfect measure of prior experience with a website. Users clear their history and use multiple devices without syncing their history. In these cases, the user's decision will be misattributed to the "new site" distribution instead of the "visited site" distribution.

The "new site" and "visited site" distributions might contain multiple impressions from the same users, both within and across distributions. We rely on our very large sample size to mitigate this source of potential bias.

3.4 Results

Over the 28-day time period, users were twice as likely to ignore warnings about websites that were already in their browsing history. The average CTR for previously visited websites was 25.4%, whereas the average CTR for new websites was 13.1%. Our evidence supports H_1: the difference between the two average CTRs is statistically significant ($p < 0.0001$, one-tailed Z-test of proportions).

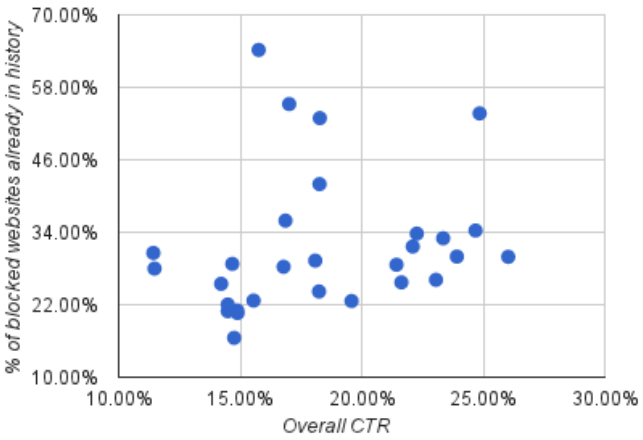


Figure 2: The relationship between the CTR and percentage of blocked websites that were already in the user's browsing history. Each point is a day. For 28 days in January-February 2014.

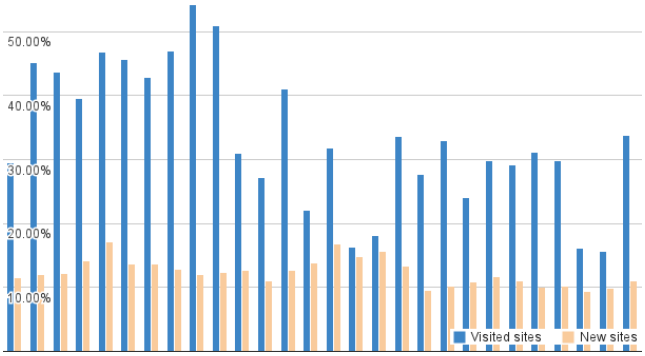


Figure 3: Daily CTR, separated by whether the website was already in the user's browsing history. For 28 days in January-February 2014.

However, the daily CTR is not correlated with the fraction of blocked websites that were previously visited. Figure 2 shows the lack of positive correlation. This means that the number of previously visited websites on the Safe Browsing list is not the cause of the daily variance. A linear regression gave a slope (0.07495) that was not significantly different from 0 ($t = 1.069$, $p = 0.295$), so we fail to reject the null hypothesis for H_2. This is a surprising result: we had expected that H_2 would follow from H_1.

Figure 3 illustrates why our data supports H_1 but not H_2. The CTR for warnings on new websites remains fairly stable over time (9.3% to 17.2%; stdev=2.1%), but the CTR for warnings on previously visited websites varies quite widely (15.6% to 54.3%; stdev=10.9%). Most of the daily variance in the overall CTR can be attributed to the variance within the visited website warnings. This suggests that a second unknown factor — such as reports from the media, word of mouth, or the quality or desirability of the website — may also be influencing user behavior. The unknown factor has a greater effect on user decisions when the destination website is already in the user's browsing history.

`http://answers.yahoo.com/question/index?qid=20130209134718AAhnNZX`, `http://www.reddit.com/r/Malware/comments/187of3/malware_warning_popping_up_everywhere_today/`

4. MTURK: METHODOLOGY

Section 3 showed that users are more likely to ignore a warning if they have visited the destination website before. We hypothesize that this is because prior positive experiences contribute to a website's reputation, and users are less likely to believe malware warnings for high-reputation websites. To explore the role of reputation in malware warning decisions, we set up an online, survey-based experiment.

We asked 1,397 Mechanical Turk workers to tell us how they would react to screenshots of Google Chrome malware warnings. In one experiment, we asked participants to respond to warnings on high- and low-reputation websites (YouTube or an unknown personal blog). In another experiment, we asked participants to respond to warnings that were linked from high- and low-reputation referrers (a friend's Facebook status or a low-quality lyrics website). We also tested minor variations of both experiments to evaluate how the specific question wording affected responses.

4.1 Research Questions

We focus on two questions related to reputation:

- Does the reputation of the referrer (i.e., the source that linked to the warning) affect how users respond to malware warnings?
- Does the reputation of the destination (i.e., the site with the warning) affect how users respond to malware warnings?

Reputation refers to a perception of quality. It can be established via prior personal experience (i.e., browsing history), brand recognition, word of mouth, or other factors.

4.2 Experiment Scenarios

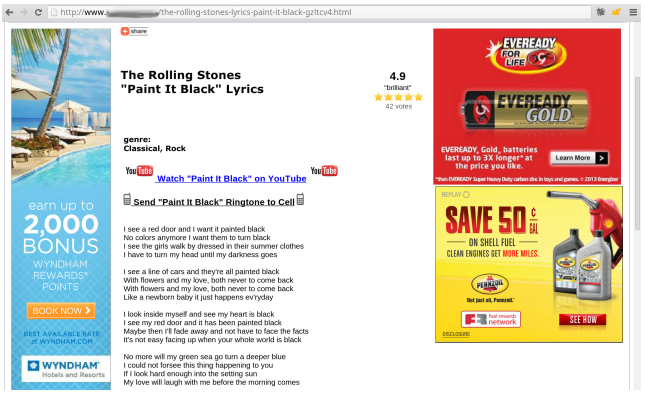
We presented participants with scenarios in which a referrer links them to a destination website with a warning. We created three scenarios (Figure 4):

1. Low-reputation referrer (lyrics website) → high-reputation destination (YouTube)
2. High-reputation referrer (friend's Facebook status) → high-reputation destination (YouTube)
3. High-reputation referrer (friend's Facebook status) → low-reputation destination (low-reputation blog)

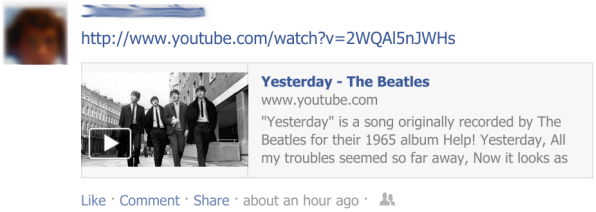
We ran two within-subjects experiments with these scenarios. The **referrer experiment** asked participants about scenarios 1 and 2 in a random order to evaluate the effect of the referrer's reputation. The **destination experiment** asked participants about scenarios 2 and 3 in a random order to evaluate the effect of the destination's reputation.

For the referrer experiment, we chose a friend's Facebook status to represent a high-reputation referrer because Facebook is a common way of exchanging links. We used a lyrics website for the low-reputation referrer because lyrics websites have poor reputations as sources of malware and unwanted advertisements [26, 38].

For the destination experiment, we chose YouTube as an example of a high-reputation destination because it is a highly popular, family-friendly website. We selected a little-trafficked personal blog to represent a low-reputation destination. A branding question at the beginning of the survey



(a) Low-reputation referrer that links to a high-reputation destination



(b) High-reputation referrer that links to a high-reputation destination

(c) High-reputation referrer that links to a low-reputation destination; its URL has been partly obscured for the paper

Figure 4: Screenshots from the scenarios used for the experiments. Each was followed by a screenshot of a Chrome malware warning.

confirmed that participants were familiar with YouTube but not the blog (100% and 0.01% of participants said they were familiar with the two websites, respectively).

We could have also tested a fourth scenario with a low-reputation referrer and a low-reputation destination. However, a pilot study suggested that this was not necessary because participants' self-reported click-through rates for scenario 3 were already close to 0%. As a result, we did not think that a fourth scenario would yield additional results; we decided to focus on the other three scenarios to increase our sample sizes within our budget.

4.3 Wording Choices

Our survey wording could influence the results. To account for this, we tested multiple versions of the two experiments. Prior work has similarly run multiple versions of experiments to look for biases [24, 32]. We tested five between-subjects versions of the destination experiment (three roles, priming, and interactive) and three between-subjects versions of the referrer experiment (three roles).

Roles. We asked participants to imagine the scenario as if they were personally experiencing the situation, advising a friend, or pretending to be someone else.

- Personal experience is a natural way to frame a scenario, e.g., *"Imagine that you are visiting..."*
- Researchers use the "helping a friend" role to reduce social desirability bias [14]. We asked participants to help their best friend decide what to do about a warning. For example, *"Imagine that your best friend is visiting www.facebook.com to check friends' latest updates."*
- We asked participants to pretend to be someone else who is completing a task. Researchers use this type of role to reduce the risk of an experiment, reduce social desirability bias, and/or motivate participants to complete an imaginary task [32, 33, 41]. Having a task to complete is intended to mimic real life situations. For example, *"Imagine that you are a moderator of a 'Music Video of the Day' Facebook group that only your friends can join. Your friends post YouTube videos they like to the group, and you visit them to record the number of views. The winner of the day is the most viewed video. Imagine that you are visiting www.facebook.com to check the videos posted to the group today."*

Priming. Prior work offers conflicting guidance on the effects of priming on security research [12, 32, 40]. Thus, we took care to avoid mentioning risk and used neutral language (e.g., "page" or "red page" instead of warning) in all but one version of the experimental survey. One survey variation intentionally began with a paragraph that discussed malicious software and potential risks in order to prime participants. It also used the word "warning" in the prompts.

Interactivity. In one variant of the destination experiment, we provided participants with the ability to read more information about the warnings before deciding. This variant was interactive: participants could choose from any of the available buttons and walk through a series of screenshots until reaching a decision. For example, a participant could select the option "click on 'Advanced' link" to see additional options (Figure 5 shows the additional options). From there, the participant could choose "click on 'Details about the problems on this website"' to see the diagnostic page (Figure 7). This increased the length and complexity of the survey but allowed us to study the effect of providing all of the available options.

4.4 Survey Walkthrough

We created eight surveys: five variations of the destination experiment, and three variations of the referrer experiment. All of the surveys were similarly structured, although the variants had slightly different wording for the scenarios. Each survey had two scenarios. The following illustrates the survey's outline (with a full example in the appendix):

1. Brand familiarity. "Which of these websites have you heard of?" We alphabetically listed the three websites that appear in the survey (Facebook, the blog, and YouTube) and four decoy websites.

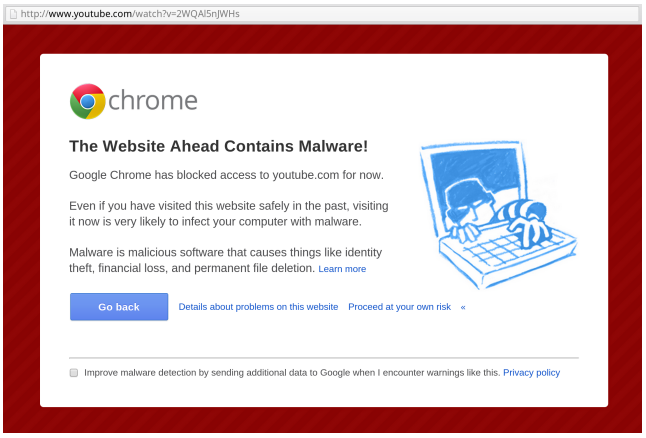


Figure 5: The malware warning, with the "Advanced" options exposed

2. Scenario introduction. For example, "Imagine that you are searching for the lyrics for the song 'Paint It Black'. You find the lyrics on the website shown below. [screenshot]" We then asked a comprehension question to ensure that participants looked at the screenshot. For example, "Which band recorded the song shown in the screenshot above?"
3. Reaction to warning. The survey instructed the participant to imagine she had clicked on a link. It then displayed a screenshot of a Chrome malware warning and asked: "What would you do?" (multiple choice) and "Why?" (short essay).
4. Second scenario. Steps 2 and 3 were repeated for a different scenario.
5. Ramifications. Questions about the ramifications of clicking through the warnings, e.g., "Which outcome is the most likely if you clicked through the red page to proceed from the lyrics website to youtube.com?"
6. Real world behavior. "How would you typically react if you saw a similar red page when trying to visit a website in your day-to-day life?" Also, "Before this survey, had you ever seen a similar red page when trying to visit any website?" If yes: "What happened the last time you saw a similar red page when trying to visit a website?"
7. Demographics. Demographic questions to measure their reputation with the websites in the survey, technical ability, and security knowledge. Also basic demographic information such as age and education level.

The questions were a mix of closed- and open-ended questions, giving us a mix of quantitative and qualitative data.

We randomly assigned participants to one of eight versions of the survey and randomized the order in which the scenarios were displayed. We also randomized the choice order for multiple-choice questions, with a caveat: we kept the choice order constant between similar questions to avoid confusion. (For example, if "Go back" were the first choice for the first scenario's warning question, it would also be the first choice for the second scenario's warning question).

4.5 Recruitment

We used Amazon Mechanical Turk to recruit participants. We posted a task that invited Mechanical Turk workers to "take a survey about how you browse the web." Participants were compensated $1 each for a survey that took 6 minutes on average to complete. We limited the survey to people aged 18 and older in the United States. Participants also had to have an approval rate of 95% or higher. In the instructions, we said that the survey was limited to Google Chrome users. To enforce this constraint, we discarded responses from users who said that they do not use Google Chrome; however, we still paid non-Chrome users to avoid incentivising lying.

We discarded responses from participants who appeared to be cheating. For example, we excluded participants who responded more than once or tried to use incorrect survey completion codes. Each survey also contained two scenario comprehension questions with gold standard answers (see Pages 3 and 4 in the appendix).

4.6 Demographics

We had a total of 1,386 survey responses after excluding submissions that did not meet our requirements. Table 2 shows a summary of participants' demographics. A majority of participants are active Facebook and YouTube users who reported checking Facebook and watching a YouTube video at least once in the week prior to the survey.

Our sample population is likely more tech-savvy than the average Internet user. To assess participants' technical abilities, we asked a multiple-choice question: *"What would you do if your wireless router at home were not working?"* 73% of participants reported that they would fix the problem themselves, which we assume is indicative of relatively high technical confidence. We also asked participants two multiple choice security questions: *"What is a computer firewall?"* and *"What is a public key certificate?"* 44% of participants answered both security questions correctly.

We also asked participants about their highest level of education. About 10% of participants have a post graduate degree, 35% have a bachelor's degree, 31% have some college, 12% have an associates degree, and 11% have a high school diploma or GED. The rest have some high school education without obtaining a diploma.

4.7 Statistical Analysis

We used logistic regression to test for statistical significance of our experimental treatments (destination, referrer, and wording variants). We fitted two logistic regression models, one for the destination experiment and one for the referrer experiment. Except where otherwise noted, p-values for significance testing come from Wald tests at the $\alpha = 0.05$ level of whether the fitted regression coefficients are significantly different from zero. Logistic regression is similar to ANOVA analysis in that it automatically accounts for multiple statistical tests, but unlike standard ANOVA, allows us to model experiments with a binary outcome (in our case, the binary outcome is whether the participant would click through the warning or not).

5. MTURK: LIMITATIONS

Our results must be viewed within the context of the limitations of this type of study.

5.1 Generalizability

Our demographic questions show that most participants are active Internet users who would feel comfortable tinkering with a wireless home router. As such, caution should be exercised in generalizing our results, especially to others with lower levels of Internet exposure. However, our survey population represents an important demographic because active web browsing increases the chances of seeing a warning. Future work could extend this research to groups of users who use the Internet less and are less comfortable with technology.

5.2 Interpretation of Study CTRs

Our experiment asked participants how they would react to warnings under hypothetical circumstances. These artificial conditions differ from real life; our online tasks lacked the urgency that participants might experience in real life, and our experiment posed no real risk. To distinguish our experimental survey results from field data, we refer to the rate at which participants say they would proceed through a warning as the *self-reported click-through rate* (SRCTR).

The primary goal of our work is to evaluate how the reputations of referrers and destinations influence behavior. To this end, we compare SRCTRs between high- and low-reputation conditions. Any bias inherent in our study methodology applies equally to the different conditions, so participants were not biased in favor of any particular condition. In addition, the effect of any inherent bias is minimized by randomizing the order of the tasks (e.g. low-reputation task first), the careful wording of the survey (e.g. using different roles, priming vs. no priming), and the random assignment of participants to different conditions. We therefore interpret SRCTRs as being able to reveal differences between conditions even though they may not be indicative of the absolute value of real-world CTRs.

Despite these limitations, we consider participants' statements to be representative of thoughts that would occur in real encounters with malware warnings, even though they might ultimately act differently due to competing priorities.

5.3 Wording Choices

We were concerned that the wording of our survey instrument would introduce bias. To try to account for this, we tested multiple versions of the survey instrument. Table 3 breaks down the results pertaining to the different survey instrument versions by condition and variation.

Roles. The role did not change most participants' responses. We do not observe a significant difference in SRCTR between roles for the high-reputation destination (37%, 38%, 36%), high-reputation referrer (31%, 31%, 33%), or low-reputation referrer conditions (27%, 28%, 24%). However, playing someone else leads to a higher SRCTR for the low-reputation destination condition (3%, 3%, 8%). The difference is small but statistically significant (p=0.04).

Priming. The type of priming that we used did not influence participants' decisions. The "priming" and "personal" variants are identical except for the presence or absence of priming text and the use of the word "warning" in the prompts instead of the word "red page". The priming variant yields a slightly lower SRCTR (31% vs. 37%) for the high-reputation destination condition, but the difference is not statistically significant (p=0.28). For the low-reputation

Table 2: Characteristics of online, survey-based experiment participants

Experiment	Word Variant	N	% Male	Mean Age	Tech Confident	Security Savvy	% actively use... Facebook	YouTube
Destination exp	Personal ("you")	174	58%	30	76%	42%	87%	96%
Destination exp	Helping a friend	174	54%	30	67%	38%	82%	94%
Destination exp	Playing someone else	173	62%	30	72%	42%	83%	94%
Destination exp	Priming + personal	175	59%	32	74%	59%	86%	94%
Destination exp	Interactive + personal	174	59%	33	75%	47%	87%	93%
Referrer exp	Personal ("you")	172	54%	31	73%	49%	89%	96%
Referrer exp	Helping a friend	171	56%	31	72%	41%	88%	98%
Referrer exp	Playing someone else	173	67%	31	75%	46%	79%	93%

Table 3: Results for the online, survey-based experiment. Darker shaded values indicate higher SRCTRs.

Experiment	Wording Variant	High-Reputation SRCTR	N	Low-Reputation SRCTR	N	Aggregate SRCTR
Destination experiment	Personal ("you")	37%	159	3%	171	19%
Destination experiment	Helping a friend	38%	158	3%	160	20%
Destination experiment	Playing someone else	36%	151	8%	156	22%
Destination experiment	Priming + personal	31%	158	4%	169	17%
Destination experiment	Interactive + personal	15%	162	2%	167	9%
Referrer experiment	Personal ("you")	31%	163	27%	161	29%
Referrer experiment	Helping a friend	31%	166	28%	165	30%
Referrer experiment	Playing someone else	33%	162	24%	161	28%

destination condition, the two SRCTRs are within a percent. This suggests that the type of priming that we used has little effect on participants' responses. This finding is similar to some prior findings about priming in security studies [12,32], although it conflicts with others [40].

Interactivity. For the high-reputation destination, participants in the interactive variant were less likely to proceed than participants in the non-interactive "personal" variant (15% vs. 37%). The difference is statistically significant ($p<0.0001$). The difference is most likely explained by an extra step that participants in the interactive variant had to take in order to click through the warning: they had to first choose an "Advanced" option, while those in non-interactive conditions had the option to click through on the first screenshot they saw.

6. MTURK: RESULTS

We present the results of our experiments in terms of self-reported click-through rates (SRCTRs) and participant quotes. We also present common misconceptions and points of frustration from the short essay responses.

6.1 Destination Reputation

We asked participants to respond to warnings on high- and low-reputation sites, and we find that the destination's reputation strongly affects how participants react to hypothetical warning scenarios. As Table 3 shows, many more participants claim that they would ignore the warning for a high-reputation destination and heed a warning for a low-reputation destination. The difference between the two SRCTRs is statistically significant overall ($p<0.0001$).

Many participants discussed brand reputation and prior personal experience. E.g.,

> *I have never heard of this site [the blog] so I wouldn't trust it.*
>
> *YouTube is well known website. I'd assume that the malware block is in error.*
>
> *Because I frequent youtube.com a lot and I have never gotten any malware*
>
> *Youtube.com is a trusted site that I use almost everyday and have not had any problems with.*

A small number of participants also noticed that the blog is hosted on Blogspot. They said that they would proceed to the blog because they trusted Blogspot.

Additionally, there was a correlation between the reputation of the destination and participants' perceived risk of ignoring the warning. We asked participants about the ramifications of ignoring the malware warning (e.g., "Which outcome is the most likely if you clicked through the red page to proceed to youtube.com?"), and the answers differ based on the type of destination. Table 4 shows the percentage of participants who think a bad outcome (i.e., "My computer would be infected by malware.") is most likely to occur. Fewer participants believe there will be a bad outcome when the destination is high-reputation ($\chi^2=265.35$, df=1, $p<0.0001$).

6.2 Referrer Reputation

We asked participants to respond to warnings on sites linked from high- and low reputation referrers. We find that the referrer's reputation had only a weak effect on how participants reacted to the warning scenarios. As Table 3 shows, slightly more participants claim that they would ignore a warning on a site linked from a high-reputation re-

Table 4: Perceived risk of ignoring a malware warning

Experiment	Scenario	% Bad Outcome	N
Destination experiment	High-reputation (YouTube)	34%	823
Destination experiment	Low-reputation (blog)	77%	792
Referrer experiment	High-reputation (Facebook friend)	51%	454
Referrer experiment	Low-reputation (lyrics site)	58%	462

ferrer. However, the difference between the two SRCTRs is not statistically significant (p=0.36).

In the open-ended question responses, some participants said that their trust in friends or mistrust of lyrics sites would influence their decision. For example,

> *Malware is dangerous, and most of those lyrics sites are shady*
>
> *I find it harder to believe [the warning] when my facebook friend just posted it and had no problems.*
>
> *I presume that visiting youtube from a facebook link would be safe.*

One participant summarized the difference between the Facebook status update from a friend and the lyrics website:

> *This [lyrics] website is less reliable than my friend who posted the link so I don't know if I should trust it.*

Some participants' responses indicated that they were considering both the reputation of the referrer and the reputation of the destination (YouTube). For example:

> *[I] trust youtube, but I don't necessarily trust the lyrics website*

There was also a weak relationship between the reputation of the referrer and participants' risk perception. We asked participants about the ramifications of ignoring the malware warning (e.g., "Which outcome is the most likely if you clicked through the red page to proceed to youtube.com?"), and the answers differ slightly based on the type of referrer. Table 4 shows the percentage of participants who think a bad outcome ("My computer would be infected by malware.") is most likely to occur if they ignore the warning. Fewer people believe there will be a bad outcome when the referrer is high-reputation (χ^2=4.13, df=1, p<0.05). Although the difference is statistically significant, the practical difference between the conditions is small.

Overall, we found little difference between the high- and low-reputation referrer conditions.

6.3 Getting More Information

There are two ways to get more information about a Chrome malware warning. First, the warning includes a "Learn more" link in the last paragraph. This leads to Google's general online security guide (Figure 6). Second, clicking on the "Advanced" link triggers the appearance of a link named "Details about problems on this website." That link leads to a diagnostic page with technical information (Figure 7).

The interactive variant of the destination experiment allowed participants to access additional information about

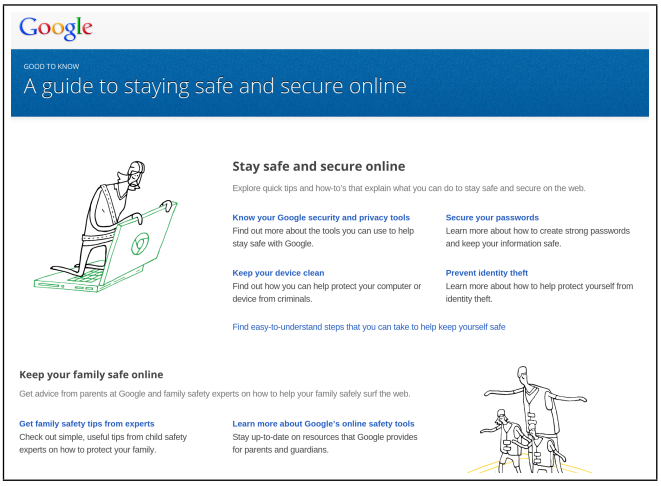


Figure 6: Google's online security guide

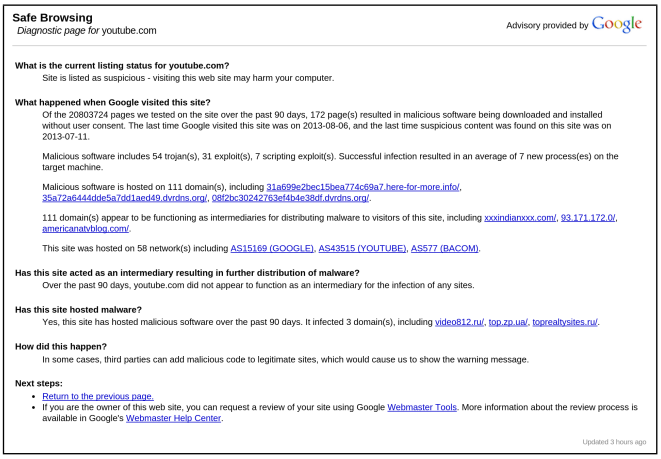


Figure 7: Safe Browsing diagnostic page

warnings before making a decision. Participants were more likely to want additional information in the high-reputation destination scenario. 16% of participants in the interactive condition navigated to the online security guide (6%) or the diagnostic page (9%). In contrast, only 3% of participants sought more information from either source when the warning was for a low-reputation destination.

Unfortunately, participants who saw more information proceeded at the same rate as other participants. Furthermore, participants said they were not satisfied with the content or amount of information on the pages that provide more information. Participants felt that the online security guide was too generic or too lengthy to be helpful:

> *Close out [the page], I would want to know more*

specifically why the warning was brought up for the particular site.

I would probably ignore it and just go to youtube site. There's too much general information on this page for it to be helpful.

It's too much to read.

Although the diagnostic page provides more detailed information, participants were still frustrated by it. Several said that they would look elsewhere. For example,

I would likely not continue, instead I would go to a search engine there and search for the site. This warning is inconsistent with what I believe the integrity of the site is. But, it is possible that this is some sort of advanced hijacking technique.

I would close the tab and check the URL in firefox to see what info I got there. I'd probably also post to twitter and ask if anyone else was getting this info and if so had anyone seen any articles/posts about what kind of malware and who had infected it.

Additionally, several participants in the interactive variant and other variants of the destination experiment indicated that they would seek external information about the warning before making a decision. In particular, participants in the high-reputation destination scenario said they would seek external information from sources such as search engines, news articles, and social media websites.

Something is screwed up, given that it's YouTube. I would search the internet for others reporting the problem.

I would be worried that someone compromised Youtube. I'd try to research and see if this was widespread news (as it likely would be if it were true), or just a problem with Chrome.

I would reenter my search to make sure I didn't click on a link that was masked. If it still showed malware I'd watch news sites to make sure youtube wasn't compromised.

Search the net and find any information on why chrome is blocking youtube.

6.4 Misconceptions

Participants mentioned three notable misconceptions that could hinder the effectiveness of the malware warning.

Protective technology. Some participants believe that they are safe from malware because of protective technology such as anti-virus software or their operating system. E.g.:

I use Linux I'm not afraid of anything.

Because i own a mac and i dont worry about that stuff

I would still proceed knowing I have an anti virus

Other participants had similar responses. These beliefs are dangerous; anti-virus software does not prevent drive-by download attacks, and some drive-by download attacks can succeed on Linux and Mac computers.

Confusion with other warnings. Participants also confused malware warnings with the SSL warning. From their responses, it sounded like they had encountered SSL warnings that they considered to be false positives. For example, one person said:

I know and trust youtube, sometimes the internet browser doesn't have the right certificate.

I want to learn why chrome thinks the site contains malware. Sometimes it might just involve something like an expired security certificate

We also asked participants about prior warnings. About 77% of participants remembered seeing a similar warning in the past. We asked participants to elaborate, and some responses referred to the SSL warning as if it were the same warning. For example:

I believe I got [the warning] because of some discrepancy between http and https.

Identity of the destination site. In the referrer experiment, some participants suspected that the lyrics site might have linked to a site that was not actually YouTube. E.g.:

I don't trust lyrics sites very much, especially ones with those kinds of ads. They could have possibly altered that link to lead to somewhere malicious.

I don't trust redirects from lyric sites.

However, the screenshot in the survey showed a warning for `youtube.com`. The screenshot included the omnibox (which said "`http://www.youtube.com`"), and the malware warning itself includes the destination URL in the text. These participants either did not know how to check the identity of the destination site, did not think to check those identity indicators, or did not trust those identity indicators.

7. IMPLICATIONS

We discuss the implications of our findings and make suggestions for improvement to the warnings. Some of the suggestions have already been adopted by Google Chrome. We also highlight additional open questions and challenges.

7.1 Gaining Users' Trust

Our findings suggest that end users may trust other parties more than they trust the browser's malware warning. In particular, some study participants trusted the reputation of the destination site more than the warning. Some participants also trusted their anti-virus software or operating system to protect them. We recommend adjustments that could increase users' belief in the warnings.

High-reputation destinations. Many participants could not believe that a site like YouTube would be malicious, causing the SRCTR for the high-reputation destination to

be much higher than the SRCTR for the low-reputation destination.[2] Participants' open-ended responses show that this is due to trust in the brand, prior positive experiences with the site, or some combination of the two. Our field data demonstrates that this same effect happens in naturalistic settings for websites that users have previously visited.

We recommend using a special warning for high-reputation destinations. The warning would need to acknowledge that the website is usually safe but emphasize that it has been temporarily compromised. This should match users' mental model better than telling them that the website itself is malicious. One challenge is how to identify high-reputation destinations; a possible solution is to treat all sites in a user's history as high-reputation, combined with a pre-loaded list of high-reputation destinations. Prior literature on site credibility may help guide the identification of high-reputation destinations [8, 16, 18, 22, 27, 39].

How to communicate this information effectively is an open question. The warning already attempts to address this with its third sentence: "Even if you have visited this website safely in the past, visiting it now is very likely to infect your computer with malware." It is not clear whether participants missed this information because they did not read it, or whether they simply found it unconvincing. In our future work, we will be experimenting with different approaches to address this issue.

More information. Our findings suggest that some people are conflicted when they encounter warnings on high-reputation sites and want more information to resolve this conflict. In our study, a notable minority of participants expressed a desire for more information about the warning on a high-reputation destination. There are also ample examples of users asking for more information about malware warnings on web forums and Twitter.

We have already updated the "Learn More" link and diagnostic page in response to this concern. Our participants complained that the general online security guide was too vague, so we modified the "Learn More" link to point to the Safe Browsing Transparency Report. The Transparency Report provides more specific information about why Chrome blocks websites. This change will take effect in Chrome 37. Although the diagnostic page was intended to be more specific, participants found it confusing and unsatisfying. We have built a new version of the diagnostic page that should better address participants' needs. It will be launched in July 2014. Future work is needed to determine whether the new "Learn More" link and diagnostic page will sway undecided users.

Protective technology. Some participants thought that they did not need to heed the malware warning because their anti-virus software or operating system would protect them. Such (often inaccurate) beliefs could expose people to very real risks. We recommend that the warning should specify that neither is an adequate defense against a malicious site.

In the hope of reaching Mac users, Chrome for Mac OS X changes the phrase "your computer" to "your Mac" in the warning. A limitation of our study is that we showed the default PC version ("your computer") to all participants. However, we recommend that this should be made more explicit. People may not notice the subtle reference to Macs.

Role of the referrer. The reputation of the referrer played only a minor role in participants' decisions. We consider three possible reasons: (1) participants do not use the referrer's reputation to make a decision, (2) our experiment lacked the necessary statistical power to identify a small effect, or (3) participants did not consider Facebook statuses to be high-reputation because of the prevalence of Facebook spam. With respect to the third explanation, participants had inconsistent views of the Facebook status:

> *There are always issues like this on facebook. I would not proceed.*
>
> *Someone could have hacked that person's facebook account and posted a false link to a virus.*
>
> *I would trust my friend not to post a bad link but I would be afraid to continue on based on the screen that showed up.*
>
> *from facebook i am less likely to think there is malware associated with the link, especially a youtube link.*

It is possible that more of a difference would appear if the high-reputation referrer were a news website, text message, or other mode of delivery.

We do not have any recommendations to offer about the referrer at this time. However, future work could further investigate the role of different referrers.

7.2 Differentiate Malware and SSL Warnings

Some participants confused Chrome's malware and SSL warnings. This is undesirable because SSL warnings are often false positives; we worry that this devalues user perception of the malware warning. Furthermore, malware warnings put the security and privacy of the whole computer at risk, not just the confidentiality and integrity of a single domain. Ideally, malware warnings should be taken more seriously than SSL warnings.

A possible solution is to make the two warnings more distinct. At the time of our study, both warnings had predominantly red color schemes. We modified Google Chrome's SSL warning to have a yellow-orange background, starting in Chrome 32. In future work, we will investigate if further changes may still be needed to help end users distinguish between the two types of warnings.

7.3 Survey Wording

The type of role and priming with risk information made little difference in participants' responses. Our finding on priming with risk information reinforces similar findings in prior studies [12, 32]. However, interactivity changed participants' responses to our scenarios.

In all but one variant, we asked participants to choose between proceeding and returning to the previous page. In the interactive variant, participants were able to view additional information before deciding. The additional choices significantly decreased the SRCTR in the high-reputation destination condition. However, this was not due to the additional information itself; people who viewed the additional information chose to proceed at the same rate as

[2]We cannot be certain of effect size because the interactive and non-interactive survey variants yielded different gaps between the high- and low-reputation destinations. However, the gap was large in all variants.

other participants. Instead, this suggests that the presence of more choices changed how participants responded to the question. Since we do not know the ground truth for these participants, we do not know whether the interactive or non-interactive variant better represents the participants' real world behavior. Future work could further investigate this effect.

7.4 Open Question: Daily Variance

The malware warning CTR in Chrome fluctuates over time in the field. Discovering the cause of this fluctuation could help warning designers reduce the CTR. Ideally, the warning would be modified to address the situations that lead to sudden increases in CTR.

Prior experience. As discussed in Section 3, we originally hypothesized that the daily variance was due to the daily rate at which familiar websites appeared on the Safe Browsing list. However, our data did not support this hypothesis. Nonetheless, we did discover one clue: the daily variance is larger for warnings on previously visited websites than for warnings on new websites. The daily variance might therefore be related to prior experience with the website. For example, it could be due to the quality of the website or how much the user likes the website.

News stories and social media. Another possible explanation for the daily variance is that high-profile news stories or social media discussions influence users' reactions to warnings. Warnings on popular websites are sometimes mentioned in the news, and we have seen people turn to social media (Twitter, message boards, etc.) to ask each other about warnings on high-profile websites. This might be more likely for previously visited websites, since users might find those warnings more puzzling. Several participants in the Mechanical Turk study said that they would search for more information if they saw a warning for YouTube.

For example, Section 3.1 describes an event on February 9 that was covered in the press and discussed by many on social media. A similar event took place the week before, on February 4, 2013. An advertising network was put on the Safe Browsing malware list because its homepage was compromised. It was initially unclear whether its advertisement serving infrastructure was compromised as well. This caused malware warnings to appear on several high-reputation sites that use the advertising network (e.g., Huffington Post, Washington Post, The New York Times). This event caused the number of warning impressions to dramatically increase: from approximately 100,000 on a "typical" day to 1,254,520 on February 4 (within the subset of the population that shares statistics with Google).

The two events on February 4 and February 9 were fairly similar. Both led to malware warnings on popular websites, made the news, and swamped social media websites. However, users responded differently to the two events: they clicked through only 8% of warnings on February 9 but 15% of warnings on February 4. What was different? On February 9, news stories and social media posts exhorted users to heed the warning. Users saw the recommendation, and the CTR decreased to 8%. In contrast, the advertising company involved in the February 4 event issued a statement saying that the warning was a "false alarm" [28], and news outlets reported that the warnings were false positives [20, 29].

Anectodal evidence is insufficient to substantiate a hypothesis, but the role of news stories and social media should be investigated further. Measuring the influence of news and social media on user behavior is left for future work.

8. CONCLUDING SUMMARY

Our goal is to understand why users ignore malware warnings. To this end, we analyzed 3,875,758 Chrome malware warning impressions from the field and ran an online, survey-based controlled experiment with 1,397 participants.

We found that users in the field are twice as likely to ignore a malware warning from Chrome if the blocked website is already in their browsing history. This suggests that users are less likely to believe a malware warning if they have prior experiences with a website. Participants in our online study echoed this sentiment: they said that they did not believe that a popular, high-quality site could be malicious. Furthermore, participants' quotes indicated that some people have misconceptions about the warning; for example, some participants confused the malware and SSL warnings.

Our primary recommendation is that malware warnings need to be changed to convey that high-reputation websites can be temporarily compromised. This will address the unfortunately common situation where malware authors take control of popular websites to spread malware. Some participants also expressed a desire for clear, contextual information to help them make a decision. To address this latter concern, we adjusted the Chrome warning's "Learn More" link and built a new diagnostic page. Our work on improving the Chrome malware warning continues.

Data Collection Ethics

All Chrome metrics are subject to privacy policies. Participants opt in to share statistics with Google, and participants can later opt out by changing their settings [19]. Our new statistics were reviewed according to the Chromium review process. We did not collect or analyze any private or personally identifiable information.

Our Amazon Mechanical Turk experiment asked participants about hypothetical scenarios and prior warning encounters. None of this data is private or sensitive. We also collected optional demographic information on age, gender, social media usage, and education. Our institution does not have an Institutional Review Board (IRB), so it was not subject to IRB review; however, multiple researchers who have received human subjects training reviewed the survey instrument prior to the experiment. We paid the study participants (Amazon Mechanical Turk workers) a rate intended to mimic California's minimum wage.

9. REFERENCES

[1] Desitvforum.com Site Info. http://www.alexa.com/siteinfo/desitvforum.net. Accessed: 2013-11-7.

[2] Php.net Site Info. http://www.alexa.com/siteinfo/php.net. Accessed: 2013-11-7.

[3] Safe Browsing API. https://developers.google.com/safe-browsing/. Accessed: 2013-11-9.

[4] Warriorforum.com Site Info. http://www.alexa.com/siteinfo/warriorforum.com. Accessed: 2013-11-7.

[5] D. Akhawe and A. P. Felt. Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness. In *Proceedings of the 22th USENIX Security Symposium*, 2013.

[6] F. Asgharpour, D. Liu, and L. J. Camp. Mental models of computer security risks. In *WEIS*, 2007.

[7] P. Briggs, B. Burford, A. De Angeli, and P. Lynch. Trust in online advice. *Social Science Computer Review*, 20(3):321–332, 2002.

[8] P. Briggs, B. Burford, A. De Angeli, and P. Lynch. Trust in online advice. *Social Science Computer Review*, 20(3):321–332, 2002.

[9] R. Dhamija, J. Tygar, and M. Hearst. Why Phishing Works. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2006.

[10] S. Egelman, L. F. Cranor, and J. Hong. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1065–1074, New York, NY, USA, 2008. ACM.

[11] S. Egelman and S. Schechter. The importance of being earnest [in security warnings]. In *The International Conference on Financial Cryptography and Data Security*, FC '13, 2013.

[12] S. Fahl, M. Harbach, Y. Acar, and M. Smith. On The Ecological Validity of a Password Study. In *Proceedings of the Symposium On Usable Privacy and Security*, 2013.

[13] A. P. Felt, R. W. Reeder, H. Almuhimedi, and S. Consolvo. Experimenting at scale with google chrome's ssl warning. 2014.

[14] R. J. Fisher. Social Desirability Bias and the Validity of Indirect Questioning. *Journal of Consumer Research*, 20(2), September 1993.

[15] B. Fogg, J. Marshall, O. Laraki, A. Osipovich, C. Varma, N. Fang, J. Paul, A. Rangnekar, J. Shon, P. Swani, et al. What makes web sites credible?: a report on a large quantitative study. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 61–68. ACM, 2001.

[16] B. Fogg, J. Marshall, O. Laraki, A. Osipovich, C. Varma, N. Fang, J. Paul, A. Rangnekar, J. Shon, P. Swani, et al. What makes web sites credible?: a report on a large quantitative study. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 61–68. ACM, 2001.

[17] B. Fogg, C. Soohoo, D. R. Danielson, L. Marable, J. Stanford, and E. R. Tauber. How do users evaluate the credibility of web sites?: a study with over 2,500 participants. In *Proceedings of the 2003 conference on Designing for user experiences*, pages 1–15. ACM, 2003.

[18] B. Fogg, C. Soohoo, D. R. Danielson, L. Marable, J. Stanford, and E. R. Tauber. How do users evaluate the credibility of web sites?: a study with over 2,500 participants. In *Proceedings of the 2003 conference on Designing for user experiences*, pages 1–15. ACM, 2003.

[19] Google Chrome support. Usage statistics and crash reports. https://support.google.com/chrome/answer/96817?hl=en.

[20] R. Greenfield. Why Malware Warnings Took Over the Internet Today. http://www.theatlanticwire.com/technology/2013/02/google-chrome-malware-warnings/61774/, February 2013.

[21] J. Haber. SmartScreen Application Reputation in IE9, May 2011.

[22] J. Kim and J. Y. Moon. Designing towards emotional usability in customer interfaces–trustworthiness of cyber-banking system interfaces. *Interacting with computers*, 10(1):1–29, 1998.

[23] J. Kim and J. Y. Moon. Designing towards emotional usability in customer interfacesÑtrustworthiness of cyber-banking system interfaces. *Interacting with computers*, 10(1):1–29, 1998.

[24] T. H.-J. Kim, P. Gupta, J. Han, E. Owusu, J. Hong, A. Perrig, and D. Gao. Oto: online trust oracle for user-centric trust establishment. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 391–403. ACM, 2012.

[25] E. Limer. Google Chrome Is Blocking a Bunch of Major Sites for Malware, Even YouTube, February 2013.

[26] McAfee. The web's most dangerous search terms. Report, 2009.

[27] M. J. Metzger. Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, 58(13):2078–2091, 2007.

[28] NetSeer. Twitter post. https://twitter.com/NetSeer/status/298498027369402368, February 2013.

[29] J. C. Owens. Google Chrome's NetSeer malware warning blocks websites, company says no virus distributed. http://www.mercurynews.com/ci_22515730/malware-warning-citing-netseer-blocks-google-chrome-users, February 2013.

[30] N. Provos. Safe Browsing - Protecting Web users for 5 Years and Counting. http://googleonlinesecurity.blogspot.com/2012/06/safe-browsing-protecting-web-users-for.html, June 2012.

[31] N. Provos, D. McNamee, P. Mavrommatis, K. Wang, N. Modadugu, et al. The ghost in the browser analysis of web-based malware. In *Proceedings of the First Workshop on Hot Topics in Understanding Botnets*, 2007.

[32] S. E. Schechter, R. Dhamija, A. Ozment, and I. Fischer. The Emperor's New Security Indicators. In *Proceedings of the IEEE Symposium on Security and Privacy*. IEEE, 2007.

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

[34] K. Šolic and V. Ilakovac. Security perception ot a portable pc user (the difference between medical doctors and engineers): A pilot study. *Medicinski Glasnik*, 6(2), 2009.

[35] A. Sotirakopoulos, K. Hawkey, and K. Beznosov. On the Challenges in Usable Security Lab Studies: Lessons Learned from Replicationg a Study on SSL Warnings. In *Proceedings of the Symposium on Usable Privacy and Security*, 2011.

[36] T. C. Sottek. Malware warnings ripple across the web just five days after last major incident. `http://www.theverge.com/2013/2/9/3971766/major-websites-hit-with-malware-warning`, February 2013.

[37] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F. Cranor. Crying Wolf: An Empirical Study of SSL Warning Effectiveness. In *Proceedings of the USENIX Security Symposium*, 2009.

[38] Y. Wang, D. Beck, X. Jiang, R. Roussev, C. Verbowski, S. Chen, and S. T. King. Automated Web Patrol with Strider HoneyMonkeys: Finding Web Sites That Exploit Browser Vulnerabilities. In *Proceedings of the Network and Distributed System Security Symposium*, 2006.

[39] C. N. Wathen and J. Burkell. Believe it or not: Factors influencing credibility on the web. *Journal of the American society for information science and technology*, 53(2):134–144, 2002.

[40] T. Whalen and K. M. Inkpen. Gathering Evidence: User of Visual Security Cues in Web Browsers. In *Proceedings of the Conference on Graphics Interfaces*, 2005.

[41] M. Wu, R. C. Miller, and S. L. Garfinkel. Do Security Toolbars Actually Prevent Phishing Attacks? In *ACM SIGCHI Conference on Human Factors in Computing Systems*, 2006.

APPENDIX

The following is a full example of the destination experiment survey, using the "personal" wording variant.

Mechanical Turk HIT Description

How do you browse the web?

We are conducting a survey about how you browse the web. This survey will ask you about how you would react to different situations on the web. The whole survey should not take more than 10 minutes. Please answer each question carefully and answer honestly. We will pay you $1 for your participation.

To participate:

1. You must be 18 years old or older.
2. You must be a Chrome user.
3. You must be in the United States while taking the survey.
4. You must be an English language speaker.
5. You must NOT participate in the survey more than once.

To be paid, follow these steps:

1. Click on the link below to go to the survey:
2. The link will appear here when you accept this HIT.
3. After completing the survey you will receive a confirmation code in the last page.
4. Enter the code in the box below and we will approve your payment.
5. Please enable Javascript to perform this HIT.

Enter code here: [ ]

For questions and problems, please contact us through Mechanical Turk's contact functionality.

Thank you!

Researchers at Google

Page 1

How do you browse the web?
Thank you for your interest in participating in our survey. Please click "Continue" to start the survey.

Page 2

Which of these websites have you heard of? (check all that apply)

- Diaspora
- Facebook
- FunFactsOfLife
- SnackWorks
- Vimeo
- Wikipedia
- YouTube

Page 3

Imagine that you are visiting www.facebook.com to check friends' latest updates. You see the status update shown below.
[Figure 4(b)]
Which band recorded the song shown in the status update?

- Four Men
- Weezer
- The Beatles
- The Clash

After clicking on the link to watch related videos, you see the page shown below.
[Figure 1]
What would you do?

- Proceed to youtube.com.
- Go back (do not proceed to youtube.com).
- Other. Please specify:

Why? (short-essay)

Page 4

Imagine that you are visiting www.facebook.com to check friends' latest updates. You see the status update shown below.
[Figure 4(c)]
What is the name of the blog shown in the status update?

- Monkeys
- TechCrunch
- The Fast Runner
- Fun Facts Of Life

After clicking on the link to read the full blog post, you see the page shown below.
[Figure 1, but with the blog as the URL]
What would you do?

- Proceed to [blog URL]
- Go back (do not proceed to [blog URL]).
- Other. Please specify:

Why? (short-essay)

Page 5

Which outcome is the most likely if you clicked through the red page to proceed to youtube.com?

- I would be able to watch videos with no malware.
- My computer would be infected by malware.
- Other. Please specify:

Which outcome is the most likely if you clicked through the red page to proceed to [blog URL]?

- I would be able to read the blog post with no malware.
- My computer would be infected by malware.
- Other. Please specify:

Page 6

How would you typically react if you saw a similar red page when trying to visit a website in your day-to-day life?

- I would typically proceed to the website.
- I would typically go back (wouldn't proceed to the website).
- Other. Please specify:

Page 7

Before this survey, had you ever seen a similar red page when trying to visit any website?

- Yes
- No
- I don't remember

If the respondent chooses "Yes", then:
What happened the last time you saw a similar red page when trying to visit a website? (What was the website? What did you do?) (short essay)

Page 8

In the past week, how many times have you checked Facebook?

- I have never heard of Facebook.
- I have heard of Facebook but I do not have a Facebook account.
- Zero times in the past week
- Once in the past week
- Twice in the past week
- Three times or more in the past week

In the past week, how many videos have you watched on YouTube?

- I have never heard of YouTube.
- None in the past week
- 1 video in the past week
- 2 videos in the past week
- 3 or more videos in the past week

What would you do if your wireless router at home were not working?

- I do not know what a wireless router is.
- I would call the provider's technical support to fix it.
- I would call a friend to help me to fix it.
- I would fix it myself.
- Other. Please specify:

What is a computer firewall?

- I do not know what a computer firewall is.
- Software that locates the nearest fire station.
- Software that encrypts personal files.
- Software that controls network traffic to/from a computer.
- Other. Please specify:

What is a public key certificate?

- I do not know what a public key certificate is.
- An electronic document that shows a computer is virus-free.
- An electronic document that shows a website is using 2-factor authentication.
- An electronic document that shows the identity of a website.
- Other. Please specify:

Page 9

What is your gender?

- Male
- Female

What is your age? (free response)
What is your highest completed level of education?

- Professional doctorate (e.g., MD, JD, DDS, DVM, LLB)
- Doctoral degree (e.g., PhD, EdD)
- Masters degree (e.g., MS, MBA, MEng, MA, MEd, MSW)
- Bachelors degree (e.g., BS, BA)
- Associates degree (e.g., AS, AA)
- Some college, no degree
- Technical/Trade school
- Regular high school diploma
- GED or alternative credential
- Some High School
- Other. Please specify:

Which operating systems do you normally use? (check all that apply)

- Windows
- Mac OS
- Linux
- iOS
- Android
- I don't know
- Other. Please specify:

Which web browsers do you normally use? (check all that apply)

- Microsoft Internet Explorer (IE)
- Mozilla Firefox
- Google Chrome
- Apple Safari
- Opera
- I don't know
- Other. Please specify:

Which web browser do you use the most on your personal computer(s)?

- Microsoft Internet Explorer (IE)
- Mozilla Firefox
- Google Chrome
- Apple Safari
- Opera
- I don't know
- Other. Please specify:

Page 10

If you have any additional comments, please write them here. (short essay)

Page 11

Please copy the following code and paste into the text box in the HIT before clicking "Submit".
Check that this is your Amazon worker ID
Submit