An Experience Sampling Study of User Reactions to Browser Warnings in the Field

Robert W. Reeder[1], **Adrienne Porter Felt**[1], **Sunny Consolvo**[1],
Nathan Malkin[2], **Christopher Thompson**[2], **and Serge Egelman**[2,3]
[1]Google, Mountain View, CA
[2]University of California, Berkeley, CA
[3]International Computer Science Institute, Berkeley, CA
{rreeder,felt,sconsolvo}@google.com, {nmalkin,cthompson,egelman}@cs.berkeley.edu

ABSTRACT

Web browser warnings should help protect people from malware, phishing, and network attacks. Adhering to warnings keeps people safer online. Recent improvements in warning design have raised adherence rates, but they could still be higher. And prior work suggests many people still do not understand them. Thus, two challenges remain: increasing both comprehension and adherence rates. To dig deeper into user decision making and comprehension of warnings, we performed an experience sampling study of web browser security warnings, which involved surveying over 6,000 Chrome and Firefox users *in situ* to gather reasons for adhering or not to real warnings. We find these reasons are many and vary with context. Contrary to older prior work, we do not find a single dominant failure in modern warning design—like habituation—that prevents effective decisions. We conclude that further improvements to warnings will require solving a range of smaller contextual misunderstandings.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous; K.6.5 Management of Computing and Information Systems: Security and Protection

Author Keywords

Usable security; browser security; web security; warnings

INTRODUCTION

When someone encounters a browser security warning, they need to make a security-critical decision. Should they adhere to the advice in the warning or proceed to the website despite the risk of an attack? Browsers warn about malware, HTTPS errors, and social engineering attacks. These warnings are a key part of browsers' security strategies; threat detection systems are only useful if end users heed their warnings.

CHI 2018 April 21–26, 2018, Montreal, QC, Canada

ACM ISBN 978-1-4503-5620-6/18/04.
DOI: https://doi.org/10.1145/3173574.3174086

Due to their importance, browser warnings have received a great deal of attention. Initially, browser warnings gained a reputation for low adherence rates [17, 19, 37, 40, 41]. Vendors responded to this research with improvements. In 2013, the Chrome and Firefox teams released data showing that contemporary warnings had relatively high adherence rates—except for the Chrome HTTPS error warning, which still suffered from low adherence (only 30% of users adhered to the warnings) [2]. After further work, Chrome's HTTPS error warning caught up, and now up to 70% of users adhere to the warnings [21, 22, 44]. Despite these improvements, HTTPS warning adherence rates can still be improved.

Past work on warnings has addressed a few classic problems that were presumed to prevent all, or nearly all, users who encountered warnings from adhering to them. These problems included incomprehensible warning text [7, 22, 40, 41], passive warnings that users did not notice [17, 19, 48], and habituation to recurring warnings [4, 5, 8, 10, 28]. These classic problems were initially hypothesized by researchers, then demonstrated in lab studies. But, due to methodological limitations, it was not known for sure what user decision making was like in the wild until telemetry studies came along. Although some researchers had speculated that warnings were a counterproductive [28] or hopeless [31] security measure, Akhawe and Felt [2] showed, in a telemetry study of millions of real user warning decisions, that warnings *can* be effective in practice. Their work came after years of research and vendor improvements, so it would seem that the work on the classic problems has largely been effective. Telemetry methodology thus has yielded some impressive findings [2, 22], but it has its limits; it could generate statistics about *how* users behave in the wild, but not *why*. Thus, several questions remain, including: (1) if the classic problems have largely been solved, why, in some situations, do users still not adhere to or comprehend warnings?; and (2) when users do adhere to warnings, why do they do so?

To address these questions, we used the Experience Sampling Method (ESM) [34] to study how Chrome and Firefox users make decisions about contemporary warnings. ESM gives us elements of the ecological validity of an in-the-wild telemetry study plus the explanatory power of a lab study. We recruited 5,041 Chrome users and 1,251 Firefox users to install and use extensions we developed for this study for several months.

When they encountered warnings in the course of their normal web browsing, our extensions recorded their decisions and prompted them to answer surveys about the events. We thus collected participants' reasons for their warning decisions in the moment, thus minimizing recall bias.

We show that participants took a wide range of factors into account when deciding whether to adhere to a warning. The breadth of our participants' responses illustrates the diversity of contexts in which warnings appear and the numerous factors that people may take into account. Our results both shed light on the current state of some of the classic problems in warning design and reveal new user decision-making factors that have not previously been addressed. In particular, we find:

- Habituation was not a major factor in participants' decision making; in fact, a majority of our participants showed strong evidence of *not* being habituated to contemporary warnings.
- Site reputation was a major factor in many participants' decisions. Over-reliance on site reputation is a misconception that could be corrected through future work on warning designs and educational materials.
- Some participants proceeded through a warning for a trusted site where they had not seen a warning before. In fact, it would be best for users to pay the most attention to warnings on trusted sites that do not usually evoke warnings. So, future work should focus effort on improving this.
- Some participants responded to an HTTPS warning on a site by visiting the site via HTTP; future work could investigate how to better inform users of the risks of downgrading the protocol.
- Some of the reasons participants adhered to warnings were that they trust the warnings themselves; warnings alerted them to typos and misclicks; warnings pushed them toward safer alternatives; and warnings discouraged frivolous tasks.

We conclude that warning designers should address a variety of specialized issues to improve warning adherence and user comprehension. This represents a notable shift in the landscape of warning research. In the early days of browser warning research, designers needed to address low-hanging fruit, like the reading levels of warnings and habituation. Our results suggest that there are no more widespread problems that can be addressed to significantly improve warnings for many people. Instead, browser designers should begin to look at the context and personal experience of warnings if they wish to further improve adherence and comprehension rates.

RELATED WORK

Past work has used various methods to understand how users make decisions when presented with warnings about security-critical events and what can be done to improve the efficacy of warnings.

Reducing Information Overload

Early work on software license agreements by Good et al. [23, 24] sought to reduce the amount of text presented to users for security-relevant decisions. Good et al. tested short summary notices in contrast to long, legalistic license agreements. The summary notices did lead to fewer software installs, though user comprehension remained a challenge.

Passive vs. Active Warnings

For over a decade, one of the most active online threats has been from phishing websites. Phishing attacks can often be prevented if users pay very careful attention to the contents of their URL bar (and even this is not always sufficient). However, Dhamija et al. found that users generally paid attention only to the look and feel of the website, ignoring cues from the browser itself [17].

There exist third-party browser toolbars that provide more explicit indicators of a page's safety or trustworthiness [13]. Yet, Wu et al. found that, even when asked to pay attention to such toolbars, participants failed to look at them and ended up falling for phishing attacks [48].

One problem with these toolbars was that the indicators were *passive*: they displayed their warnings to the user without interrupting their flow. When developers started adopting these features directly into browsers, they too left some warnings as passive. However, research showed that such warnings were considerably less effective. In a between-subjects lab experiment with 60 participants, Egelman et al. compared active and passive warnings used by Internet Explorer 7 for phishing pages [19]. They found that only 13% followed the advice of a passive warning, compared with 79% for the active warning. Since then, most browsers have adopted active, blocking warnings as the default.

Habituation

However, an active warning does not guarantee that a user will actively engage with it. Because of the frequency with which people encounter warnings that turn out to be false alarms, many have become habituated to them, learning to automatically ignore them, rather than assessing the situation. Böhme observed a similar habituation to license agreements by testing various designs for consent dialogs with 80,000 users of an online privacy tool [8]. Anderson et al. studied the mechanism by which this occurs by showing people warnings while scanning their brains using functional magnetic resonance imaging (fMRI) [4]. They found that, with repeated exposure to warnings, neural activity decreases and concluded that this is a consequence of the brain's biology, rather than "due to users' laziness or carelessness."

A number of researchers have focused on reducing habituation to the warnings. Brustoloni et al. designed polymorphic dialogs, which forced users to pay attention by continuously changing the required user input [10]. Anderson et al. varied size, color, highlighting, and option order, among other attributes [5]. Measuring their participants using fMRI and mouse cursor tracking, they found that these changes slowed the habituation process in users.

HTTPS Warnings

Browsers warn users when they try to connect to a site over HTTPS, but the certificate is self-signed, expired, or otherwise invalid. Users encounter such errors significantly more often—by an order of magnitude [2]—than phishing warnings.

Compared to a phishing warning, a user's decision when faced with an HTTPS warning is much more complicated. When a warning is a true positive, the warning indicates to the user that

they are experiencing a man-in-the-middle attack. Yet, this is probably much less likely than the relatively benign alternative: that the site's certificate—or, more frequently, the client's computer [1]—is misconfigured. As a result, users at one point had been ignoring such warnings in large numbers. Herley argued that such behavior is rational [28]. The first evidence of users ignoring HTTPS warnings in the laboratory was obtained by Dhamija et al., whose participants tried to identify phishing websites; in the process, 68% of their subjects clicked through an HTTPS warning they encountered, without reading it [17].

These results hold across browsers: while Dhamija et al.'s experiment used an early version of Mozilla Firefox, Schechter et al. used Internet Explorer 7 when they conducted an experiment also focused on phishing [37]. Here as well, participants encountered an HTTPS warning and many clicked through, including 36% of participants in a condition where they were using their own, real credentials to log into a banking website.

Sunshine et al. also tested HTTPS warnings using Internet Explorer 7, as well as Firefox 2 and 3 [41]. (The experiment was later replicated by Sotirakopoulos et al. with Internet Explorer 7 and Firefox 3.5 [40].) Clickthrough rates were high, but Sunshine et al.'s study also demonstrated that the design of the warning could play a major role. While 90% of Internet Explorer 7 and Firefox 2 users clicked through the warnings, the corresponding number for Firefox 3 was only 55%. Between the two versions, the warning had changed; the new design deemphasized the option to proceed and made it more complicated, requiring more clicks.

Further Design Improvements

Since the initial warnings, researchers and browser makers have iterated on designs in order to reduce the rates at which users click through the warnings. For example, Sunshine et al., in the study above, also tested a redesigned dialog that estimated the risk level by soliciting the website type (financial, e-commerce, or other) from the user and displayed a succinct warning with urgent coloring based on the category [41].

Biddle et al. focused on increasing users' understanding of the decision they were facing [7]. Egelman and Schechter showed that changes to the look and feel of the Internet Explorer phishing warning resulted in more people noticing it [20].

Felt et al. experimented with a variety of warning designs, in order to explain differences in warning compliance between Google Chrome and Mozilla Firefox [22]. In addition to the warning's design, Weinberger et al. found that the browser's warning storage policy—how long the browser remembers a user's choice to proceed through a warning on a particular website—explains the bulk of remaining differences between the browsers [44]. In a separate study, Felt et al. redesigned Chrome's HTTPS warning to enhance user comprehension and improve adherence [21]. Adherence increased because of the opinionated design, but the authors concluded that they "ultimately failed at [their] goal of a well-understood warning."

Varying Methods

Past studies on browser security warnings have typically either involved laboratory experiments or surveys structured around contrived scenarios (e.g., [9, 19, 20, 21, 37, 41]) or the use of telemetry data collected from users *in situ* (e.g., [2, 21, 22]). The former allows researchers to collect deep insights into user behavior in controlled environments, but has limited ecological validity, whereas the latter is ecologically valid, but lacks explanatory power. Our goal was to bridge this methodological gap.

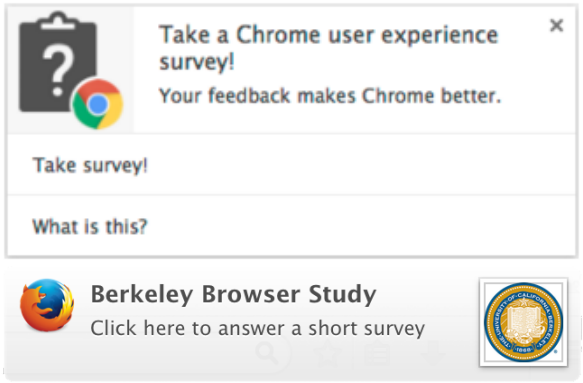

Figure 1. Survey prompt shown when the Chrome (above) and Firefox (below) extensions detected the participant had responded to a warning.

The Experience Sampling Method is a method for collecting *in situ* explanatory data by prompting participants to complete short surveys immediately after performing behaviors of interest [34] and has been used by various researchers in usable privacy and security [15, 16, 27, 30, 39]. For example, researchers have used ESM prompts to study participants' privacy concerns surrounding hypothetical or emergent ubiquitous computing devices [12, 25, 29]. Carrascal et al. prompted study participants about the value that they placed on personal information that they had recently transmitted to various websites [11]. Others have used ESM prompts to gather participants' reactions to different types of location-sharing requests [6, 16, 35], as well as to examine participants' willingness to share information with smartphone apps [14, 38, 46], and users' decisions to use various security mechanisms on their mobile devices [26].

In this work, we used ESM to study participants' reactions to web browser security warnings. In particular, we used it to gauge the effectiveness of current warning interfaces that were naturally encountered during participants' day-to-day web browsing behavior, thereby bridging the current gap in the security warning literature between laboratory and survey studies structured around contrived scenarios and telemetry data collected in the field.

METHODOLOGY

We surveyed participants responding to warnings in the wild and in the moment. We recruited two samples: one of Chrome users and another of Firefox users. We developed an extension for each browser that would detect when the browser had shown a security warning and would prompt the user to take a survey immediately after the user had responded to the warning (i.e., either by adhering to it and navigating away from the suspicious website, or proceeding to the website against the warning's recommendation). The survey notifications appeared as "toast" notifications on the desktop (Figure 1) so as

to be unobtrusive. Besides choice of browser, our two samples also varied in participant recruitment and compensation. Otherwise, we treated the samples as similarly as possible, including showing similar survey content to both the Chrome and Firefox participants. For both samples, we collected data for the 91-day period (approximately 3 months) from May 18, 2015 through August 16, 2015. Participants in both samples installed their respective browser extensions and ran the extension for at least this 91-day period. We disabled and uninstalled all browser extensions by September 10, 2015.

Procedure

Upon installing one of our extensions, we asked participants to consent to running the extension and receiving prompts to take surveys from time to time. After consenting, participants filled out a demographic survey. The extension then ran silently in participants' browsers, waiting to detect warning events.

Survey Prompts

When our extensions detected that a warning had been shown and the participant had reacted to it, they sometimes showed a system notification (Figure 1) to the participant prompting them to take a survey. Survey prompts were not always shown because we limited them to 2 per day and 4 per week per participant on Chrome and once per week on Firefox,[1] to avoid annoying participants by bombarding them with survey requests. Participants could dismiss or ignore prompts, so they were not required to complete a survey on every notification.

Warning Conditions

We detected 3 warnings: SSL/TLS,[2] malware, and phishing warnings. Malware and phishing warnings appear for sites on the Google Safe Browsing list of sites hosting malware and phishing. For each warning, a participant may have proceeded (by choosing an option on the warning to proceed) to a site despite the warning, or they may have not proceeded (by choosing a "Back to safety" option or by closing the tab/window). The 3 warning types times 2 possible responses leads to 6 conditions for which users may have answered our surveys, which we describe throughout as follows:

- *ssl-proceed*: The participant was shown an SSL warning, but they chose to proceed to the website anyway.
- *ssl-noproceed*: The participant was shown an SSL warning, and chose to adhere to it (by not proceeding to the website).
- *malware-proceed*: The participant was shown a malware warning, but they chose to proceed to the website anyway.
- *malware-noproceed*: The participant was shown a malware warning, and chose to adhere to it (by not proceeding).
- *phishing-proceed*: The participant was shown a phishing warning, but they chose to proceed to the website anyway.
- *phishing-noproceed*: The participant was shown a phishing warning, and chose to adhere to it (by not proceeding).

Survey Content

If the participant clicked on a survey prompt, the survey appeared in a new tab in the browser. Except for cosmetic differences, surveys were almost identical[3] in question content and ordering for the two browser samples. Response order for multiple-choice questions was randomized (or randomly reversed, for ordered-scale questions). Each survey displayed a screenshot of the warning the participant had just seen, though with blurred text. Our objective in showing the blurred screenshot was to remind the participant of which warning the survey was asking them about, but not to provide so much detail that they re-thought their decision-making process by re-reading the warning text. The screenshot was fixed in position at the top of the survey, while the questions in the survey were in a scrollpane below. Figure 2 shows an example of an *ssl-noproceed* condition survey.

Surveys were similar for all 6 conditions, with a few context-appropriate variants per condition. We asked 5 survey questions related to warning comprehension and reasons for the decision to proceed or not. Surveys also asked questions that collected more specific context (such as URL) on each decision, but due to limited space, we do not present results of those questions here. The questions on which we report are:

- *Choice* question: This question's wording was different for *proceed* versus *no-proceed* conditions:
  - *Proceed:* You chose "Proceed to <website>" instead of "Back to safety." How did you choose between the two options? *[Open-ended response]*
  - *No-proceed:* You chose the "Back to safety" option, or you closed the page. Why did you choose not to proceed to <website>? *[Open-ended response]* (Figure 2)
- *Visited-site* question: Have you visited <website> before? *[Multiple choice: Yes/No/I'm not sure]*
- *Seen-warning* question: Have you seen a page like the one shown above when trying to visit <website> before? *[Multiple choice: Yes/No/I'm not sure]*
- *Account* question: Do you have an account on <website>? *[Multiple choice: Yes/No/I'm not sure/I prefer not to answer]*
- *Trust-in-site* question: How much do you trust <website>? *[Multiple choice: Strongly distrust/Somewhat distrust/Neither trust nor distrust/Somewhat trust/Strongly trust/I don't know]*

Note that the survey wording referred to a warning as a "page"; we used "page" as a neutral term to avoid the social acceptability bias that might have been introduced had we used "warning" (since participants might assume that adhering to warnings is a socially desirable behavior). Note also that the placeholder <website> indicates places where questions used the actual host and domain name of the URL the participant was trying to visit. However, we only recorded the URL if the participant explicitly consented for us to do so in one of the survey questions not reported here. For Firefox participants, if they elected to not share the URL with us, we saved a hashed copy, so that we could still examine whether participants visited the same URLs multiple times during the course of the study.

[1] The Firefox rate limit was never encountered during the study.

[2] We henceforth use "SSL" for consistency with older literature.

[3] Due to a technical error, two questions were omitted from the Firefox sample: whether they had seen the warning before, and whether they had seen it before on the particular website they were trying to visit.

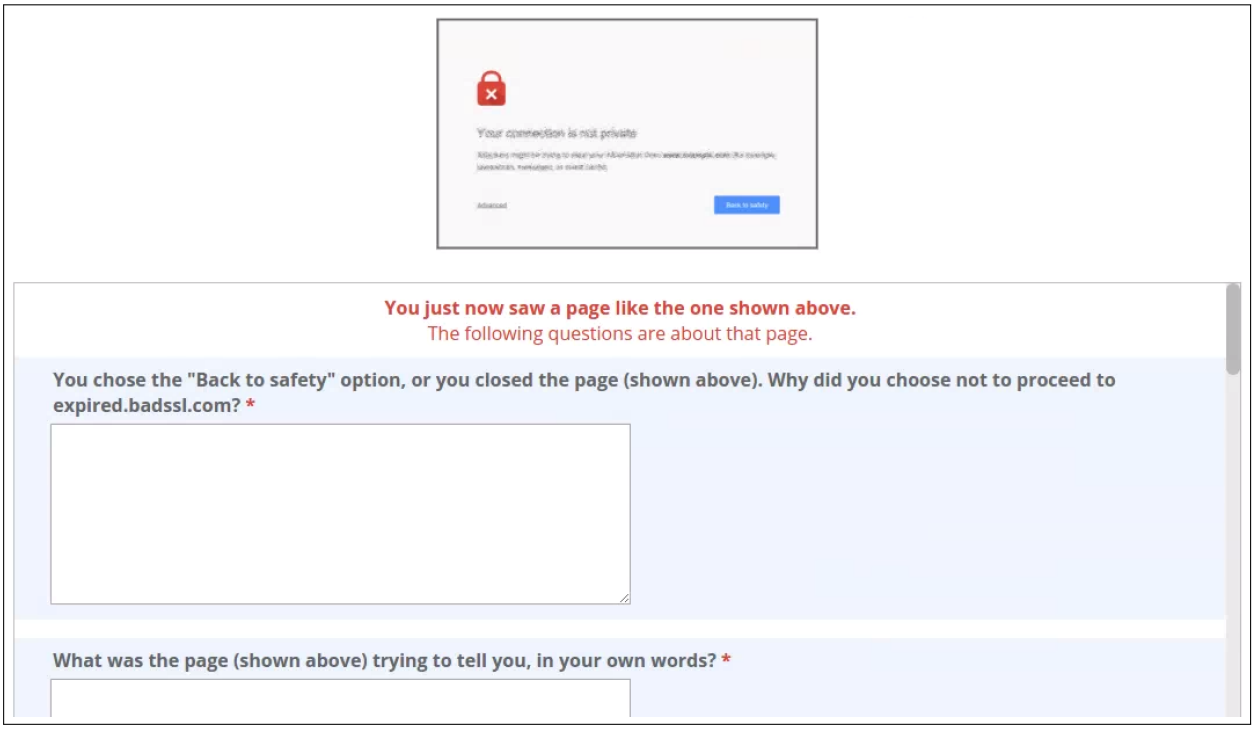

Figure 2. Example of survey shown after a participant responded to an SSL warning and clicked on the survey prompt. A blurred version of the warning was shown above the survey, in order to clarify the survey's questions, without priming participants to specific warning text.

Recruitment, Compensation, and Demographics

Recruitment and compensation presented considerable challenges. The incidence of SSL, malware, and phishing warnings is very low (3.4% of Chrome users during one week for SSL, 0.4% for malware, and 0.2% for phishing), and in addition, we anticipated a low survey response rate. To account for both infrequent survey notifications and low completion rates, we estimated needing a sample of several thousand participants running our extension over our 91-day study period to get a reasonable number of survey responses. After dismissing several recruiting methods as infeasible, we decided on using a press release to recruit for the Chrome sample and Amazon Mechanical Turk for the Firefox sample. Both recruiting methods introduced bias, but different kinds of bias. Thus, the two different samples provide different points in the space of the general browser-using population, and therefore we do not use inferential statistics to directly compare the two samples.

Compensation also introduced challenges. Paying participants a fixed rate might incentivize some to install our extensions but not answer any surveys. On the other hand, paying per survey completed could incentivize gaming, in which participants might try to unnaturally induce and respond to security warnings, which could spoil the ecological validity of our experience sampling. Furthermore, any compensation scheme would require keeping track of participants' identities, and we preferred to offer our participants anonymity, since they were providing potentially sensitive data. As with recruitment, we used two different compensation schemes for our two samples. Chrome participants were not compensated; they were volunteers and were allowed to remain anonymous. Firefox participants were paid $0.50 for installing the extension and an additional $3 for keeping the extension installed for the entire study period. (Using Mechanical Turk for the Firefox recruitment allowed us to separate study data from payment information, so as to keep participants anonymous.)

Chrome Sample Demographics

The Chrome press release announced our Chrome extension as a way for users to provide in-the-moment feedback on their Chrome experience. While our recruiting method provided the potential to reach a large, broad sample of Chrome users, it may have provided a biased sample, since we could not control who ran our press release. Demographic survey responses indicated the most common places where respondents heard about our study were the Chrome Web Store itself (where the extension was published for download), TechCrunch, omgchrome.com, and Reddit. Since these sites cater to technology enthusiasts, it is likely our audience was heavily composed of tech-savvy users. Furthermore, since we provided no compensation, it is likely participants were motivated by a desire to help Chrome and/or by technical curiosity. Since the sample is not representative of the general browser-using population, we must use caution in generalizing our results or comparing them with the Firefox sample. Nevertheless, our sample of 5,041 participants is large enough to represent an important subset of the general browser-using population; the tech-savvy are users too. Moreover, the reasons we found for proceeding or not proceeding through warnings likely do exist throughout the general population, just in different proportions than in our sample (our sample is, after all, a part of the general population). Finally, our results on reasoning

	Chrome sample		Firefox sample	
	Resp. (n=508)	Installs (n=5,041)	Resp. (n=136)	Installs (n=1,251)
Male	92.6%	81.0%	58.8%	56.8%
Female	5.5%	14.2%	38.2%	39.8%
Other or not specified	1.9%	4.8%	2.9%	3.4%
Age 18-24	26.4%	25.8%	17.6%	24.0%
Age 25-34	43.1%	33.9%	40.4%	44.4%
Age 35-44	17.2%	20.0%	19.9%	16.0%
Age 45-54	7.9%	10.1%	14.7%	10.0%
Age 55-64	3.7%	6.3%	6.6%	4.2%
Age 65 or over	1.7%	3.8%	0.7%	1.4%
Some High School	3.7%	7.0%	2.9%	1.4%
HS or equiv	43.1%	45.9%	32.4%	38.8%
College degree	29.9%	24.1%	39.7%	45.4%
Graduate degree	20.9%	16.6%	23.5%	12.5%
Prefer not to answer	2.4%	6.4%	1.5%	1.8%
US	42.1%	27.8%	84.6%	80.1%
India	2.4%	3.4%	8.1%	11.1%
France	8.5%	6.8%	—	—
UK	7.0%	4.0%	—	—
Germany	6.7%	3.7%	—	—
Canada	3.5%	3.6%	—	—
Other	29.8%	51.0%	7.3%	8.8%

Table 1. Chrome and Firefox sample demographics. The Chrome sample includes the 508 respondents who completed 637 surveys and the superset of 5,041 participants who installed our Chrome extension. The Firefox sample includes 136 respondents who completed surveys and 1,251 participants who installed our Firefox extension.

for warning decisions can be seen as technically sophisticated. Our sample's misconceptions likely occur at even greater rates in the general population.

Besides likely bias toward a tech-savvy population, the demographics of our population, shown in Table 1, also suggest some bias relative to the general population. Compared with the representative sample of US Internet users collected by Wash and Rader [43], for example, our sample skews young, more educated, and heavily male. Our sample is not limited to the US, so our population is not the same as theirs, but we still note that our sample skews especially young and male.

Firefox Sample Demographics

Since the half of our team responsible for implementing the Firefox portion of the study did not have the resources to issue a widely-circulated press release, we experimented with paid advertisements on Facebook and Twitter. We also commented on the existing Reddit thread regarding the Chrome recruitment effort, inviting Firefox users to participate in the Firefox version of the study. These efforts resulted in many page views for our study recruitment page, but ultimately yielded few installs of our extension. As a result, we decided to pay participants on Mechanical Turk, which was how we recruited the majority of our Firefox participants (86%). Like the Chrome recruitment efforts, we advertised the Firefox study as being about improving web browser usability, and made no mention of the security focus, so as to avoid priming.

We requested demographic information from everyone who installed the browser extension, even if they never saw a warning from their browser (i.e., the demographic survey was displayed immediately upon a successful installation of the extension). This resulted in 1,251 completed demographics surveys (Table 1).[4] According to responses to demographic questions in our survey, 59% of our participants had some higher education with 49% of them holding bachelors degrees or higher. Participants held a wide range of occupations including teachers, managers, and engineers; however, the plurality of our participants (12%) were students.

Data Analysis

Qualitative responses were coded using a general inductive approach [42]. Two of the authors read through Chrome responses for all 6 warning conditions and worked together to form initial codebooks for each condition. Two other authors did the same for Firefox responses and a common codebook for each warning condition was agreed upon. We then used three raters for the Chrome sample and two for the Firefox sample to code responses. They coded in 3 rounds, updating the codebook between rounds. Every response was coded by at least 2 raters and some Chrome responses by 3 raters. Codes for the third round were considered final. While raters could assign multiple codes per response, we required at least one code per response. One of the authors served as a tie-breaking rater to ensure all responses were assigned at least one code.

Inter-rater reliability (IRR) was computed using the Kupper-Hafner statistic [32]. For the *Choice* question, we computed IRR for each pair of raters. After eliminating *malware-proceed* and *phishing-proceed* conditions from IRR analysis due to their small sample sizes, we had 12 IRR scores for Chrome and 4 for Firefox, all of which were at or around 0.6 (and 11 of which were above 0.7), so would be considered "substantial" agreement [33].

Limitations

Our large-sample experience sampling methodology, while providing advantages in ecological validity and coverage, also has some limitations. We noted that our samples are skewed compared to the general population. The Chrome sample in particular skews tech-savvy, young, and male, and the skew toward more young and male participants continues between installers and those who actually answered surveys. The Firefox sample had a better gender balance, but also skewed towards more educated and younger than the population as a whole.

Most of our findings come from answers to the *Choice* question. This question was always first on the survey, by design — we wanted to capture respondent's unvarnished, initial feelings of the reasons behind their decisions. However, since the question was free-form, it has the drawback that recall is never complete. Thus, some reasons may go under-reported compared to their actual frequency in the sampled population.

Our response rates were low, which may indicate the possibility of a skew in the warning events for which we received survey responses. The rarity of malware and phishing warnings, combined with our low response rates, led to a very small sample of *malware-proceed* and *phishing-proceed* survey responses (n=4 for *malware-proceed* in Chrome and n=2 in Firefox, and n=3 for *phishing-proceed* in Chrome and n=0

[4] Since the demographic survey took place at the end of the installation, we refer to this number as the total number of *complete* installs.

in Firefox). We still present those results for completeness, but they should be considered anecdotal.

Ethics

Our Firefox experimental protocol received approval from an Institutional Review Board (IRB). Our Chrome experiment was conducted within an organization that does not have an IRB, so the study was not subject to IRB review; however, researchers who have received human subjects training reviewed the study protocol and survey instrument prior to the experiment. Participants provided informed consent to having the extension run in their browser, collect data, and prompt them to take surveys during the study. In both samples, a pseudonymous identifier was used so that study data was not personally identifiable; in the Chrome sample, since no compensation was provided, participants were not required or asked to identify themselves at all (in the Firefox sample, Mechanical Turk worker IDs were collected for payment purposes, though they were not linked to study data). Raw survey data access was restricted to members of the research team.

RESULTS

Our sources of data come from warning events automatically collected by the extensions and from the event-triggered survey data participants provided when they responded to a survey prompt. We start by presenting basic data on warning exposures and user decisions on those warnings (adherence rates), we present response rates to our survey prompts, and then we present actual survey data.

Adherence Rates

Adherence rates are the rates at which users do not proceed through a warning, i.e., the rate at which they choose the safer option [2]. Tables 2 and 3 show the number of warnings shown to our participants during the study period in each of our 6 warning conditions. We compute adherence rates by dividing the *ssl-noproceed*, *malware-noproceed*, and *phishing-noproceed* numbers by the total number of SSL, malware, and phishing warnings displayed.

Our 5,041 Chrome participants adhered to 37.6% of SSL warnings, 76.9% of malware warnings, and 79.9% of phishing warnings. They adhered at lower rates than the general Chrome population at the time. During that same time period in 2015, Chrome users adhered to 50% of SSL warnings, 87% of malware warnings, and 96% of phishing warnings. The adherence rates were slightly higher among our 1,251 Firefox participants: 52.3% for SSL, 89.3% for malware, and 96.4% for phishing.

Survey Response Rates

Tables 2 and 3 show, for each warning condition, the number of instances observed, number of survey prompts shown to participants, number of survey responses, and response rate (responses divided by prompts). Our response rates, all under 10% for Chrome and 10-20% for Firefox (except for the *phishing-noproceed* condition, in which neither of the two warning exposures resulted in a survey response), were low compared with rates reported in several other experience sampling studies [15]. We suspect that the lack of compensation for Chrome users and low rate for Firefox users were factors. While low response rates can indicate bias in the sample of events for which responses were received, we note that our responses are distributed roughly proportionally across the 6 warning conditions. We note further that our responses were not dominated by any particular respondent; for example, 312 unique Chrome respondents provided the 373 *ssl-proceed* responses and 204 unique Chrome respondents provided the 229 *ssl-noproceed* responses. We collected, across all warning conditions, 637 responses from Chrome respondents and 216 from Firefox respondents. The mean, minimum, and maximum number of responses per respondent were 1.25, 1, and 5 for Chrome and 1.65, 1, and 12 for Firefox.

Condition	Warnings	Prompts	Resp.	Rate
SSL-proceed	7688	5072	373	7.4%
SSL-noproceed	4638	2342	229	9.8%
malware-proceed	214	78	4	5.1%
malware-noproceed	712	268	16	6.0%
phishing-proceed	152	67	3	4.5%
phishing-noproceed	604	195	12	6.2%

Table 2. Chrome: response rate data for each condition

Condition	Warnings	Prompts	Resp.	Rate
SSL-proceed	572	572	96	16.8%
SSL-noproceed	627	627	94	15.0%
malware-proceed	14	14	2	14.3%
malware-noproceed	117	117	17	14.5%
phishing-proceed	2	2	0	0%
phishing-noproceed	53	53	7	13.2%

Table 3. Firefox: response rate data for each condition

Experience Sampling Survey Results

We cover results from our experience sampling surveys in two main parts. First, we present responses to the free-form *Choice* question; these results give us a broad overview of respondents' many reasons for their decisions to proceed through warnings or not. Second, we present results from responses to fixed-response questions that explored specific decision factors that have been suggested by past work, particularly habituation and site reputation.

Reasons for Proceeding or Not Proceeding

Responses to the *Choice* question show a range of reasons for proceeding or not proceeding through warnings. Table 4 shows reasons given by 2% or more of respondents in each of the 6 warning conditions for the *Choice* question. Each reason corresponds to a code from our codebooks. In the text below, we place these codes in italics.

Reasons for proceeding

For all warnings and browsers, the top reasons for proceeding related to site reputation. In *ssl-proceed* responses, participants reported that they were *familiar with the site* and therefore trusted it; this accounted for half of the Firefox responses. In Chrome, more than a third noted that they were seeing the warning on a site that was *internal or their own*. For *malware-proceed* and *phishing-proceed* warnings, the top reason was another variant on site reputation, *knows the site is safe* (though most respondents did not specify why they believed this).

Beyond site reputation, common reasons offered for proceeding through warnings were variants on confidence in the respondent's own judgment or knowledge of the situation; for

Reason	Chrome (%)	FF (%)	Representative quotes
ssl-proceed	*(n=373)*	*(n=96)*	
Familiar with the site	34.3	50.0	"Because it is a site I visit frequently and trust it."
Site was internal or their own	33.2	2.1	"It is a web page on company Intranet"
Understands the risks	10.7	9.4	"I know what I'm doing, and I know the consequences of it."
Wanted to get something done	7.2	26.0	"I need to perform the task this site is intended for."
Expected broken connection	6.4	2.1	"I was expecting to receive this error"
Literal actions to proceed	6.4	5.2	"I clicked show advanced settings and then proceed."
Won't enter personal info	4.8	2.1	"I was not planning to enter any information that could be taken and did not feel the need for a secure site"
Other	4.0	8.3	"I'm on a ship at sea and everyone is getting this message."
ssl-noproceed	*(n=229)*	*(n=94)*	
Safety or security, in general	16.6	20.4	"I don't want to risk my safety."
Trusted the warning/browser	12.2	2.2	"I abide my Google warnings!"
Unintended site	10.5	4.3	"Misclicked the link in the first place."
Reverted to HTTP	8.3	2.2	"Actually, I removed the HTTPS."
Task wasn't important	7.4	16.1	"I thought 'never mind', it was work-related and didn't need to be there. Just curious"
Didn't trust the site	6.1	16.1	"I was not sure of the identity of the website."
Connection: insecure connection	3.9	1.1	"I got a warning my connection wasn't secure"
Did proceed to the site	3.9	11.8	"I actually did, but I did it incognito."
Used an alternative site	3.9	7.5	"I decided to play it safe and go somewhere else."
Didn't trust the referrer	2.2	1.1	"I was not confident in the sending page, and would just use search to find the page if I wanted to go back."
Led by the UI	2.2	6.4	"because the button was the biggest"
Problem on a trusted site	2.2	3.2	"I have visited this website before without seeing this error"
malware-proceed	*(n=4)*	*(n=2)*	
Knows the site is safe	75.0	0	"I know this website as a legitimate one, plus I'm a tech-savvy user (and a developer)."
Is an expert user	50.0	0	"I'm a Google Top Contributor and I was visiting a site a user created and asked us info about this problem."
Other	0	100	"I just needed the files on the next page."
malware-noproceed	*(n=16)*	*(n=17)*	
Trusted the warning/browser	31.2	0	"because there was a scary red message with an "X" and Stop sign "
Don't want to get malware	25.0	23.5	"Because I'm afraid this site has malware that may damage my PC."
Unintended site	18.8	5.9	"I mistyped the URL"
Safety or security, in general	12.5	41.2	"I value my safety"
Don't trust the site	6.2	17.6	"I don't know the site I was trying to visit (it was part of a search result)"
phishing-proceed	*(n=3)*	*(n=0)*	
Knows the risk	100	N/A	"I knew the risks and decided to go ahead and see what would happen."
Needed something on site	33.3	N/A	"I decided I needed the file, and am confident that I will not click on fishing links..."
Will take precautions	33.3	N/A	"...will scan any files I download with antivirus."
phishing-noproceed	*(n=12)*	*(n=7)*	
Site is unsafe	41.7	14.3	"Warning about phishing alerted me it was a dangerous site."
Didn't want to take the risk	16.7	28.6	"I have no idea if this site is valid or safe. Therefore, I just closed it as soon as possible"
Trusted the warning/browser	16.7	14.3	"I was warned not to proceed further"

Table 4. Reasons respondents gave, in response to the *Choice* question, for their decision to proceed or not proceed through warnings.

SSL warnings, these reasons were coded as *understands the risks* and *expected broken connection* and for malware warnings, these participants described themselves as *an expert user*.

Other frequent reasons for proceeding through warnings included wanting to complete a task despite the risks (*wanted to get something done* for SSL warnings and *needed something on site* for phishing warnings), and proceeding with a plan to reduce risk (*won't enter personal info* for SSL and *will take precautions* for phishing). Table 4 also shows there were some responses of *literal actions to proceed*; these were apparently from respondents who misinterpreted the *Choice* question to be about what UI steps they took to proceed.

Reasons for not proceeding

Reasons for not proceeding through warnings were even more varied than reasons for proceeding. The top reasons for not proceeding were related to preserving security or safety. These were *safety or security, in general* for SSL and malware warnings, and *didn't want to get malware* for malware and *site is unsafe* and *didn't want to take the risk* for phishing warnings.

Among our Chrome participants, another common reason was trust in the warning itself, or its source—the browser or browser manufacturer (*trusted the warning/browser*). Also common, for SSL and malware warnings, was the reason that the respondent had made an error, such as a typo or misclick, and thus that it was an *unintended site* that triggered the warning. The remaining reasons for not proceeding were observed only for SSL warnings, where we had many more responses than for malware or phishing warnings.

We were surprised to find that 8.3% of Chrome and 2.2% of Firefox *ssl-noproceed* respondents reported that they had *reverted to HTTP*, i.e., they proceeded to a site using HTTP after not proceeding through the SSL warning. Similarly, 11.8% of Firefox and 3.9% of Chrome respondents also indicated that they *did proceed to the site* after not proceeding through the warning, but didn't explicitly mention downgrading the protocol to HTTP.

One other reason for not proceeding, *task wasn't important*, was the flip side of the *wanted to get something done* reason in the *ssl-proceed* condition. It seems participants who provided these answers were weighing the risk of proceeding through a warning against the importance of their primary task. Instead, 7.5% of Firefox and 3.9% of Chrome participants responded that they *used an alternative site*, so were able to avoid proceeding through an SSL warning but still complete their primary task.

A small but noteworthy fraction of users of both browsers explicitly mentioned the warning's user interface design (i.e., they were *led by the UI*); for example one element was dominant, or because they could not figure out how to proceed.

	Proceed	No-proceed	ρ or ϕ	p-value
Chrome				
Account	66%	30%	0.35	<0.001
Visited-site	79%	47%	0.33	<0.001
Seen-warning	81%	43%	0.37	<0.001
Trust-in-site	5	4	0.87	<0.001
Firefox				
Account	12.9%	9.4%	0.08	0.29
Trust-in-site	4	3	0.34	<0.001

Table 5. Results for fixed-response survey questions for *ssl-proceed* and *ssl-noproceed* responses. For *Account*, *Visited-site*, and *Seen-warning* questions, the first two columns show percentage of "Yes" answers. For the *Trust-in-site* question, those columns show median value on a 1-5 scale from 1="Strongly distrust" to 5="Strongly trust". Correlations are computed using the ϕ coefficient for binary questions and Spearman's ρ for the *Trust-in-site* question.

Finally, 2–3% of participants mentioned a *problem on a trusted site* as a reason not to proceed. This response is the flip side of the *familiar with the site* reason to proceed. Some participants indicated trusting a site as a reason to proceed through a warning, but, in fact, if a trusted site does not normally trigger warnings, a warning is more likely a sign of real danger, such as a man-in-the-middle attack or a site compromise.

Specific Decision Factors

Table 5 shows results from the *ssl-proceed* and *ssl-noproceed* conditions for the *Account*, *Visited-site*, *Seen-warning*, and *Trust-in-site* questions (we only analyzed these questions for SSL due to the low number of *malware-proceed* and *phishing-proceed* survey responses). We computed the ϕ correlation between the decision to proceed or not and a "Yes" answer to the *Account*, *Visited-site*, and *Seen-warning* questions, and Spearman's ρ between the decision to proceed and answers to the *Trust-in-site* question. We found significant effects for Chrome participants in all cases, but only found significance for Firefox users with regard to the effect of trusting websites on decisions to proceed past warnings.

DISCUSSION

Our overarching observation from our results is that there were many factors at play in participants' decisions to proceed or not proceed through warnings. As our observed adherence rates show, it is not the case that all participants behaved the same when confronted with warnings, and as our qualitative data on users' reasons for their decisions shows, they have varied and diffuse reasons for their decisions. We thus confirm Akhawe and Felt's conclusion that modern browser warnings *can* be effective and that stereotypes of the "oblivious user" are severely misleading [2].

The variety of reasons we found for participants proceeding through warnings also suggests that the low-hanging fruit in browser warning design has largely been addressed. In the early days of browser warning design, when adherence rates as low as 10% were observed [19, 41], there were clear reasons why warnings failed, such as the use of passive warnings that failed to get users' attention [19] or warnings that had false alarm rates near 100% [28]. Our data do not show evidence of such issues remaining; instead, improving adherence rates may require addressing numerous smaller, more contextual issues. For example, users seem to consider decision factors like importance of their primary task and whether alternative sites with similar content are available; warnings could perhaps tip some decisions toward adherence by nudging users away from trivial tasks or pointing them toward alternative sites.

We continue our discussion by first addressing key decision factors identified in prior work: habituation, comprehension, and trust-in-site. We then discuss new decision factors identified in our data that future work should focus on.

Habituation

One of the prevailing theories about why people ignore warnings has been habituation. For example, Krol et al. concluded that users warned about potentially harmful PDF files "mainly ignored the warnings because they have been desensitised by the high number of security warnings" [31]. False positives do remain an issue, though only for SSL warnings: the false positive rate for malware and phishing warnings is minuscule [36]. However, the data we have collected does not support habituation as a major factor in modern browser warnings. The breadth of thoughtful reasons for warning decisions we collected argues against widespread habituation.

We performed an additional analysis on our data to test whether habituation was a major decision factor. If it were, we would expect that most users behave consistently, i.e., that they always proceed or never proceed through warnings. We looked at our warning event data—data from all participants, not just survey respondents—to see what percentage behaved consistently for the warnings seen during our study period. Of participants who saw more than one warning during the study period, 64.1% of Firefox participants and 50.6% of Chrome participants acted inconsistently, i.e., they proceeded at least once and did not proceed at least once on different websites. We thus observe that at least a majority (and a large majority for Firefox) showed strong evidence of *not* being habituated to the Firefox and Chrome SSL, malware, and phishing warnings. We further note that these are only lower bounds on the percentages of non-habituated participants; a participant who saw two warnings and proceeded or did not proceed both times may still be considering their decisions carefully.

Making inconsistent decisions suggests that, rather than acting out of habit, participants made decisions based on the specific circumstances and the context of each warning.

Our *Seen-warning* question asked Chrome respondents if they had seen a warning before on the page they were about to visit. As shown in Table 5, the decision to proceed correlates significantly with having previously seen a warning on that page. This may suggest a certain kind of site-specific habituation (possibly a rational one, for respondents who knew their sites were misconfigured), for instance, 33% of Chrome respondents mentioned seeing the SSL warnings on either their own or other internal websites; but we found strong evidence against an over-arching habituation to all warnings.

Comprehension

Unfortunately, just because people make different choices when faced with the warning, we are not guaranteed that these are well-informed decisions. And fostering comprehension is an important part of, and perhaps the next big challenge in,

warning design [20, 21]. While the detailed reasons in our results suggest widespread basic comprehension of warnings (insofar as they represent impending security threats), our data do suggest two potential areas for comprehension improvement. First, many respondents cited their trust in a site as a reason to proceed despite an SSL warning. However, SSL warnings are not warnings about the site *per se*. They are warnings about the connection to a site or the the identity of the site being connected to. Participants who proceeded through an SSL warning because they trusted the site they were trying to connect to likely misunderstand the warning and may have been making a mistake. There may be an opportunity for SSL warnings and browsers more broadly to better educate users about the security properties of SSL.

Second, participants who reverted to HTTP to circumvent SSL warnings may not have known the advantages of HTTPS over HTTP and/or may not appreciate some of the risks of connecting to a site via HTTP (e.g., "supercookies" [47]). In most cases, proceeding past an HTTPS warning to a misconfigured site still offers greater protection than simply avoiding HTTPS altogether by visiting the website's HTTP counterpart.

Site Reputation, History, and Trust

Reasons related to site reputation or a respondent's familiarity with a site were the most-mentioned reasons for proceeding through warnings in responses to the *Choice* question. This result confirms prior work showing that site reputation is an important factor in users' warning decisions [3, 18]. Answers to the *Account*, *Visited-site*, and *Trust-in-site* fixed-response questions provide further support to this finding, showing that having an account at a site, having previously visited a site, and trusting a site are all correlated with proceeding through warnings to that site.

Participants' apparent willingness to proceed through a warning because they trust a site or have an account or visit history with a site may be in need of correction. As previously discussed, SSL warnings are not really about problems with a site itself. And malware warnings can appear on a legitimate site that has been made to host malware against its knowledge. In fact, seeing an SSL or malware warning on a site where warnings had not previously been shown is one expected symptom of a real man-in-the-middle or temporary-malware-hosting attack. It was thus heartening to see those who did not proceed through warnings due to *Problems on a trusted site*, but far fewer participants mentioned this reason not to proceed than mentioned site reputation or experience with a site as reasons to proceed. Thus, it may help if warnings make it more clear that when a warning appears on a trusted site, it's a good time *not* to proceed. This idea of basing warnings on SSL consistency has been previously proposed [45], though has not been widely adopted.

New Decision Factor Findings

Our use of an experience sampling methodology over a large sample of warning contexts revealed some new warning decision factors that, to our knowledge, have not been previously published. First, there is some good news about modern browser warnings. Many respondents who did not proceed through warnings reported trusting the warnings or their browsers. Respondents also reported that their typos and misclicks, which led to unintended and presumably malicious URLs, were caught by the warnings. This is a beneficial function for warnings we had not been aware of. Other reasons for not proceeding through warnings indicate that warnings push safe alternatives and discourage frivolous tasks.

Our finding, already discussed, that some respondents who did not proceed through warnings actually did ultimately proceed to sites—for example, by reverting to HTTP—is also new, as far as we know, and points to one direction in which browsers may help users fix possible misunderstandings about the risks of unencrypted HTTP and benefits of SSL. With the advent of free trusted certificates from the Let's Encrypt project,[5] and ever-increasing processing power, there is little reason why websites cannot forward users from HTTP to HTTPS counterparts, thereby preventing this situation from occurring.

Sample Comparison

While our two samples were recruited using different mechanisms, which led to the Chrome sample being more technically savvy, we were surprised that the reasons for proceeding or adhering to warnings were relatively similar across both samples. This suggests that regardless of technical expertise, many users have similar experiences when interacting with browser security warnings. More importantly, across both samples, our participants presented us with a variety of reasons for not adhering to warnings, which reinforces our conclusion that there is no one-size-fits-all solution for improving current web browser warnings at this point in time.

CONCLUSION

We performed a large-scale study of web browser warning behavior using the Experience Sampling Method (ESM). In so doing, we observed that when encountering warning messages *in situ*, participants have a wide variety of reasons for choosing to adhere to or proceed past a given warning. Based on our qualitative data, we conclude that warnings have improved to the point that additional gains in adherence rates are likely only to be made by examining contextual factors and a wider variety of users' concerns, rather than through one-size-fits-all improvements. Similarly, given that users make inconsistent decisions regarding whether to proceed or adhere to a warning, our results suggest that habituation plays a smaller role in user decision making than previously thought.

ACKNOWLEDGMENTS

We wish to acknowledge Mustafa Emre Acer, Arjun Baokar, Helen Harris, Ashkan Hosseini, Iulia Ion, Patrick Gage Kelley, Kris Maglione, Elisabeth Morant, Ahir Reddy, Martin Shelton, Parisa Tabriz, Jorge Villalobos, Tanvi Vyas, Amanda Walker, as well as the Chrome team for contributions to building and reviewing our browser extensions, coding our data, recruiting, and reading through drafts. Thank you all. The Berkeley team was supported by a Google Faculty Research Award.

[5] https://letsencrypt.org/

REFERENCES

1. Mustafa Acer, Emily Stark, Adrienne Porter Felt, Sascha Fahl, Radhika Bhargava, Bhanu Dev, Matt Braithwaite, Ryan Sleevi, and Parisa Tabriz. 2017. Where the Wild Warnings Are: Root Causes of Chrome Certificate Errors. In *Proceedings of the 24th ACM SIGSAC Conference on Computer and Communications Security*. https://research.google.com/pubs/pub46359.html
2. Devdatta Akhawe and Adrienne Porter Felt. 2013. Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness.. In *USENIX security symposium*, Vol. 13. https://www.usenix.org/system/files/conference/usenixsecurity13/sec13-paper_akhawe.pdf
3. Hazim Almuhimedi, Adrienne Porter Felt, Robert W. Reeder, and Sunny Consolvo. 2014. Your reputation precedes you: History, reputation, and the chrome malware warning. In *Symposium on Usable Privacy and Security (SOUPS)*, Vol. 4. 2. https://www.usenix.org/system/files/conference/soups2014/soups14-paper-almuhimedi.pdf
4. Bonnie Anderson, Tony Vance, Brock Kirwan, David Eargle, and Seth Howard. 2014. Users aren't (necessarily) lazy: using NeuroIS to explain habituation to security warnings. Auckland, New Zealand.
5. Bonnie Brinton Anderson, C. Brock Kirwan, Jeffrey L. Jenkins, David Eargle, Seth Howard, and Anthony Vance. 2015. How Polymorphic Warnings Reduce Habituation in the Brain: Insights from an fMRI Study. ACM Press, 2883–2892. DOI: http://dx.doi.org/10.1145/2702123.2702322
6. D. Anthony, T. Henderson, and D. Kotz. 2007. Privacy in Location-Aware Computing Environments. *IEEE Pervasive Computing* 6, 4 (Oct 2007), 64–72. DOI: http://dx.doi.org/10.1109/MPRV.2007.83
7. Robert Biddle, Paul C. Van Oorschot, Andrew S. Patrick, Jennifer Sobey, and Tara Whalen. 2009. Browser interfaces and extended validation SSL certificates: an empirical study. In *Proceedings of the 2009 ACM workshop on Cloud computing security*. ACM, 19–30. http://dl.acm.org/citation.cfm?id=1655012
8. Rainer Böhme and Stefan Köpsell. 2010. Trained to accept?: a field experiment on consent dialogs. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2403–2406. http://dl.acm.org/citation.cfm?id=1753689
9. Cristian Bravo-Lillo, Lorrie Faith Cranor, Julie Downs, Saranga Komanduri, and Manya Sleeper. 2011. Improving Computer Security Dialogs. In *Proceedings of the 13th IFIP TC 13 International Conference on Human-computer Interaction - Volume Part IV (INTERACT'11)*. Springer-Verlag, Berlin, Heidelberg, 18–35. http://dl.acm.org/citation.cfm?id=2042283.2042286
10. Jose Carlos Brustoloni and Ricardo Villamarin-Salomon. 2007. Improving security decisions with polymorphic and audited dialogs. In *Proceedings of the 3rd symposium on Usable privacy and security*. ACM, 76–85. http://dl.acm.org/citation.cfm?id=1280691
11. Juan Pablo Carrascal, Christopher Riederer, Vijay Erramilli, Mauro Cherubini, and Rodrigo de Oliveira. 2013. Your Browsing Behavior for a Big Mac: Economics of Personal Information Online. In *Proceedings of the 22Nd International Conference on World Wide Web (WWW '13)*. ACM, New York, NY, USA, 189–200. DOI: http://dx.doi.org/10.1145/2488388.2488406
12. Mauro Cherubini and Nuria Oliver. 2009. A refined experience sampling method to capture mobile user experience. *arXiv preprint arXiv:0906.4125* (2009).
13. Neil Chou, Robert Ledesma, Yuka Teraguchi, and John C. Mitchell. 2004. Client-Side Defense Against Web-Based Identity Theft. In *Proceedings of the 11th Annual Network and Distributed Systems Security Symposium (NDSS '04)*.
14. Karen Church and Barry Smyth. 2009. Understanding the Intent Behind Mobile Information Needs. In *Proceedings of the 14th International Conference on Intelligent User Interfaces (IUI '09)*. ACM, New York, NY, USA, 247–256. DOI: http://dx.doi.org/10.1145/1502650.1502686
15. S. Consolvo, F.R. Bentley, E.B. Hekler, and S.S. Phatak. 2017. Mobile User Research: A Practical Guide. In *Synthesis Lectures on Mobile and Pervasive Computing*. Morgan and Claypool Publishers, Chapter 4.
16. Sunny Consolvo, Ian E. Smith, Tara Matthews, Anthony LaMarca, Jason Tabert, and Pauline Powledge. 2005. Location Disclosure to Social Relations: Why, when, & What People Want to Share. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. ACM, New York, NY, USA, 81–90. DOI:http://dx.doi.org/10.1145/1054972.1054985
17. Rachna Dhamija, J. Doug Tygar, and Marti Hearst. 2006. Why phishing works. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 581–590. http://dl.acm.org/citation.cfm?id=1124861
18. Julie S. Downs, Mandy B. Holbrook, and Lorrie Faith Cranor. 2006. Decision Strategies and Susceptibility to Phishing. In *Proceedings of the Second Symposium on Usable Privacy and Security (SOUPS '06)*. ACM, New York, NY, USA, 79–90. DOI: http://dx.doi.org/10.1145/1143120.1143131
19. Serge Egelman, Lorrie Faith Cranor, and Jason Hong. 2008. 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*. ACM, 1065–1074. http://dl.acm.org/citation.cfm?id=1357219

20. Serge Egelman and Stuart Schechter. 2013. The importance of being earnest [in security warnings]. In *International Conference on Financial Cryptography and Data Security*. Springer, 52–59.

21. Adrienne Porter Felt, Alex Ainslie, Robert W. Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettes, Helen Harris, and Jeff Grimes. 2015. Improving SSL Warnings: Comprehension and Adherence. ACM Press, 2893–2902. DOI:http://dx.doi.org/10.1145/2702123.2702442

22. Adrienne Porter Felt, Robert W. Reeder, Hazim Almuhimedi, and Sunny Consolvo. 2014. Experimenting at scale with Google Chrome's SSL warning. ACM Press, 2667–2670. DOI: http://dx.doi.org/10.1145/2556288.2557292

23. Nathaniel Good, Rachna Dhamija, Jens Grossklags, David Thaw, Steven Aronowitz, Deirdre Mulligan, and Joseph Konstan. 2005. Stopping spyware at the gate: a user study of privacy, notice and spyware. In *Proceedings of the 2005 symposium on Usable privacy and security*. ACM, 43–52. http://dl.acm.org/citation.cfm?id=1073006

24. Nathaniel S. Good, Jens Grossklags, Deirdre K. Mulligan, and Joseph A. Konstan. 2007. Noticing notice: a large-scale experiment on the timing of software license agreements. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 607–616. http://dl.acm.org/citation.cfm?id=1240720

25. Jonna Häkkilä, Farnaz Vahabpour, Ashley Colley, Jani Väyrynen, and Timo Koskela. 2015. Design Probes Study on User Perceptions of a Smart Glasses Concept. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia (MUM '15)*. ACM, New York, NY, USA, 223–233. DOI: http://dx.doi.org/10.1145/2836041.2836064

26. Marian Harbach, Emanuel von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. 2014. It's a hard lock life: A field study of smartphone (un)locking behavior and risk perception. In *Symposium on Usable Privacy and Security (SOUPS)*.

27. Eiji Hayashi and Jason Hong. 2011. A diary study of password usage in daily life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2627–2630.

28. Cormac Herley. 2009. So long, and no thanks for the externalities: the rational rejection of security advice by users. In *Proceedings of the 2009 workshop on New security paradigms workshop*. ACM, 133–144. http://dl.acm.org/citation.cfm?id=1719050

29. Giovanni Iachello, Khai N. Truong, Gregory D. Abowd, Gillian R. Hayes, and Molly Stevens. 2006. Prototyping and Sampling Experience to Evaluate Ubiquitous Computing Privacy in the Real World. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 1009–1018. DOI: http://dx.doi.org/10.1145/1124772.1124923

30. Philip G Inglesant and M Angela Sasse. 2010. The true cost of unusable password policies: password use in the wild. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 383–392.

31. Kat Krol, Matthew Moroz, and M. Angela Sasse. 2012. Don't work. Can't work? Why it's time to rethink security warnings. In *risk and security of internet and systems (CRiSIS), 2012 7th International conference on*. IEEE, 1–8. http://ieeexplore.ieee.org/abstract/document/6378951/

32. Lawrence L. Kupper and Kerry B. Hafner. 1989. On Assessing Interrater Agreement for Multiple Attribute Responses. *Biometrics* 45, 3 (Sept. 1989), 957. DOI: http://dx.doi.org/10.2307/2531695

33. J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 1 (March 1977), 159. DOI: http://dx.doi.org/10.2307/2529310

34. Reed Larson and Mihaly Csikszentmihalyi. 1983. The Experience Sampling Method. *New Directions for Methodology of Social & Behavioral Science* 15 (1983), 41–56.

35. Sameer Patil, Roberto Hoyle, Roman Schlegel, Apu Kapadia, and Adam J. Lee. 2015. Interrupt Now or Inform Later?: Comparing Immediate and Delayed Privacy Feedback. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1415–1418. DOI: http://dx.doi.org/10.1145/2702123.2702165

36. Niels Provos. 2012. Safe Browsing—Protecting Web Users for Five Years and Counting. https://www.blog.google/topics/safety-security/safe-browsingprotecting-web-users-for/. (June 19 2012). Accessed: September 18, 2017.

37. Stuart E. Schechter, Rachna Dhamija, Andy Ozment, and Ian Fischer. 2007. The emperor's new security indicators. In *Security and Privacy, 2007. SP'07. IEEE Symposium on*. IEEE, 51–65. http://ieeexplore.ieee.org/abstract/document/4223213/

38. Fuming Shih, Ilaria Liccardi, and Daniel Weitzner. 2015. Privacy Tipping Points in Smartphones Privacy Preferences. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 807–816. DOI: http://dx.doi.org/10.1145/2702123.2702404

39. Manya Sleeper, Rebecca Balebako, Sauvik Das, Amber Lynn McConahy, Jason Wiese, and Lorrie Faith Cranor. 2013. The post that wasn't: exploring self-censorship on facebook. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 793–802.

40. Andreas Sotirakopoulos, Kirstie Hawkey, and Konstantin Beznosov. 2011. On the challenges in usable security lab studies: lessons learned from replicating a study on SSL warnings. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*. ACM, 3. http://dl.acm.org/citation.cfm?id=2078831

41. Joshua Sunshine, Serge Egelman, Hazim Almuhimedi, Neha Atri, and Lorrie Faith Cranor. 2009. Crying Wolf: An Empirical Study of SSL Warning Effectiveness. In *USENIX security symposium*. 399–416. http://static.usenix.org/legacy/events/sec09/tech/full_papers/sec09_browser.pdf

42. David R. Thomas. 2006. A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation* 27, 2 (June 2006), 237–246. DOI: http://dx.doi.org/10.1177/1098214005283748

43. Rick Wash and Emilee J. Rader. 2015. Too Much Knowledge? Security Beliefs and Protective Behaviors Among United States Internet Users.. In *SOUPS*. 309–325.

44. Joel Weinberger and Adrienne Porter Felt. 2016. A week to remember: The impact of browser warning storage policies. In *Symposium on Usable Privacy and Security*. https://www.usenix.org/system/files/conference/soups2016/soups2016-paper-weinberger.pdf

45. Dan Wendlandt, David G. Andersen, and Adrian Perrig. 2008. Perspectives: Improving SSH-style Host Authentication with Multi-Path Probing. In *USENIX '08: Proceedings of the 2008 USENIX Annual Technical Conference*. USENIX Association, Berkeley, CA, USA.

46. Primal Wijesekera, Arjun Baokar, Lynn Tsai, Joel Reardon, Serge Egelman, David Wagner, and Konstantin Beznosov. 2017. The Feasability of Dynamically Granted Permissions:
Aligning Mobile Privacy with User Preferences. In *Proceedings of the 2017 IEEE Symposium on Security and Privacy (Oakland '17)*. IEEE Computer Society.

47. Will Wiquist. 2016. FCC settles Verizon "supercookie" probe, requires consumer opt-in for third parties. Federal Communications Commission press release. Mar 7, 2016. https://apps.fcc.gov/edocs_public/attachmatch/DOC-338091A1.pdf. (2016).

48. Min Wu, Robert C. Miller, and Simson L. Garfinkel. 2006. Do security toolbars actually prevent phishing attacks?. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 601–610. http://dl.acm.org/citation.cfm?id=1124863