Personalized Security Messaging: Nudges for Compliance with Browser Warnings

Nathan Malkin*, Arunesh Mathur†, Marian Harbach‡ and Serge Egelman*‡
*University of California, Berkeley †Princeton University ‡International Computer Science Institute
nmalkin@cs.berkeley.edu amathur@cs.princeton.edu mharbach@icsi.berkeley.edu egelman@cs.berkeley.edu

***Abstract*—Decades of psychology and decision-making research show that everyone makes decisions differently; yet security messaging is still one-size-fits-all. This suggests that we can improve outcomes by delivering information relevant to how each individual makes decisions. We tested this hypothesis by designing messaging customized for stable personality traits—specifically, the five dimensions of the General Decision-Making Style (GDMS) instrument. We applied this messaging to browser warnings, security messaging encountered by millions of web users on a regular basis. To test the efficacy of our nudges, we conducted experiments with 1,276 participants, who encountered a warning about broken HTTPS due to an invalid certificate under realistic circumstances. While the effects of some nudges correlated with certain traits in a statistically significant manner, we could not reject the null hypothesis—that the intervention did not affect the subjects' behavior—for most of our nudges, especially after accounting for participants who did not pay close attention to the message. In this paper, we present the detailed results of our experiments, discuss potential reasons for why the outcome contradicts the decision-making research, and identify lessons for researchers based on our experience.**

I. Introduction

Today's systems often call on users to make security decisions, and outcomes show that the resulting choices are frequently suboptimal. Browser warnings are one example of such a situation. Web browsers warn users if they are about to visit phishing pages, attack sites, and domains with invalid TLS certificates. Given the option to ignore these warnings, many do, putting their computers and data at risk. Outright preventing users from visiting these websites may lead to worse outcomes: they may switch to a different web browser that does not yet detect the threat, or they may permanently disable the security features that appear to be the impediments. As a result, messaging needs to be crafted in a way that leads users towards making correct choices without forcing them; users themselves should want to make the right decision. It remains an open question how to present this information in a way that leads to users making better decisions.

One approach is choice architecture, an idea originating in the field of behavioral economics and popularized by Thaler and Sunstein in their book *Nudge* [40]. Because of naturally occurring heuristics and cognitive biases, there will always be options people are more likely to select—for example, the default answer on a questionnaire or the closest item in a buffet line. The choice architect—the person responsible for designing the experience—can therefore improve outcomes by setting that option to be the one that is most beneficial to either the individual user or society at large. In doing so, they *nudge* decision-makers towards better choices, without taking away their freedoms; this idea is known as "libertarian paternalism" or "soft paternalism."

While choice architecture suggests a general approach, it does not prescribe a particular solution. Setting the default choice in a menu of options is one effective strategy, though one that is already being used: the user sees a security warning and not their intended website, and must take additional actions (i.e., click through the warning) to proceed with their original plan. Another option is to focus the messaging on certain aspects of the choice. Since any interface is subject to limited space and attention, it makes sense for the interface designer to use it to deliver the most effective message possible.

But what is an effective message? The answer is complicated, because it depends on the person making the choice, and everyone makes decisions differently. Fortunately, psychologists and marketers have been able to identify trends and systematize decision-making procedures, allowing them to measure which group each of us belongs to. For example, "dependent" decision-makers take into account what others are doing before deciding; some people are guided by "rational" reasons, while emotions drive others. Ideally, messaging could be customized to each of these styles. In fact, this is already happening, with marketers and political campaigns making active use of behavioral advertising. Yet, security messaging remains one-size-fits-all.

What would security messaging look like if it were customized to how each of us makes decisions? Someone who looks to experts for guidance could be shown their advice, and someone who wants to see the statistics for themselves could get the numbers they crave (e.g., a quantifiable risk metric). The goal of our study is to test the efficacy of this approach by developing and deploying personalized nudges within the web browser.

We implemented our nudges as modifications to existing browser warnings—specifically, warnings about invalid TLS certificates. Ultimately, we chose to target HTTPS warnings for several reasons:

- It is a common security decision—almost any Internet

EuroUSEC '17, 29 April 2017, Paris, France

http://dx.doi.org/10.14722/eurousec.2017.23008

user will encounter one of these warnings sooner or later [3].

- Warnings about TLS certificates have low compliance rates, especially compared to other browser warnings [44].
- Most users lack the understanding of network security (such as TLS certificates, their trust models and failure modes) needed to make a fully informed decision. As a result, they may be more willing to accept the information and guidance offered in a nudge.
- Compliance with a warning is limited in scope, affecting only the current site for a limited duration, as opposed to the repeated, lasting effects of a decision such as choosing a password or a lock-screen type. Consequently, users may have fewer competing considerations when making the decision.

A second major design decision is the evaluation strategy. Traditional laboratory experiments allow for the greatest flexibility and customization. However, they cannot adequately simulate real-world security decisions, because they function to reduce risk: participants are usually not dealing with real personal information, have limited consequences for their actions, and even sometimes use devices provided by researchers (rather than personal equipment). Lab experiments also feature lowered uncertainty, as instructions and guidance from researchers may establish expectations.

We therefore chose to validate our nudges in a scenario most similar to the natural situation in which a user would encounter a warning from their browser. Workers on a crowdsourcing platform (Mechanical Turk) were asked to complete a task that involved visiting and reviewing multiple web pages. One of these presented them with a simulated warning that appeared to come from their browser, customized to include one of our nudges. We measured participants' compliance with this warning as the key dependent variable in our experiment.

Our analysis of the results of this experiment showed that some nudges were more effective for certain personality traits. For example, a nudge that cited statistics increased compliance among more rational decision-makers. However, most of the other predicted effects were not observed. Furthermore, the effects that were observed were not robust: they did not manifest in a follow-up validation study, where we also checked whether participants could recall the warning they had seen. Among those who did, compliance did not correlate with personality traits as hypothesized.

While these results appear to contradict decision-making research, which predicts that people *should* respond differently to the varying messages, there are a number of possible explanations for our null result; we explore these in detail in our discussion. In addition to this, our contributions include a series of security-focused nudges based on the General Decision-Making Style scale and a methodology for evaluating browser warnings under realistic conditions.

II. Related work

In this section, we present prior research on the design of web browser security warnings, nudging, personalization, and connections between security decisions and personality.

A. Browser warnings and warning adherence

Modern browsers warn their users when they are about to enter an unsafe situation. These can include browsing to sites known to host malware or phishing pages, or connecting to a website whose TLS certificate could not be properly validated. The intent of the warnings is to communicate the danger present and to persuade the user to turn back, unless they are certain the target site is benevolent. Research in usable security has shown that browsers have consistently struggled with both of these goals. Users may override the browser's recommendation and proceed through the warning, and they frequently do.

Early incarnations of both phishing warnings [15] and SSL warnings [38] were especially unsuccessful at keeping users from danger. Design improvements based in part on these findings have led to incremental improvements in the warnings' design—and, consequently, their adherence rates. As a result, more recent data suggests that phishing and malware warnings have generally high adherence rates, at around 90%, with adherence rates for HTTPS errors noticeably lower, at 30–70% [3], [44].

Researchers have investigated how the use of imagery, extra steps before the user can proceed, and style choices can improve warning adherence [22]. They have also evaluated how related UI features, such as the connection security indicator, can be used most effectively to communicate the connection status [21]. Recent work has also uncovered that significant variations in the adherence rates among browsers can be attributed to their storage policy: how long exceptions are remembered [44].

More broadly, studies have shed light on why people ignore security warnings in general—not just from browsers. Across a variety of software, users believe, often falsely, that they are able to recognize and react to the true threats; moreover, frequent false alarms have desensitized users to the messages' advice [31]; The effects of such habituation have even been observed at the neurological level [5], [6]. Consequently, while precise numbers are not available, evidence suggests that people read warnings and other security messages only infrequently: for example, Felt et al. found that only 17% of Android users paid attention to permissions before installing an app [20]. Another perspective suggests that users are insufficiently incentivized to comply with warnings or, indeed, adhere to most security advice [25].

B. Nudging

Thaler and Sunstein define nudges as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options" [40]. A majority of the work to make security and privacy features more usable falls under this definition and can consequently be viewed through the lens of behavioral economics [1].

A smaller but sizable subset of research has been directly inspired by behavioral economics research, seeking to incorporate its ideas into interventions. Egelman et al., drawing on the ideas of choice architecture, showed that when privacy information was presented in a way that allowed for easy comparison, people would incorporate it in their purchase

decisions [16]. Inspired by the notion of framing effects, Choe et al. designed security visualizations that had identical meanings, differing only in the way they were framed [13].

A number of studies worked to bring nudges directly into user interfaces. Balebako et al. proposed implementing nudges for location sharing and Twitter privacy settings [7]. Besmer et al. added social cues to Facebook apps' permission requests, displaying the percentage of people who granted the permission in question [10]. Also modifying the Facebook interface, Wang et al. nudged users to consider what they posted and with whom they shared it [43]. Almuhimedi et al. used nudges to raise users' awareness about the data their smartphone apps collected about them [4]. These interventions were generally successful; for example, in the latter study, 95% of participants reassessed app permissions, and 58% added restrictions.

Our study builds on the successes of nudges in human-computer interaction but focuses on security decisions, in contrast to the prior studies, which primarily focused on privacy choices.

C. Personalization

While personalization is rarely encountered in real-world security systems, researchers have explored the idea in various contexts. Privacy has been a popular target, since privacy preferences and decisions have been shown to differ significantly among people [45].

Harbach et al. personalized apps' permission requests by showing examples of their own information that would be accessed and shared by the app [24]. Another approach to personalization has been to create location privacy preference recommendations through collaborative filtering [47], [48]. Other people's decisions can also be used in the form of crowdsourcing; Ismail et al. used it to find the minimal permission set that preserves the app's usability [26]. Recently, Liu et al. developed a Personalized Privacy Assistant, which recommended Android permission settings based on the user's existing permissions and responses to tailored questions [32]. However, to date, no study has focused on personalization through an individual's personality traits.

D. Security & personality

Researchers have explored the connections between security and personality traits, with most of the early work focusing on the Five Factor Model, a widely used scale measuring openness, conscientiousness, extraversion, agreeableness, and neuroticism [27]. Halevi et al. studied correlation between the Big Five personality traits and susceptibility to email phishing [23]. Uebelacker et al. proposed the traits as predictors of susceptibility to social engineering attacks [42]. Minkus and Memon used them to personalize Facebook privacy preferences [33].

However, our group's recent work has shown that the Big Five traits are comparatively weak predictors of privacy attitudes, while other psychometrics offer better predictive power; these include scales measuring Need for Cognition (NFC), Domain Specific Risk-Taking (DoSpeRT), General Decision Making Style (GDMS), and Consideration for Future Consequences (CFC) [17], [18]. Our study builds on these findings by exploiting these differences to personalize nudges for browser warnings.

III. Nudge Development

Our methodology for developing nudges consisted of two steps. First, we selected the traits we wished to target. Then, we developed messaging designed specifically to appeal to those who have these traits.

A. Selecting the psychometrics

To develop our nudges, we sought psychometrics that satisfied the following requirements:

- *Substantiated*: There must be ample research evidence of systematic differences among subsets of the population.
- *Stable*: They should not significantly vary due to a person's affect or otherwise change over time.
- *Relevant*: The trait that demonstrates systematic differences should play a role in the decision process.

Based on these criteria, we selected the General Decision Making Style (GDMS) and Need for Cognition (NFC) instruments. Prior work by Egelman and Peer also found that these scales correlated with users' computer security intentions [19].

GDMS assesses the way in which individuals approach decision situations. Each individual is measured along five different sub-scales, representing different decision styles [35]:

- Rational: "A thorough search for and logical evaluation of alternatives."
- Intuitive: "A reliance on hunches and feelings."
- Dependent: "A search for advice and direction from others."
- Avoidant: "Attempts to avoid decision making."
- Spontaneous: "A sense of immediacy and a desire to get through the decision-making process as soon as possible."

NFC is a single scale that measures a person's "tendency to engage in and enjoy thinking" [11]. For example, people who measure highly on this scale tend to agree with statements like "I prefer my life to be filled with puzzles that I must solve" [12].

B. Designing the nudges

From GDMS, we decided to target dependent, rational, and avoidant decision-making, which were the three sub-scales found to correlate with noticing security indicators [19].

Dependent people look to others for guidance. This could be others in their social circle, high-status individuals (e.g., experts), or people in general. We could not plausibly use social information (e.g., how many people known to the user visited the same website), but for the other cases it was plausible that a browser maker could obtain and share this data. Thus, we designed two nudges: a `social` message, providing

TABLE I. NUDGES AND THEIR ASSOCIATED PSYCHOMETRICS

Nudge	Trait	Text
`Social`	GDMS-Dependent	81.3% of people who saw this warning clicked "Back to safety."
`Expertise`	GDMS-Dependent	Security experts strongly recommend clicking "Back to safety."
`Positive Frame`	GDMS-Rational, NFC	Clicking "Back to safety" significantly reduces the risk of online fraud.
`Negative Frame`	GDMS-Rational, NFC	Not clicking "Back to safety" significantly increases the risk of online fraud.
`Statistics`	GDMS-Rational, NFC	Clicking "Back to safety" has been found to reduce the risk of online fraud by 81.3%.
`Difficulty`	GDMS-Avoidant	Clicking "Back to safety" is enough to protect you from online fraud in many cases.

information about the behavior of people in general, and one appealing to the authority of `experts`.

While, in practice, a social warning could use actual click-through rates by aggregating them anonymously, for consistency and reproducibility, we opted for a static social value (fixed at 81.3%). This number was chosen to be sufficiently large to demonstrate the alleged majority's clear preference, but not so high (or round) as to arouse suspicion by being implausible. We planned to test different values in further experiments, to determine, for example, what threshold yields greatest compliance, whether negative framing is more effective (e.g., 81.3% obeyed the warning vs. 18.7% ignored the warning), and so on.

Rational decision-makers like to logically evaluate alternatives and consequently find data appealing. We therefore designed a data-centric nudge (`Statistics`) allegedly substantiating the safety benefits of complying with the warning (see below).

However, for a rational decision-maker, data is not strictly necessary. A logical and persuasive argument should be sufficient. It may therefore be enough to state the benefits of warning compliance (for example, that it reduces the risk of fraud) without citing statistics. Thus, we also included a nudge with just that statement (`Positive Frame`).

A separate body of literature, also in behavioral economics, has documented that the framing of a decision can have an impact on which option people choose—even if the outcomes of the two presentations are identical. Specifically, prospect theory suggests that people may be more risk-averse or risk-seeking depending on whether they perceive the outcome as a gain or loss (and its magnitude) [41]. To check for framing effects, we tested a variant of the previous nudge where the same information was framed as a loss (`Negative Frame`), rather than a gain.

We further hypothesized that the nudges aimed at rational decision-makers would be moderated by Need for Cognition (NFC).

Finally, we designed a nudge that emphasized that safety can be achieved with little effort, simply by choosing not to proceed through the warning. We expected this message to appeal to people with an avoidant decision-making style, by giving them an option where inaction was framed as positive.

The nudges, their exact wordings, and the traits they targeted are summarized in Table I.

IV. METHODOLOGY

In this section, we describe our experimental design for implementing and evaluating our nudges, ethical considerations, a follow-up experiment to examine awareness of the nudges, and finally, our method of obtaining each participant's psychometrics.

A. Experimental Design

The goal of our study was to examine the compliance rates with our new browser warnings. To accomplish this, we developed a methodology to test simulated warnings under maximally realistic conditions.

We drew participants from the Amazon Mechanical Turk crowdworking marketplace. MTurk is a common source of participants for social science and human-computer interaction research, including research in usable privacy and security [30]. Studies have shown that subjects from the Mechanical Turk population are more diverse than those drawn from traditional university participant pools [9] and produce results similar to those recruited from other sources [8], including nationally representative samples [37].

Workers on Mechanical Turk were hired to complete a task unrelated to computer security: we asked them to visit different websites to view their designs and answer questions about them, similar to many other website reviewing and categorization tasks that are routinely posted on Mechanical Turk. We asked our participants to answer questions such as, "what was the first thing you noticed on the website?" Each participant visited five websites for this task. The survey presented the websites in a random order; most of them belonged to real banks. We chose this theme to provide a security-sensitive context—as consumers generally consider financial information to be especially sensitive. However, at no point were participants expected to enter any personal information or otherwise interact with the websites. We assumed that the act of visiting a financial website might be sufficient to prime participants to being sensitive to security.

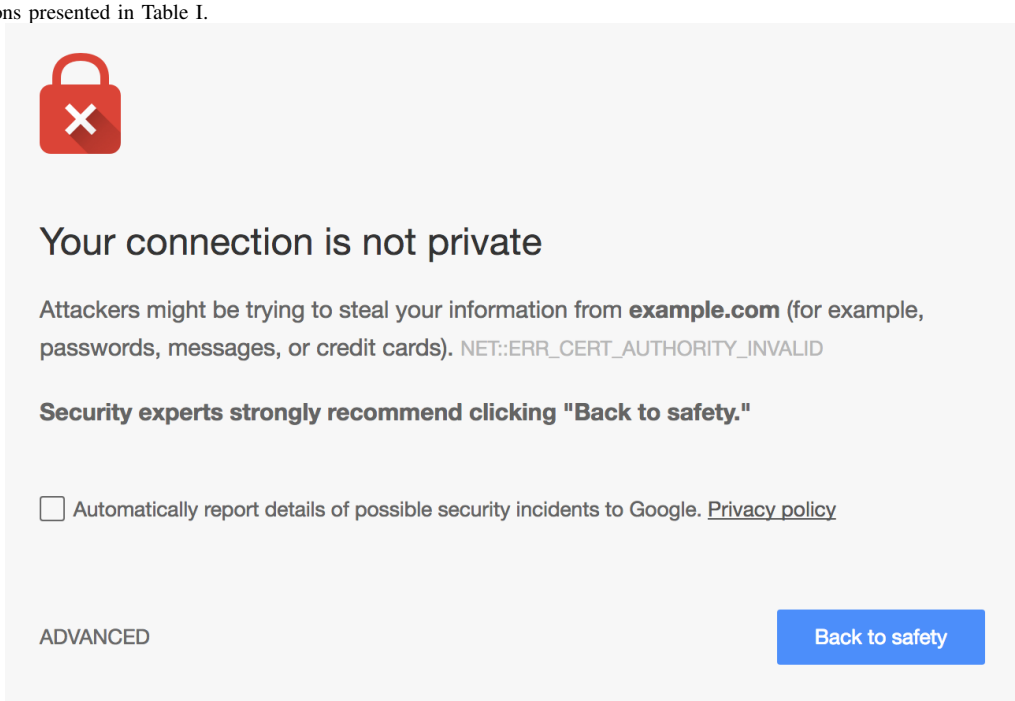
Rather than belonging to a real bank, one of the websites was a decoy page hosted by us. It detected the browser its visitor was using and served a page designed to look like that browser's HTTPS error.[1] Supported browsers included Google Chrome, Microsoft Edge and Internet Explorer, and Mozilla Firefox (including two generations of warning designs).[2] Participants who clicked through the warning saw what appeared to be a bank website, though created by us.

We also included an additional website that yielded a DNS error: the website could not be resolved, and therefore could not be accessed. This had the benefit of making the study appear to be about errors in general (i.e., not just security

[1]The specific error was that the presented certificate was self-signed, but most browsers do not differentiate the specific TLS errors or provide them only as "Advanced" information.

[2]Participants using other browsers encountered a real HTTPS warning to ensure a consistent experience. However, data from them is excluded from our analysis.

Fig. 1. The `Expertise` nudge, as seen in the Chrome browser. The nudge itself appears below the description of the warning and was randomly chosen from the six conditions presented in Table I.



warnings), as well as for us to examine whether participants would actually report an error, rather than making up a fake review of an inaccessible website.[3]

Specifically, participants visited the following websites:

- http://www.ucbfirst.com/ (decoy, unable to load)
- http://www.reddingfirst.com/ (decoy, HTTPS warning)
- http://www.bancfirst.com/
- http://www.midfirst.com/
- http://www.simmonsfirst.com/

We randomly assigned participants to one of the six nudge conditions or the control (the absence of any nudge text). If in a nudge condition, the standard HTTPS warning was augmented with the nudge text on a separate line in bold font, but otherwise matching the style of the warning. Figure 1 shows a sample warning in the `Expertise` condition. The text of the nudge was customized to match the text of the UI (e.g., referring to clicking on "Back to safety" in Chrome, but "Get me out of here" for Firefox). We instrumented the warning page to record the visitors' behavior, including whether they decided to heed the warning or ignore it, and how much time they spent viewing it.

Based on prior work by Sunshine et al. [38], we wanted to give participants an alternative way of completing the task after viewing the warning, so that they would not feel compelled to click through it to complete the tasks. Thus, instead of answering questions about the website's design, participants could report a problem with it ("Were you able to access the website?"), which did not impact their compensation for the task. As a result, there was no special incentive for ignoring the warning. Consequently, we believe that our participants did not feel coerced into bypassing the warning and their decisions are comparable to ones they make under more "natural" circumstances.

B. Ethical considerations

Our study utilized limited deception by simulating a warning that came from participants' web browsers, rather than having it actually display one. Since warnings are a regular occurrence on the web, we do not believe we caused our participants extra distress or harm by showing it to them. As described above, we asked the workers if they encountered any problems with the site before asking any further questions about it, and compensation was not affected by their response, so the participants should not have felt coerced into ignoring the warning. By controlling the domain in question, as well as the link that referred to it, and collecting all other information over HTTPS, this procedure did not expose our participants to any potential harm. We therefore judged that a debrief at the end of the study would cause greater distress and prevent replicability, and obtained a waiver allowing us to omit it. However, we obtained consent from all subjects to participate and collect their data. This procedure was approved by the UC Berkeley Committee for Protection of Human Subjects.

C. Follow-up study: testing recall

A benefit of our study's design is that, even at its conclusion, participants likely believed that the warning they saw came from their browser, allowing for similar follow-up experiments.[4] The downside of this approach is that we cannot

[3]Two participants failed to note this error, however, since both correctly identified the security warning, we still included their data.

[4]In our study, however, we prevented subjects who completed the first study from participating in the follow-up.

effectively tell whether they read and internalized the contents of the warning, including our nudge. Immediately asking them questions about the warning would have revealed the purpose of the study, thereby potentially tainting future participants. While we collected statistics such as the time spent on the page, this turned out to be a poor metric for engagement, as many participants kept the page open for an extended period of time (for example, by returning to the primary survey without closing the new window).

To attain a better understanding of our participants' engagement with the warning—in particular, whether they read and remembered it—we performed a follow-up experiment on a new sample, using the original procedure with a few small modifications:

- At the conclusion of the study, we presented participants with all of the possible nudges and asked which, if any, they had encountered. Participants could also state, without a penalty, that they could not remember if they had seen any of the options.
- To reduce the likelihood of forgetting what they had seen prior to completing this question, the HTTPS warning was always the last of the websites visited.
- The survey automatically closed the opened website before proceeding, to prevent "cheating" on the recall question.
- To compensate for the length of the task, the total number of websites was reduced from five to three, removing one legitimate banking website and the inaccessible decoy, and the irrelevant questions on website design were slightly shortened.

The intent of testing participants' recall was to allow us to limit analysis to those who—we could say with reasonable certainty—had read the nudge. While we would not expect everyone who was influenced by the nudge to be able to recall it, we can be quite certain that anyone who did not read the message would not be influenced by it. Analyzing only the subset of participants who we knew read the nudge can therefore tell us more about the effectiveness of the nudge itself, instead of the warning design or any other contextual information.

D. Obtaining psychometrics

Only people who completed a separate Mechanical Turk qualification task, consisting of the GDMS and NFC psychometric scales, were eligible to participate in our studies. The connection between the two tasks was not advertised, and we enforced a waiting period between collecting the psychometrics and entering our primary experiment. This was done in order to reduce the effects of priming on our participants.

V. Results

A. Participants

We recruited 680 workers on Mechanical Turk for the initial study. After excluding those who did not see a simulated browser warning (for example, because they used an unsupported browser or did not follow the directions in the survey), we were left with 662 complete data points. Participants were restricted by the platform to be from the United States and those who had completed 500 previous tasks with a 95% approval rate or above. Our participants were evenly split between female and male (49.8% and 49.7%, respectively). The age of our subjects ranged from 20 to 75, with the median at 33 and a standard deviation of 11. Our participants were relatively well educated: 54.9% reported having a bachelor's degree or higher. Regarding web browser usage, 68.0% of our participants used Chrome, 24.8% used Firefox, with the remainder using other browsers. The number of participants per randomly-assigned condition varied from 85 to 104 (see Table II).

TABLE II. Participants per condition, in each experiment

Condition	Study 1	Study 2
Control	85	91
Social	88	76
Expertise	96	92
Positive Frame	98	75
Negative Frame	92	104
Statistics	99	96
Difficulty	104	96

TABLE III. Warning adherence rate

Condition	Study 1	Study 2
Control	71.8%	78.0%
Social	78.4%	78.9%
Expertise	65.6%	79.3%
Positive Frame	79.6%	76.0%
Negative Frame	83.7%	66.3%
Statistics	76.8%	78.1%
Difficulty	72.1%	75.0%

B. General results

Participants interacted with the simulated browser warning as if it were real, and many chose to adhere to its advice. Overall, 75.4% obeyed the warning and did not visit the website; This number differed based on the condition: 71.8% participants in the control condition adhered to the warning, and the adherence rate was higher for most (but not all) nudges (see Table III), varying from 65.6% to 83.7%. However, our analysis showed that, controlling for all factors including personality, the differences between the conditions were not significant (see Appendix).

C. Hypotheses tested

For each of the six nudges, we hypothesized that one or more traits would be correlated with participants obeying the warning. To test each hypothesis, we performed a logistic regression on a subset of the data, consisting of all participants in the given condition as well as those in the control condition. Logistic regression was a natural choice for this test, as it is able to predict a binary outcome (proceeding through the warning), while accounting for multiple explanatory variables, including interaction between these.

The participant's decision whether or not to proceed was the dependent variable; as independent variables, we used the targeted GDMS sub-scale and its interaction effect with the condition. If the latter turned out to be statistically significant, this would suggest the effectiveness of our nudge, as it would indicate that the targeted trait was a significant predictor of behavior *only in the presence of that nudge*.

TABLE IV. LOGISTIC REGRESSION FOR STATISTICS NUDGE, STUDY 1.
(** DENOTES $p < 0.01$, INTERACTION TERMS DENOTED WITH +)

Variable	Estimate	Std. Error	z value	$Pr(>\|z\|)$	
(Intercept)	-4.5920	2.0848	-2.203	0.02762	*
Browser: Edge	-16.8421	2232.3539	-0.008	0.99398	
Browser: Firefox	0.1545	1.2375	0.125	0.90064	
Browser: Firefox (New)	-0.8078	0.5004	-1.614	0.10644	
Browser: Internet Explorer	-17.7700	3956.1804	-0.004	0.99642	
Browser: Other	-17.0404	1370.3876	-0.012	0.99008	
Condition: Statistics	7.5064	2.9062	2.583	0.00980	**
GDMS-Rational	0.7849	0.5044	1.556	0.11965	
NFC	0.2090	0.3561	0.587	0.55736	
Statistics + Rational	-2.0226	0.7469	-2.708	0.00677	**
Statistics + NFC	0.1190	0.5063	0.235	0.81412	

TABLE V. LOGISTIC REGRESSION FOR NEGATIVE FRAME NUDGE, STUDY 2.
(* DENOTES $p < 0.05$, INTERACTION TERMS DENOTED WITH +)

Variable	Estimate	Std. Error	z value	$Pr(>\|z\|)$	
(Intercept)	-4.7288	2.3698	-1.995	0.0460	*
Browser: Edge	-17.4667	2235.4582	-0.008	0.9938	
Browser: Firefox	-16.2233	3956.1804	-0.004	0.9967	
Browser: Firefox (New)	-0.3866	0.4538	-0.852	0.3944	
Browser: Internet Explorer	-17.1520	1923.4461	-0.009	0.9929	
Browser: Other	-17.0150	1134.2781	-0.015	0.9880	
Condition: Negative Frame	3.8275	2.9327	1.305	0.1918	
GDMS-Rational	0.5876	0.5400	1.088	0.2765	
NFC	0.4093	0.4641	0.882	0.3778	
Negative Frame + Rational	0.2413	0.6946	0.347	0.7282	
Negative Frame + NFC	-1.2454	0.6104	-2.040	0.0413	*

Prior research has shown that different browsers demonstrate varying adherence rates [3], an effect at least partially explained by the warning's design and content [22], [44]. To control for this variation, we also included the user's browser as an independent variable.[5]

We formulated the null hypothesis as follows: adherence rates in the presence of the given nudge are *not* correlated with the targeted psychometric scale. We could therefore reject the null hypothesis for a particular nudge if the interaction effect in the associated logistic regression was found to be statistically significant. This would indicate that, in the presence of the nudge, the targeted trait had a significant effect on whether the participant clicked through the warning or not.

In our analysis, we could reject the null hypothesis in this way for only one of the six nudges. In the `Statistics` condition, the interaction with the *rational* sub-scale of GDMS was significant ($p < 0.01$, see Table IV). The interpretation of this result is that, among those who see the `Statistics` nudge, the odds of ignoring the warning decrease multiplicatively by 0.3 for every one-point increase in the (five-point) GDMS-rational scale.

For each of the other nudges, the interaction coefficient was not significant, and consequently we could not reject the null hypothesis for any of them.

D. Did participants pay attention to the warnings?

One potential explanation for the non-results seen above is that participants did not read the nudges. This could have happened for a variety of reasons: users may be in the habit of clicking through or away from a warning, they may not read through the entire warning before making a decision, or they may react to non-textual cues on the page (such as the colors or icons).

In our original study, we do not have a reliable way of distinguishing those who read the nudge from those who did not. While we collected the time each participant spent on the page, longer stays do not necessarily imply careful reading: we recorded dwell times of minutes or longer, suggesting that participants left the page open and moved on to other tasks. On the opposite end of the spectrum, some durations may be too short for someone to plausibly read the entire message during this time. We experimented with lower bounds between 5 and 10 seconds, but found that this affected only a small number of participants, and excluding their data did not materially change the results.

Therefore, to test our hypothesis, we developed and executed the "follow-up study" described in Section IV-C, asking participants to recall which nudge they saw, immediately after completing the decoy task for the website with the warning. We received 614 complete responses for this follow-up experiment (see Table II for a breakdown by condition). The demographics of the sample closely resembled those of the first experiment: 51.8% female, median age of 31, and 48.7% with a bachelor's degree or higher.

Replication experiment: Omitting the responses to the recall question, this experiment serves as a replication of the original study.[6] To test the robustness of the first study's results, we repeated the associated analysis with the new data, testing the hypothesis for each of the nudges. We found that we could only partially replicate the original results with the new data. Similarly to the original study, for most nudges, we found no correlation between the target traits and warning adherence. This extended to the `Statistics` nudge: unlike in the first study, the correlation with the GDMS-rational decision-making style was not significant, at $p > 0.05$. However, the model found a different nudge to be significant: among participants who viewed the `Negative Frame` nudge, an increase in NFC was associated with greater adherence to the warning (see Table V).

Recall rate: Only a relatively small fraction of our participants—35.0%—was able to successfully answer the recall question by correctly identifying the message they had been shown as part of the HTTPS warning. Of the others, approximately a third (20.8% of all participants) admitted that they did not remember which (if any) message they had seen. The remainder, 44.1% of the participant pool, answered the recall question incorrectly, with slightly less than half claiming that they did not see any of the presented nudges, and the others choosing the wrong one.[7]

The rather low recall rate—despite the question appearing shortly after the nudge was presented—suggests that many of

[5] In the resulting models, some browsers have a very large standard error (e.g., in Tables IV and V) due to the small number of participants who used them.

[6] While there were small procedural differences, detailed above, we do not believe they were likely to substantially affect how the participants made their decisions. The changes to the survey instrument, wherein we explicitly asked participants about the nudge, were only shown to participants after they had made a decision about whether to obey the warning.

[7] A participant randomly guessing would have a one-in-seven chance of correctly answering the recall question. Consequently, random chance could account for some but not all of the correct answers.

our participants did not read the warning they saw carefully, making their decision on the basis of prior knowledge or beliefs about HTTPS warnings, rather than the details of our nudge.

We found that the memorability of the conditions varied significantly: as low as 18.8% in the `Difficulty` condition, compared to 51.6% in the control. Only the `Social` condition had a higher recall than the control, at 65.8%. Despite this condition's memorability, participants who were in this group ignored the warning at a higher rate than those in all other conditions (when looking only at those who passed the recall question). In general, however, participants who passed the recall question had an adherence rate that was almost 10% higher than the group as a whole.

Incorporating recall information: We continued our analysis by excluding all participants who could not remember which nudge they had seen, as well as anyone who answered the recall question incorrectly. This left us with 215 participants, 35.0% of the original sample.[8] We then repeated the hypothesis testing on this limited dataset, once again looking for correlation between traits and conditions. In this analysis, none of the interaction effects were found to be significant, and therefore we could not reject the null hypothesis for any of our nudges.

E. Additional analysis

While our analysis did not show our nudges having the hypothesized effects, it is possible that the nudges interacted with other traits in ways we did not predict. To explore this possibility, we conducted an omnibus test, performing a logistic regression on the full dataset, including all conditions and allowing for every interaction effect. The results of this test can be seen in the appendix.

One robust result of the omnibus model is that personality factors alone do not affect compliance — i.e., on their own, the psychometrics we measured were not significant predictors of people's behavior. However, in the presence of nudges, behavior did vary based on personality in a statistically significant manner.

In addition to the hypothesized interaction between `Statistics` and the GDMS-rational sub-scale, discussed in detail above, the model found interaction effects between the `Expertise` nudge and spontaneous decision-makers, as well as with intuitive and dependent decision-makers in the `Social` condition. While the latter interaction was among those hypothesized, the direction of the effect is reversed: according to the omnibus model, being exposed to the `Social` nudge makes dependent decision-makers *less likely* to adhere to the warning. The `Expertise`/intuitive interaction is likewise counterintuitive: more intuitive decision-makers were likelier to ignore expert advice when exposed to it.

We performed a similar analysis based on the data from the follow-up study and found somewhat different effects. As such, it does not appear that the observed effects can plausibly be attributed to any underlying properties of the decision-making styles.

[8]While we did not expect the recall rate to be different between conditions, analysis using logistic regressions showed that, in fact, it varied significantly, leaving our results vulnerable to post-treatment bias [34].

VI. Discussion

For the most part, the nudges we designed did not appear to elicit varying responses among different personality types—despite being targeted at specific kinds of decision-makers—at least among the sample sizes that we examined and in a way that could be replicated. There are a number of reasons why we may have failed to see an effect.

A. Small sample sizes and small effects

One explanation is the nudges do have an effect, but it was too small to be detected in our sample size. The personality traits we used are stable and well-studied, but there is no effective method for predicting the frequency of specific decisions and outcomes based on the generalized traits. Since effect sizes in psychological literature tend to be small, it is possible that, had we utilized a larger sample, we would have seen an effect.

Indeed, a power analysis we conducted found that, assuming small effect sizes (linear correlation coefficient of 0.1 [14]) and using the conventional power value of 0.8 (and alpha value of 0.05), we would need a sample size of over 781 (for each comparison between the control and a nudge) to obtain significance (compared to only 84 if assuming a medium effect size, e.g., $r = 0.3$).

A small effect size would still represent a promising result: since these warnings are seen by millions of people, a small but consistent improvement in the adherence rate translates to many better decisions every day.

B. Rational rejection

Adhering to HTTPS warnings is, in general, a good decision and good advice for the public, in the absence of additional contextual information: it can help protect one's data and connection from man-in-the-middle attackers. However, in practice, most HTTPS errors are caused by site misconfiguration, rather than active attackers [2]. Furthermore, under certain circumstances, there may be no additional risk to the user by ignoring the warning. For example, in the case of the website our participants were visiting, they had no existing session data with this site and would not be sending it any information. Therefore, visiting it in circumvention of a browser's warning would present no greater risk than loading it over an unsecured HTTP connection—which would be harmless given the lack of sensitive data. Thus, a fully informed and educated visitor could make the rational choice to ignore the warning they saw under these circumstances. A possible explanation for the non-result is that many of our subjects made this decision: they understood that they were not at risk, and therefore made a rational decision to ignore the warnings. This follows Herley's prior work on the rational rejection of certain security advice [25].

On the other hand, our data shows that many people did read the warning and adhered to it, so this theory cannot account entirely for our results.

C. Transparent deception

For similar reasons, the explanation that participants noticed that they were not dealing with a real HTTPS warning

is not entirely satisfactory. While the design of our warnings matched those in the target browsers, there were still discrepancies, which could be spotted by careful inspection (and knowing what to expect). Most prominently, the warning was served from an `http://` page, rather than a URL beginning with `https://`. In Chrome, the simulated warning also lacked a connection security indicator (i.e., a broken lock) in the URL bar. (In Firefox, we were able to simulate one using a favicon.) These differences are relatively minor and are not very likely to be spotted by casual users. In addition, the participants had an open-ended form where they could describe the page they visited and any problems with it, and none noted the fact that the warning was not real.

D. Reactant behavior

If the participants did read our nudge, they could have chosen to ignore it simply as a reaction to the nudge itself (rather than its specific content). In recent work, Jung and Mellers documented that *reactant* individuals (those who are "annoyed or angry when someone else imposes goals on them") reported that they would do the opposite of what a nudge tells them to do, if they find out they are being nudged [28]. A potentially related phenomenon is the skepticism of experts, and a rejection of their advice, found in recent political surveys and events (e.g., [39], [46], [36]). This may explain why the "expert" advice may have had a negative effect on some participants.

E. Heuristic-based decision-making

Even if our participants were not angered or annoyed by a perceived attempt to influence their decision, it is possible that they did not find the core message compelling. Several of our nudges (`Positive Frame`, `Negative Frame`, `Statistics`, `Difficulty`) emphasized "the risk of online fraud" as a potential consequence of not adhering to the warning. Since the participants were not entering personal information, they may have deemed the contents of the nudge irrelevant. But rather than simply ignoring the warning (the "rational" option, as discussed previously), they made their decision based the page's other elements (the primary warning text or the icons) or their past experiences with such warnings.

In fact, past experiences could have dominated our participants' decisions. A well-documented effect in psychology is the distinction between what Daniel Kahneman in *Thinking Fast and Slow* refers to as System I and System II [29]. The latter refers to the cognitive processes that are active when we make careful, deliberate, reasoned, and rational decisions. In contrast, System I makes quick, "gut" decisions on the basis of heuristics and shortcuts. Since deliberate decision-making is mentally taxing, many day-to-day decisions are made automatically by System I, especially when they can fall back to rules and prior experiences. Since HTTPS warnings are a reasonably frequent occurrence, it is plausible that many of our participants rely on heuristics they have intuited and adopted to make a System I decision. Future work can test this hypothesis by presenting the participants with nudges in a manner that induces System II decision-making.

F. Habituation

An extreme form of System I behavior is habituation. If people are repeatedly forced to perform the same action to reach a goal, any decision-making—including even simple rules—may be removed from the process, with the actions becoming an automatic response. This is particularly dangerous in the context of security decisions. Once an action becomes a habit, changing the behavior becomes much harder, and nudges may not suffice. One possible explanation of our results is that HTTPS warnings have reached that stage; the low recall rate observed in our follow-up study, suggesting that most participants did not fully read the warning they saw (insofar as they could not immediately recall the nudge text), may support this hypothesis. It may therefore be the case that people are so thoroughly habituated to HTTPS warning messages, that no amount of nudging will be sufficient to increase compliance, short of completely redesigning the user experience in more drastic ways. Future research may investigate to what extent this is true and look for ways to overcome ingrained bad habits, whether it is clicking through HTTPS warnings or other security decisions.

G. Lessons learned

Since the nudges we developed could not be shown to fulfill their intended goal, our work constitutes a negative result. Such results are generally still valuable to the scientific community, as they aid in avoiding duplicate work, help formulate better hypotheses, and disseminate potentially promising research methods. In our case, the takeaways of our experiment can be useful to researchers wishing to improve security behaviors, beyond simply warning compliance, and study real-world applications of nudges.

Overall, we found our experimental design to be an effective methodology for testing warning designs: people appeared to believe that they were seeing real warnings and behaved accordingly. Ironically, the success of this approach may have undercut our ability to measure whether the nudges themselves worked. With only a third of participants paying close attention to the warning's text, our study may have been capturing people's reactions to the look and feel of existing HTTPS warnings, rather than the new text provided by our nudges. By making the decision more realistic and natural for our participants, the design of the warning became a confound. Prior research has shown that both the content and design of TLS warnings can make a large difference in compliance rates [22]; however, nudges may be too subtle a change for users habituated to seeing these warnings. To significantly improve compliance with warnings, a major redesign may be needed.

Our experience therefore suggests that researchers studying personalization may wish to pursue it in a context where people are attuned to or soliciting new information, rather than following an already established pattern of behavior. Another way forward is testing personalization in application areas where some nudges have already been shown as effective: perhaps certain parts of the population respond to them disproportionately, and their usage can therefore be optimized. Finally, researchers may wish to explore other types of nudges, for example those that directly influence behavior (such as default

settings), as their effects may be more immediate, noticeable, and larger than those of purely informational nudges.

VII. Conclusion

When a security decision is complex or nuanced enough that it cannot be made automatically, software allows its users to make a choice. While providing the freedom to choose is valuable and often inevitable, software designers may wish to guide their users to safer choices. We explored one approach to doing so by nudging users to adhere to HTTPS warnings.

Drawing on literature from psychology, marketing, and behavioral economics, we customized messaging to appeal to certain stable personality traits, such as those measured by the GDMS scale. We tested our nudges in between-subjects experiments, in which participants encountered simulated HTTPS errors while reviewing websites on Mechanical Turk.

Overall, based on the results of our experiments, we cannot conclude that people's responses to the nudges we designed varied by personality trait. Some traits did appear correlated based on the statistical tests we ran—for example, GDMS-rational and the `Statistics` nudge. However, these results were not robust, failing to manifest in a replication experiment and when eliminating participants who could not recall the nudge they had seen.

While our nudges were not obviously effective, personalized messaging is, almost by definition, a promising avenue for research and development: barring unexpected effects, it can likely improve outcomes. Our study contributes a new methodology and raises new questions about the traits to target, the interventions to implement, and the decisions to direct. More importantly, our discussion about the possible reasons why our results do not corroborate the existing decision-making literature may assist other researchers in avoiding potential pitfalls.

Acknowledgment

The authors would like to thank Eyal Peer for his input on the design, development, and analysis of the results of our study, as well as Hana Habib and Ariel Tikotsky for their help in the development of the nudges. This work was supported in part by the Center for Long Term Cybersecurity at the University of California, Berkeley; the National Science Foundation (NSF) under CNS-1528070; and the United States – Israel Binational Science Foundation (BSF) under grant #2014626.

References

[1] A. Acquisti, I. Adjerid, R. H. Balebako, L. Brandimarte, L. F. Cranor, S. Komanduri, P. G. Leon, N. Sadeh, F. Schaub, M. Sleeper *et al.*, "Nudges for privacy and security: Understanding and assisting users' choices online," *SSRN*, 2016. [Online]. Available: https://ssrn.com/abstract=2859227

[2] D. Akhawe, B. Amann, M. Vallentin, and R. Sommer, "Here's my cert, so trust me, maybe?: Understanding TLS errors on the web," in *Proceedings of the 22nd International Conference on World Wide Web*, ser. WWW '13. New York, NY, USA: ACM, 2013, pp. 59–70. [Online]. Available: http://doi.acm.org/10.1145/2488388.2488395

[3] D. Akhawe and A. P. Felt, "Alice in Warningland: A large-scale field study of browser security warning effectiveness," in *Presented as part of the 22nd USENIX Security Symposium (USENIX Security 13)*. Washington, D.C.: USENIX, 2013, pp. 257–272. [Online]. Available: https://www.usenix.org/conference/usenixsecurity13/technical-sessions/presentation/akhawe

[4] H. Almuhimedi, F. Schaub, N. Sadeh, I. Adjerid, A. Acquisti, J. Gluck, L. F. Cranor, and Y. Agarwal, "Your location has been shared 5,398 times!: A field study on mobile app privacy nudging," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: ACM, 2015, pp. 787–796. [Online]. Available: http://doi.acm.org/10.1145/2702123.2702210

[5] B. Anderson, T. Vance, B. Kirwan, D. Eargle, and S. Howard, "Users arent (necessarily) lazy: using NeuroIS to explain habituation to security warnings," in *Thirty Fifth International Conference on Information Systems*, 2014.

[6] B. B. Anderson, C. B. Kirwan, J. L. Jenkins, D. Eargle, S. Howard, and A. Vance, "How polymorphic warnings reduce habituation in the brain: Insights from an fMRI study," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: ACM, 2015, pp. 2883–2892. [Online]. Available: http://doi.acm.org/10.1145/2702123.2702322

[7] R. Balebako, P. G. Leon, H. Almuhimedi, P. G. Kelley, J. Mugan, A. Acquisti, L. F. Cranor, and N. Sadeh, "Nudging users towards privacy on mobile devices," in *Proc. CHI 2011 Workshop on Persuasion, Nudge, Influence and Coercion*, 2011.

[8] C. Bartneck, A. Duenser, E. Moltchanova, and K. Zawieska, "Comparing the similarity of responses received from studies in Amazon's Mechanical Turk to studies conducted online and with direct recruitment," *PLOS ONE*, vol. 10, no. 4, pp. 1–23, 04 2015.

[9] T. S. Behrend, D. J. Sharek, A. W. Meade, and E. N. Wiebe, "The viability of crowdsourcing for survey research," *Behavior Research Methods*, vol. 43, no. 3, p. 800, 2011. [Online]. Available: http://dx.doi.org/10.3758/s13428-011-0081-0

[10] A. Besmer, J. Watson, and H. R. Lipford, "The impact of social navigation on privacy policy configuration," in *Proceedings of the Sixth Symposium on Usable Privacy and Security*, ser. SOUPS '10. New York, NY, USA: ACM, 2010, pp. 7:1–7:10. [Online]. Available: http://doi.acm.org/10.1145/1837110.1837120

[11] J. T. Cacioppo and R. E. Petty, "The need for cognition." *Journal of personality and social psychology*, vol. 42, no. 1, p. 116, 1982.

[12] J. T. Cacioppo, R. E. Petty, and C. F. Kao, "The efficient assessment of need for cognition," *Journal of Personality Assessment*, vol. 48, no. 3, pp. 306–307, 1984, pMID: 16367530.

[13] E. K. Choe, J. Jung, B. Lee, and K. Fisher, "Nudging people away from privacy-invasive mobile apps through visual framing," in *Human-Computer Interaction – INTERACT 2013*, P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, and M. Winckler, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 74–91.

[14] J. Cohen, *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum, 1988.

[15] 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*, ser. CHI '08. New York, NY, USA: ACM, 2008, pp. 1065–1074. [Online]. Available: http://doi.acm.org/10.1145/1357054.1357219

[16] S. Egelman, A. P. Felt, and D. Wagner, "Choice architecture and smartphone privacy: There's a price for that," in *The Economics of Information Security and Privacy*, R. Böhme, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 211–236.

[17] S. Egelman and E. Peer, "The myth of the average user: Improving privacy and security systems through individualization," in *Proceedings of the 2015 New Security Paradigms Workshop*. ACM, 2015, pp. 16–28.

[18] ——, "Predicting privacy and security attitudes," *SIGCAS Comput. Soc.*, vol. 45, no. 1, pp. 22–28, Feb. 2015. [Online]. Available: http://doi.acm.org/10.1145/2738210.2738215

[19] ——, "Scaling the security wall: Developing a security behavior intentions scale (SeBIS)," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15.

New York, NY, USA: ACM, 2015, pp. 2873–2882. [Online]. Available: http://doi.acm.org/10.1145/2702123.2702249

[20] A. P. Felt, E. Ha, S. Egelman, A. Haney, E. Chin, and D. Wagner, "Android permissions: User attention, comprehension, and behavior," in *Proceedings of the Eighth Symposium on Usable Privacy and Security*, ser. SOUPS '12. New York, NY, USA: ACM, 2012, pp. 3:1–3:14. [Online]. Available: http://doi.acm.org/10.1145/2335356.2335360

[21] A. P. Felt, R. W. Reeder, A. Ainslie, H. Harris, M. Walker, C. Thompson, M. E. Acer, E. Morant, and S. Consolvo, "Rethinking connection security indicators," in *Twelfth Symposium on Usable Privacy and Security (SOUPS)*. Denver, CO: USENIX Association, Jun. 2016, pp. 1–14. [Online]. Available: https://www.usenix.org/conference/soups2016/technical-sessions/presentation/porter-felt

[22] A. P. Felt, R. W. Reeder, H. Almuhimedi, and S. Consolvo, "Experimenting at scale with Google Chrome's SSL warning," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014, pp. 2667–2670. [Online]. Available: http://doi.acm.org/10.1145/2556288.2557292

[23] T. Halevi, J. Lewis, and N. Memon, "A pilot study of cyber security and privacy related behavior and personality traits," in *Proceedings of the 22nd International Conference on World Wide Web*, ser. WWW '13 Companion. New York, NY, USA: ACM, 2013, pp. 737–744. [Online]. Available: http://doi.acm.org/10.1145/2487788.2488034

[24] M. Harbach, M. Hettig, S. Weber, and M. Smith, "Using personal examples to improve risk communication for security & privacy decisions," in *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014, pp. 2647–2656. [Online]. Available: http://doi.acm.org/10.1145/2556288.2556978

[25] C. Herley, "So Long, and No Thanks for The Externalities: The rational rejection of security advice by users," in *NSPW '09: Proceedings of The 2009 New Security Paradigms Workshop*. New York, NY, USA: ACM, 2009, pp. 133–144.

[26] Q. Ismail, T. Ahmed, A. Kapadia, and M. K. Reiter, "Crowdsourced exploration of security configurations," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: ACM, 2015, pp. 467–476. [Online]. Available: http://doi.acm.org/10.1145/2702123.2702370

[27] O. P. John and S. Srivastava, "The big five trait taxonomy: History, measurement, and theoretical perspectives," *Handbook of personality: Theory and research*, vol. 2, no. 1999, pp. 102–138, 1999.

[28] J. Y. Jung and B. A. Mellers, "American attitudes toward nudges," *Judgment and Decision Making*, vol. 11, no. 1, p. 62, 2016.

[29] D. Kahneman, *Thinking, fast and slow*. Macmillan, 2011.

[30] P. G. Kelley, "Conducting usable privacy & security studies with Amazon's Mechanical Turk," in *Symposium on Usable Privacy and Security (SOUPS) (Redmond, WA)*, 2010.

[31] K. Krol, M. Moroz, and M. A. Sasse, "Don't work. can't work? why it's time to rethink security warnings," in *2012 7th International Conference on Risks and Security of Internet and Systems (CRiSIS)*, Oct 2012, pp. 1–8.

[32] B. Liu, M. S. Andersen, F. Schaub, H. Almuhimedi, S. A. Zhang, N. Sadeh, Y. Agarwal, and A. Acquisti, "Follow my recommendations: A personalized privacy assistant for mobile app permissions," in *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. Denver, CO: USENIX Association, Jun. 2016, pp. 27–41. [Online]. Available: https://www.usenix.org/conference/soups2016/technical-sessions/presentation/liu

[33] T. Minkus and N. Memon, "Leveraging personalization to facilitate privacy," *Available at SSRN 2448026*, 2014. [Online]. Available: https://ssrn.com/abstract=2448026

[34] J. M. Montgomery, B. Nyhan, and M. Torres, "How conditioning on post-treatment variables can ruin your experiment and what to do about it," 2016.

[35] S. G. Scott and R. A. Bruce, "Decision-making style: The development and assessment of a new measure," *Educational and psychological measurement*, vol. 55, no. 5, pp. 818–831, 1995.

[36] J. Shaw, "The real reason that we don't trust experts anymore," http://www.independent.co.uk/voices/the-real-reason-that-we-don-t-trust-experts-a7126536.html, July 8 2016.

[37] D. J. Simons and C. F. Chabris, "Common (mis)beliefs about memory: A replication and comparison of telephone and Mechanical Turk survey methods," *PLOS ONE*, vol. 7, no. 12, pp. 1–5, 12 2012.

[38] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F. Cranor, "Crying wolf: An empirical study of SSL warning effectiveness." in *USENIX Security Symposium*, 2009, pp. 399–416.

[39] J. Tamny, "A revolt against the 'experts' is not necessarily an endorsement of Donald Trump," http://www.forbes.com/sites/johntamny/2016/10/22/a-revolt-against-the-experts-is-not-necessarily-an-endorsement-of-donald-trump/, October 22 2016.

[40] R. H. Thaler and C. R. Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness.* Yale University Press, 2008.

[41] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.

[42] S. Uebelacker and S. Quiel, "The social engineering personality framework," in *2014 Workshop on Socio-Technical Aspects in Security and Trust*, July 2014, pp. 24–30.

[43] Y. Wang, P. G. Leon, A. Acquisti, L. F. Cranor, A. Forget, and N. Sadeh, "A field trial of privacy nudges for Facebook," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014, pp. 2367–2376. [Online]. Available: http://doi.acm.org/10.1145/2556288.2557413

[44] J. Weinberger and A. P. Felt, "A week to remember: The impact of browser warning storage policies," in *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. Denver, CO: USENIX Association, Jun. 2016, pp. 15–25. [Online]. Available: https://www.usenix.org/conference/soups2016/technical-sessions/presentation/weinberger

[45] P. J. Wisniewski, B. P. Knijnenburg, and H. R. Lipford, "Making privacy personal: Profiling social network users to inform privacy education and nudging," *International Journal of Human-Computer Studies*, vol. 98, pp. 95 – 108, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1071581916301185

[46] G. Witte, "9 out of 10 experts agree: Britain doesn't trust the experts on Brexit," https://www.washingtonpost.com/world/europe/9-out-of-10-experts-agree-britain-doesnt-trust-the-experts-on-brexit/2016/06/21/2ccc134a-34a6-11e6-ab9d-1da2b0f24f93_story.html, June 23 2016.

[47] J. Xie, B. P. Knijnenburg, and H. Jin, "Location sharing privacy preference: Analysis and personalized recommendation," in *Proceedings of the 19th International Conference on Intelligent User Interfaces*, ser. IUI '14. New York, NY, USA: ACM, 2014, pp. 189–198. [Online]. Available: http://doi.acm.org/10.1145/2557500.2557504

[48] Y. Zhao, J. Ye, and T. Henderson, "Privacy-aware location privacy preference recommendations," in *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, ser. MOBIQUITOUS '14. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014, pp. 120–129. [Online]. Available: http://dx.doi.org/10.4108/icst.mobiquitous.2014.258017

APPENDIX

Omnibus regression results

Variable	Estimate	Std. Error	z value	$Pr(>\|z\|)$	
(Intercept)	-9.414e-02	3.616e+00	-0.026	0.97923	
conditionDifficulty	4.521e+00	4.709e+00	0.960	0.33698	
conditionExpertise	-4.504e+00	4.877e+00	-0.923	0.35582	
conditionNegativeFrame	-3.211e+00	5.148e+00	-0.624	0.53287	
conditionPositiveFrame	-3.827e+00	4.858e+00	-0.788	0.43086	
conditionSocial	1.307e+00	5.082e+00	0.257	0.79708	
conditionStatistics	8.632e-01	4.808e+00	0.180	0.85752	
browserEdge	-1.373e+00	1.115e+00	-1.231	0.21816	
browserFirefox	1.065e+00	6.691e-01	1.592	0.11150	
browserFirefoxNew	-3.422e-01	2.543e-01	-1.345	0.17849	
browserInternetExplorer	-1.645e+01	1.115e+03	-0.015	0.98823	
browserOther	-1.659e+01	7.107e+02	-0.023	0.98137	
avoidant	2.035e-02	3.117e-01	0.065	0.94794	
dependent	-7.207e-01	4.030e-01	-1.788	0.07372	.
intuitive	4.048e-02	4.356e-01	0.093	0.92597	
rational	5.930e-01	6.222e-01	0.953	0.34058	
spontaneous	-7.285e-01	6.054e-01	-1.203	0.22883	
nfc	2.031e-01	3.913e-01	0.519	0.60372	
conditionDifficulty:rational	-1.503e+00	8.220e-01	-1.828	0.06748	.
conditionExpertise:rational	-1.005e-03	8.703e-01	-0.001	0.99908	
conditionNegativeFrame:rational	-3.993e-01	8.649e-01	-0.462	0.64431	
conditionPositiveFrame:rational	-4.116e-01	8.265e-01	-0.498	0.61846	
conditionSocial:rational	-1.004e+00	8.669e-01	-1.158	0.24688	
conditionStatistics:rational	-1.858e+00	8.777e-01	-2.117	0.03422	*
conditionDifficulty:avoidant	-4.280e-01	4.446e-01	-0.963	0.33574	
conditionExpertise:avoidant	-9.151e-02	4.501e-01	-0.203	0.83888	
conditionNegativeFrame:avoidant	1.840e-01	4.506e-01	0.408	0.68303	
conditionPositiveFrame:avoidant	3.955e-01	4.984e-01	0.794	0.42740	
conditionSocial:avoidant	-8.760e-02	4.913e-01	-0.178	0.85848	
conditionStatistics:avoidant	6.333e-01	4.511e-01	1.404	0.16037	
conditionDifficulty:dependent	8.383e-01	5.232e-01	1.602	0.10908	
conditionExpertise:dependent	4.218e-01	5.212e-01	0.809	0.41839	
conditionNegativeFrame:dependent	1.193e+00	6.644e-01	1.796	0.07257	.
conditionPositiveFrame:dependent	2.693e-01	5.930e-01	0.454	0.64973	
conditionSocial:dependent	1.597e+00	6.126e-01	2.607	0.00915	**
conditionStatistics:dependent	4.764e-01	5.939e-01	0.802	0.42249	
conditionDifficulty:intuitive	-2.182e-01	5.890e-01	-0.370	0.71104	
conditionExpertise:intuitive	-4.745e-01	5.878e-01	-0.807	0.41952	
conditionNegativeFrame:intuitive	-4.633e-01	6.646e-01	-0.697	0.48575	
conditionPositiveFrame:intuitive	-1.141e-01	6.306e-01	-0.181	0.85636	
conditionSocial:intuitive	-1.726e+00	6.910e-01	-2.497	0.01251	*
conditionStatistics:intuitive	2.443e-01	5.850e-01	0.418	0.67622	
conditionDifficulty:spontaneous	5.333e-01	7.671e-01	0.695	0.48697	
conditionExpertise:spontaneous	2.338e+00	7.627e-01	3.066	0.00217	**
conditionNegativeFrame:spontaneous	7.371e-01	8.369e-01	0.881	0.37843	
conditionPositiveFrame:spontaneous	1.459e+00	8.013e-01	1.821	0.06861	.
conditionSocial:spontaneous	1.362e+00	8.231e-01	1.655	0.09795	.
conditionStatistics:spontaneous	4.603e-01	7.454e-01	0.617	0.53691	
conditionDifficulty:nfc	-1.581e-01	5.299e-01	-0.298	0.76539	
conditionExpertise:nfc	-1.662e-01	5.387e-01	-0.309	0.75763	
conditionNegativeFrame:nfc	-1.574e-01	5.293e-01	-0.297	0.76617	
conditionPositiveFrame:nfc	-9.779e-03	5.892e-01	-0.017	0.98676	
conditionSocial:nfc	-1.390e-01	5.370e-01	-0.259	0.79572	
conditionStatistics:nfc	4.057e-01	5.673e-01	0.715	0.47448	