# Control versus Effort in Privacy Warnings for Webforms

Kat Krol
University College London
Gower Street
London, WC1E 6BT, UK
k.krol@cs.ucl.ac.uk

Sören Preibusch
formerly Microsoft Research
21 Station Road
Cambridge, CB1 2FB, UK
mail@soeren-preibusch.de

## ABSTRACT

Webforms are the primary way of collecting information online. However, some users may wish to limit the amount of personal information they provide and only fill out the minimum required for the transaction. With less than one third of websites marking fields as mandatory or optional, limiting disclosure can be a daunting task. This paper reports on a large behavioural online experiment on user reactions to warnings alerting them that they are about to submit non-mandatory information. Eight warning dialogues were tested between 4,620 participants. We found that warnings mentioning security or privacy threats both significantly reduced the disclosure of personal information in the webforms used (e.g., -27 percentage points for date of birth). The most actionable warning was not the one that minimised user effort but the one that left participants most in control. We consider our study useful to establish what kind of warning messages could help users manage their privacy. In order not to contribute to the ever increasing warning fatigue, a good real-world implementation of over-disclosure indicators would be for the browser to provide users with real-time information on mandatoriness/optionality when the webform loads, for example by highlighting optional fields.

## CCS Concepts

**•Security and privacy → Usability in security and privacy;**

## Keywords

privacy; security; usability; warnings; disclosure; webforms

## 1. INTRODUCTION

Webforms were introduced into the HTML standard over twenty years ago [3] and are now part of our daily Web browsing routine. The form is the primary mechanism for collecting personal information from users, who type in personal details such as their name, date of birth, or address to complete online transactions. On the one hand, the form often gets in the users' way of successfully completing a task. On the other hand, we see that users do not keep their efforts in dealing with webforms to a minimum. We use the term 'over-disclosure' to describe users providing more personal details than required. Our previous study [16] showed that participants spent more effort and time on webforms than required as they knowingly provided optional details at 3.5 seconds per field (average time to completion, excluding initial setup time). We quantified the prevalence of over-disclosure at 57% to 87% for data items such as date of birth or favourite colour respectively.

More prevalent than voluntary over-disclosure is accidental over-disclosure: the onus is on the users to tell mandatory and optional fields apart and this can be a difficult task. When surveying 140 websites, Preibusch and Bonneau [15] found that less than one third provided visual or textual indicators which of the webform fields were mandatory. However, in a previous study we provided evidence that users make clear distinctions between mandatory and optional fields and selectively decide which fields to leave blank, if possible [16]. On a form which features a mix of mandatory and optional fields, the latter see significantly lower disclosure rates compared to the former and compared to a form featuring no mandatory fields. This might be an indication that if users knew what was required, they would only fill out fields that were mandatory and in this way limit disclosure and protect their privacy.

In the absence of visual or other hints regarding the mandatoriness of input fields, identifying the minimum set of data items can cost time and effort. It would require a user to leave a field blank and attempt to submit the form to see if the form is accepted despite the field being empty. Depending on the server logic behind the form, this would need to be done with every combination of fields. Additionally, after an unsuccessful submission, previously entered data might become erased and would require re-entry.

Where manual checks of mandatoriness are difficult, users could be assisted by a technological solution. HTML5 introduced advanced webform mark-up capabilities that include an attribute for form fields to indicate whether they are optional or required [22]. In addition, browsers could examine and interpret existing visual cues. Web browsers have access to a combination of signals to sense the mandatoriness of fields and advise the user accordingly.

This paper reports on the findings of an experiment conducted on the Amazon Mechanical Turk (mTurk) platform to assess the effectiveness of such privacy-enhancing tool



*WPES'16, October 24 2016, Vienna, Austria*



DOI: http://dx.doi.org/10.1145/2994620.2994640

support and to explore the design space in a search towards most actionable user interfaces. We trialled a range of eight warnings alerting users that they have completed some non-mandatory fields. Our results show that the warnings successfully allowed participants to limit the submission of optional, personal information on a webform. Interestingly, participants preferred not the warning that required minimal effort from them by suppressing all optional data they had entered, but the one that left them most in control over which information to submit, and allowed manual re-configuration.

The remainder of the paper is organised as follows. First, we provide the necessary background on warnings and motivate our decision for their use in the study. Then, we formulate our hypotheses and describe the design of the study, followed by a presentation of the results and their discussion. Finally, we describe the limitations to our study and suggest avenues for future research and engineering.

# 2. BACKGROUND

In this section, we describe the existing literature on warnings and indicators, starting with security, for which there is a larger stream of empirical research.

## 2.1 Passive versus active security indicators

Security indicators can be divided into two groups: passive and active. On the one hand, passive indicators are adjacent to the users' task, rather than inside the task. Passive indicators do not get in users' way, at the expense of being overlooked at times. For instance, a passive indicator for a secure connection can be placed on the margins of a website highlighted through colour. On the other hand, active indicators seek to attract the user's attention by interrupting them in their primary task (online banking, shopping, etc.) and the user has to actively interact with the indicator, as they cannot continue with their primary task unless the indicator is acknowledged.

The classic passive indicator is the padlock icon indicating an SSL connection. Studies have shown that only few users notice it [6, 23]. In a study by Schechter et al. [19], participants were asked to perform some banking tasks and were then presented with increasingly alarming cues that the connection was compromised. The researchers gradually removed HTTPS indicators, the participant's authentication image and finally replaced the login page with a warning page. Passive indicators were not effective in preventing users from logging in to a phishing site—no participant refused to enter their password when the HTTPS indicators were absent and only 8% did so when their chosen authentication image was absent. However, when a security warning page replaced the login page 47% of participants refused to enter their credentials.

In their study on phishing, Wu et al. [23] showed that participants could be tricked into submitting personal information 34% of the time. Instructing them to focus on the toolbars in the browser did not make much difference since participants reported assessing the legitimacy of the website by how it looked and felt. That is, the locus of attention and strongest cue was the page itself, not the chrome surrounding it. Participants appeared to be immersed in what they were doing; 45% of them stressed they did not pay attention to the toolbar because they wanted to finish their task. The authors' follow-up study showed that pop-up warnings that interrupted users were far more effective than passive warnings shown in the toolbar. These pop-ups were displayed over the webpage, that is in the area of the screen participants were already focusing on.

Egelman et al. [6] juxtaposed passive and active browser phishing warnings. Their results corroborate findings from previous studies. Of those confronted with the active warnings, 79% closed the phishing website but only 13% participants did so when the passive warnings were used. They found that passive indicators were not significantly different from not providing any warning at all.

In a large-scale field study of SSL warnings, Akhawe and Felt [1] used telemetry to investigate user warning behaviour (e.g., click-through rates). They found that users hardly every click on explanatory links and thus, it is important to provide users with sufficient information in the main text of the warning.

## 2.2 Privacy indicators

There is a number of privacy indicators available. However, they rarely come from the browser and are more often add-ons or plug-ins created by researchers or open-source developers. Privacy indicators tend to be passive: for example, the Privacy Bird® [17] changes the colour and the content of its speech bubble depending on whether the website's privacy policy matches the user's set privacy preferences.

There are also privacy indicators around cookies, the most popular of them is Ghostery [7]. Ghostery displays a list of companies that are tracking the user when they are visiting a particular website. This browser add-on provides the user with information on the type of data being collected and offers the option to block certain cookies. A purple bubble with a list of trackers in the corner of the screen, Ghostery is a passive privacy indicator.

Our study addresses this paucity of empirical investigations into privacy indicators and the lack of studies looking at privacy warnings alerting users they completed non-mandatory information. Although both warnings are aimed at the threats arising from cyberspace interactions, one fundamental difference between privacy and security warnings is the spectrum or duality of intended outcomes. While there are vertical security preferences (everyone prefers more security), privacy indicators must cater for horizontal privacy preferences, as individuals do not unanimously prefer revealing less data about themselves. We further explore these issues in Section 2.3.1.

## 2.3 Criticisms of warnings and our defence

Although active warnings are far more effective in alerting users and making them change their course of action, there is also a cost associated with them. For instance, most certificate warnings appear to be false positives [8]. Users heeding a certificate warning would not benefit from added security as there was no threat in the first place. But more importantly, heeding security advice often means users are prevented from completing their primary task. For instance, checking URLs is time-consuming, while attacks are rare. For that reason, Herley stresses that ignoring security advice is rational from an economic point of view [8] as human memory and attention are a finite resource [9].

The use of warnings can also be desensitising since studies have shown that users are habituated to warnings and ignore them. Krol et al. [11] showed that users disregard warnings if

they have the impression that they appear indiscriminately for all the items they attempt to download.

Despite this overwhelming evidence against warnings, we decided to implement our over-disclosure warnings as active pop-up windows that interrupt the primary task since otherwise they might not have been noticed by the participants. We took a 'fail early' approach—to know whether an over-disclosure indicator works, it is better to test it in its active form that would be noticed. If it is effective and receives user acceptance, then it should be explored if and in what form it could be implemented as a less disruptive indicator.

#### 2.3.1 Success metric

*Security* warnings may be dismissed because they are seen as false alarms but an accurate security warning applies to all users: no user should input their details into a phishing site for instance. For *privacy*, the situation is rather different: even an accurate privacy warning may still be dismissible depending on the user's specific privacy preferences. In other words, security is a yes or no, privacy is a spectrum. While in security users cannot partially get infected with a virus, in privacy they can disclose some personal information but withhold other. The users' risk analysis is more nuanced and the warning design must have a built-in feature allowing them not to follow the warning's privacy-enhancing advice.

Our study therefore adopts a success metric that does not presuppose universal privacy preferences. We consider the warnings successful not only if participants deleted previously entered information. Instead, we have two different criteria. (1) We consider the warnings effective if users returned to the form and revisited the information entered. We assess the warning's actionability by the click-through rates generated by the privacy-enhancing response option on the warning. (2) We also look at post-experimental ratings for usability to establish if users considered the warnings helpful, easy to understand and relevant.

## 3. RESEARCH HYPOTHESES

The study aims to identify the characteristics of actionable privacy alerts, focusing in particular on the wording of the warning messages, the different options offered to the user, and the characteristics of the audience. Eight hypotheses guided the analysis of the results.

### 3.1 Warning text

Warnings currently deployed in browsers typically feature a short message that mentions the diagnosis, why the warning was displayed, and optionally explains the threat (Fig. 1).

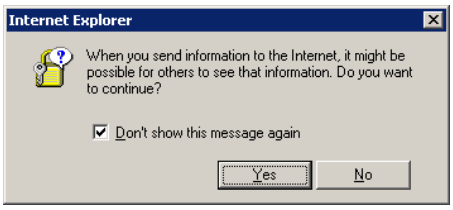


**Figure 1: Screenshot of an existing privacy warning in Internet Explorer.**

As discussed earlier, security events are associated with a more definite outcome than privacy events. In consequence, we hypothesise that users who will see a security warning will be more likely to delete information than those who will see a privacy warning. We further hypothesise that the mention of security and privacy will have a stronger impact on the user than not providing a reason for the warning.

H1a: Users presented with warnings mentioning privacy and security threats will delete previously entered data.

H1b: Users presented with a warning mentioning security will be more likely to delete previously entered data than those presented with a warning mentioning privacy.

H1c: Users presented with a warning which has no accompanying explanation will be less likely to delete previously entered information than those presented with a warning that mentions a privacy or security threat.

### 3.2 Item sensitivity

We hypothesise that the impact of the warning will depend on item sensitivity. Users who are asked to provide personal data on a webform can restrict its proliferation by either leaving the fields blank or by deleting previously entered data. Data items with low sensitivity *do not need* to be deleted; data items with high sensitivity *cannot* be deleted if users did not provide them in the first place. One can therefore predict a U-shaped relationship between data item sensitivity and its likelihood of deletion after a privacy warning.

H2: Users will be more likely to remove sensitive information while keeping the less sensitive information in place (U-shaped relationship).

### 3.3 Time and effort

Filling out a form requires an amount of work, and humans tend to be protective of what they have produced. Therefore, we hypothesise that the more time the participant spent on filling out the form, the less likely they will be to remove their entries and let all this work and time invested be for nothing. This reflects the sunk cost fallacy and the escalation of commitment [20].

H3a: The more time users spent on filling out the form, the less likely they would be to remove the information entered.

Furthermore, as current security advice instructs the user to perform time- and effort-consuming checks, tool support should be better integrated with security mechanisms to take some of the effort off the user. We therefore hypothesise that a warning that directly provides the options to remove non-mandatory information for the user will be preferred.

H3b: Users will prefer options that minimise effort and save time over other options.

### 3.4 Demographic characteristics

We hypothesise that the decision to go back and delete some information after having seen a warning could be associated with users' computer literacy as the level of knowledge might help them more accurately assess the risk linked to disclosing personal information online. We hypothesise that previous experience with viruses, fraud and scam is associated with the users' willingness to provide the requested information.

H4a: Users will differ in their deletion/altering behaviour depending on their levels of computer literacy.

H4b: Users will differ in their deletion/altering behaviour depending on their cyberthreat exposure.

# 4. METHODOLOGY

The study consisted of a behavioural experiment and a follow-up questionnaire.

## 4.1 Experiment

The webform and all instructions used were directly replicated from the preceding study [16].

Upon attempting to submit the webform, a warning was shown if the participant had filled in any non-mandatory fields (Fig. 2). No warning was displayed unless at least one optional field was filled in. The warning text typically consisted of two sentences—the statement saying the user has completed some non-mandatory fields and the mention that this can be a threat to their security or privacy.


**Figure 2: An example of a warning used in the study (condition ws1ho).**

The study was a between-subjects design. As presented as a task on mTurk, the form included two check questions that required reading and understanding the instructions. Payment of $0.50 was unconditional of participants' answers to the check questions. In the following, those who answered both check questions correctly are considered as having understood the instructions including the optionality of the form fields. The other participants who answered at least one check question incorrectly or neither are classified as not having read and/or understood the instructions.

Eight types of warnings were used in the study. They can be divided into two groups: (1) four manual warnings varying the message and (2) four one-click warnings varying the options on the buttons. Table 1 provides an overview of the warnings used in all eight conditions. The first warning (WS) mentioned security, while the second privacy (WP) as the explanation for why it was being shown. The warning in the WNX condition (no explanation) did not provide an explanation and only stated: "You have completed some non-mandatory fields". The warning in the WO condition (o as in 'orthography') served as a control: it stated that there may be spelling mistakes in some fields and gave the explanation that it can be difficult for others to read a text that contains spelling mistakes.

We implemented a control condition that displayed a warning instead of removing the intervention altogether. The rationale behind it was to confront participants with the stimulus warning to see if it impacted their altering behaviour. The idea is to benchmark against the absence of a privacy/security warning rather than the absence of a warning at all. The baseline of undisturbed disclosure rates is

| **Conditions and warning texts** |
|---|
| *Security warning* (WS):<br>You have completed some non-mandatory fields.<br>Revealing too much about yourself online can be a threat to the security of your personal information. |
| *Privacy warning* (WP):<br>You have completed some non-mandatory fields.<br>Revealing too much about yourself online can be a threat to the privacy of your personal information. |
| *Warning with no explanation* (WNX):<br>You have completed some non-mandatory fields. |
| *Orthography warning* (control condition, WO):<br>There may be spelling mistakes in some fields. It can be difficult for other people to read a text with spelling mistakes. |
| *Security warning: highlights all optional* (WS1HO):<br>Highlight optional fields |
| *Security warning: highlights filled optional* (WS1HFO):<br>Highlight optional fields |
| *Security warning: submits only mandatory* (WS1SM):<br>Submit only mandatory |
| *Security warning: clears optional* (WS1CO):<br>Clear optional fields |

**Table 1: Overview of the treatments: over-disclosure warnings used in the study.**

available from the preceding study or by observing participants' behaviour in this study, before they saw the warning.

A spelling warning is well-suited for the following reasons. First, spelling errors are a real and credible "threat". They are potentially embarrassing but are not related to security or privacy. Second, it is a plausible pop-up since spelling is something where tool support is already widely implemented. Third, manual investigation is needed from the user: similarly to the security and privacy warnings, the participants needed to go back and inspect each field.

For the one-click warnings, the same warning text was used as in the security warning (condition WS). Again, there were two buttons for each warning. The first one always said "Submit anyway" while the second one was subject to our manipulation. The four phrases used for the second button can be subdivided into two categories: highlighting and manipulating. For highlighting, "Highlight filled optional" was juxtaposed with "Highlight all optional". For the manipulating ones, "Submit only mandatory" was juxtaposed with "Clear optional fields".

When confronted with the warning, users had to make a choice, they could not proceed unless either of these two buttons was clicked. The warning in the control condition was structured in the same as in the conditions WS and WP, they all provided a diagnosis and an explanation, totalling three lines of text.

## 4.2 Follow-up questionnaire

At least one day after completing the experiment, participants were sent an invitation to fill out a feedback questionnaire for which they received an additional payment of $1.50. A reminder was sent to those who had not completed the follow-up questionnaire.

The questionnaire started by reminding the participant of the original webform with a screenshot. There was no option to enlarge it so participants could not read the original

instructions or questions. The aim was to make participants recall the original experiment rather than base their answers on what they saw.

The questions in the follow-up questionnaire can be divided into four groups. (1) The first nine questions related to the survey asking if the participant enjoyed the original task of completing the webform, how long they spent filling it out, if they had revealed any personal or sensitive data in it, and what they thought the intended purpose of the survey was. (2) Two questions related to participants' computer literacy (8 items) and cyberthreat exposure (14 items). (3) A further ten questions focused on the warning. (4) Four questions captured demographic information.

The follow-up questionnaire was completed by 87% of the original participant population. For 3,203 participants, there were the complete records of form filling behaviour plus feedback data. For a random subset of the participants, the link could not be established with the first phase.

## 4.3 Participants

Participants were workers on the mTurk platform, all recruited to be based in the United States. According to the data provided in the follow-up questionnaire, the majority of the participants were aged 18 to 24 years (36%), 25 to 29 (23%) and 30 to 39 (21%). 16% of participants were aged 40 or older; less than 4% refused to indicate their age bracket. Male participants made up 53% of the sample. Regarding formal education, 28% had completed one or more years of college and an additional 7% of participants had completed their college with an associate degree (for example, AA, AS). 30% had a bachelor's degree (for example, BA, AB, BS), and an additional 7% had a master's degree.

# 5. RESULTS

## 5.1 General results

Findings from the preceding study [16] as to the level of disclosure were corroborated. Amongst participants who were aware that none of these details were required, all personal details were disclosed by at least three quarters of participants, with the exception of date of birth, which was still revealed by more than two thirds of participants. Favourite colour was the data item volunteered most often (84%). First name occupied rank five when ordering data items by disclosure rate. In the following analysis, date of birth, first name and favourite colour are considered as representative for data items of high, medium and low sensitivity, in line with extant literature that takes disclosure rates as a proxy for sensitivity [14].

Interestingly, participants who had not read the instructions and who were thus unaware that disclosure was voluntary (accidental over-disclosure), shared more personal details with high statistical difference ($p < 0.0001$, two-tailed t-test). The average number of fields completed by participants who had not read the instructions was 9.1 and thus 1.0 items more on average. Amongst the unaware, 80% fully filled the form with all ten items of personal data compared to 64% amongst those who had read the instructions. Similarly, a blank form was submitted less often by the former than by the latter (6% versus 10%). Consequently, disclosure rates are much higher when participants ignore the optionality of the data requirements, with all of them being above 90% except date of birth with 83% (Fig. 3).

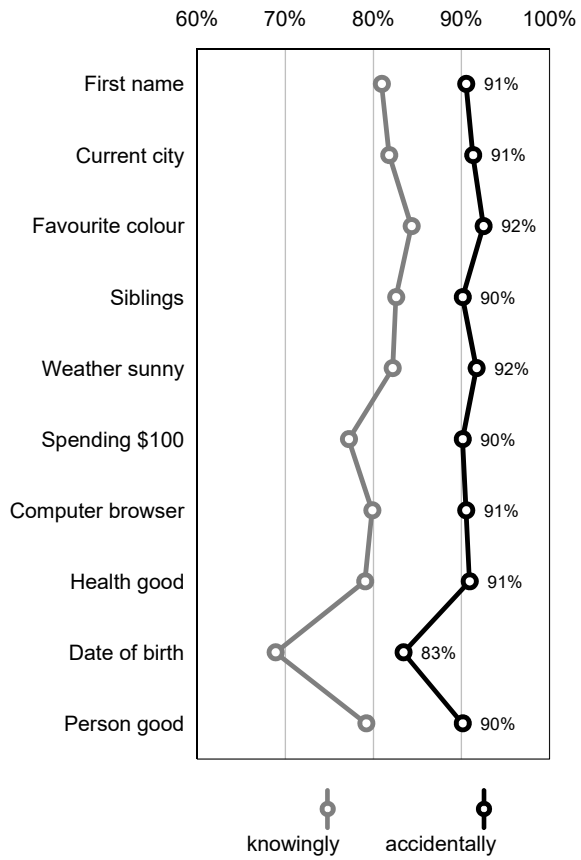


**Figure 3: Initial disclosure rates for each of the ten items of personal information collected on the form. Knowing disclosure applies to participants who had read the instructions and correctly answered the check questions; accidental disclosure applies to the other participants.**

## 5.2 Hypothesis testing

For simplification, the following analysis is mainly based on three data items to cover the whole sensitivity spectrum. The sensitivity of these items was established based on three sources. Exogenously, it was based on the results from two different previously conducted experiments which requested them and then asked participants to rate them in post-experiment questionnaires [16, 13]. Endogenously, the initial (i.e., pre-warning) completion rates are considered from this study, as outlined earlier in this section. Based on these, date of birth (DOB) was identified as the high-sensitivity data item as it had the highest proportion (19%) of participants who left it blank (69% completion rate, ranked lowest), first name as the medium-sensitivity item (81% completion rate, ranked fifth out of ten) and favourite colour as the low-sensitivity one (84% completion rate, ranked highest).

### H1a: Mentioning privacy and security

We hypothesised that warnings evoking privacy and security threats will make users delete previously entered data. When comparing the deletion ratios for the security and privacy warnings with the control condition (WO), for example

for date of birth: the deletion ratio was 26% for WS, 25% for WP but only 4% for WO. The comparison for the three items of differing sensitivity showed that warnings evoking security and privacy threats made users delete previously entered data ($p < 0.0001$ at least for all items, G-test of independence). H1a is therefore supported.

## H1b: Security vs. privacy

We hypothesised that those confronted with a warning mentioning security will be more likely to delete previously entered data than those confronted with a warning mentioning privacy. In the post-experiment questionnaire, participants indicated the security warning as being more severe than the privacy warning (rated 4.6 versus 4.4 out of 7, $p = 0.03$, two-tailed t-test). The distribution of severity ratings is shown in Figure 4. Looking at the behavioural evidence, the deletion ratios for the security warning (WS) were 27% for DOB, 26% for first name, and 14% for favourite colour. For the privacy warning, deletion rates were very similar (Fig. 5). We hypothesised that the security warning could have a stronger impact on the users than the privacy one. However, this difference was not statistically significant for any of the three levels of sensitivity: high ($p = 0.92$), medium ($p = 0.50$) and low ($p = 0.78$).

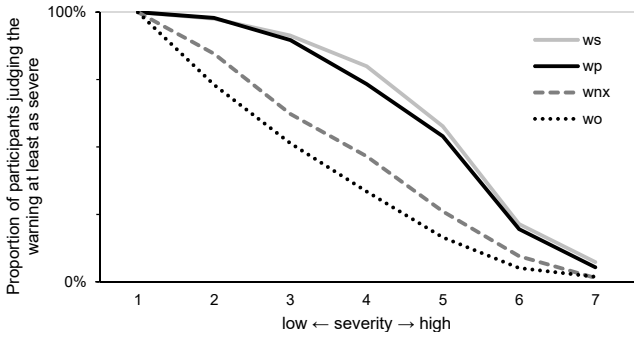


**Figure 4: Perceived severity of the warning by condition, as rated by the participants in the follow-up questionnaire. The graph shows the proportion of participants who perceived the corresponding warning at least as severe.**

| | First name | Current city | Favourite colour | Siblings | Weather sunny | Spending $100 | Computer browser | Health good |
|---|---|---|---|---|---|---|---|---|
| ws | 26% | 31% | 14% | 16% | 10% | 14% | 13% | 11% |
| wp | 25% | 29% | 14% | 17% | 10% | 14% | 12% | 10% |
| wnx | 12% | 10% | 6% | 6% | 6% | 6% | 7% | 6% |
| wo | 4% | 5% | 1% | 2% | 1% | 1% | 1% | 0% |

**Figure 5: Deletion ratios per data item and condition after the warning was displayed. Numbers give the percentage of participants who deleted that field in the condition, amongst participants who had read the instructions.**

## H1c: Providing no explanation

We hypothesised that a warning providing no explanation will have a lower deletion ratio than warnings giving an explanation. The analysis compared deletion ratios in WS to WNX. The proportion was more than double across all levels of item sensitivity and there were significant differences across all items (low: $p = 0.00001$; medium: $p < 0.00001$; high: $p < 0.00001$). Additionally, the warning that provided no explanation was classified as the least usable as per the follow-up questionnaire, achieving an even lower score than the orthography warning.

Although the analyses of results are based on participants who had read and understood the instructions, it is interesting to assess post-warning deletion patterns for those who had not read the instructions. In the combined WS and WP conditions, deletion ratios are significantly lower amongst those who had not read the instructions (for date of birth: $p = 0.002$; for first name: $p < 0.001$; Fisher's exact test; n.s. for favourite colour). We hypothesise this might reflect lower overall engagement with the task at hand. Not only did participants not invest the time to read the instructions, but they also wanted to submit the form as quickly as possible, thereby not spending time on re-visiting their entries. In the WNX condition, the difference in deletion ratios between those who had read the instructions and those who had not is similarly pronounced as for the privacy/security warnings (WS/WP) that provided an explanation. The WNX warning stimulated a deletion action only for current city and date of birth.

## H2: Item sensitivity

We hypothesised that more sensitive items will be removed more often forming a U-shaped relationship. Data sensitivity ratings were taken exogenously from a previous study [13], where participants indicated their willingness to disclose a variety of data items online. Ratings were recorded on a four-point Likert scale, coded '1' (happy to provide) to '4' (unhappy to provide). The data items asked on the webform were mapped to those for which data item sensitivity ratings were available (e.g., 'good' was mapped to 'health'). The relationship was then approximated by a polynomial trend. Highest deletion ratios were observed for items with medium willingness to disclose; low deletion ratios were observed for items with low or high willingness to disclose, such as favourite colour or health respectively. A parabolic fit explains 73% of the variance ($R^2 = 0.7317$). H2 is therefore supported.

## H3a: Sunk cost fallacy

We hypothesised that in line with the sunk cost fallacy the more time a participant spent on filling out the form, the less they will be likely to remove the information entered. On average, participants saw the warning after 86 seconds on the form (range between 82 seconds for WS1HO and 91 seconds for WO). As expected, there is no statistically significant difference across the conditions (two-tailed t-test: $p = 0.12$). The data across all conditions are therefore considered. Participants who took more time completing the form were more likely to delete information than those who had spent less time on the form ($p = 0.02$, G-test of independence). This is the exact opposite of what we hypothesised, H3a is therefore rejected.

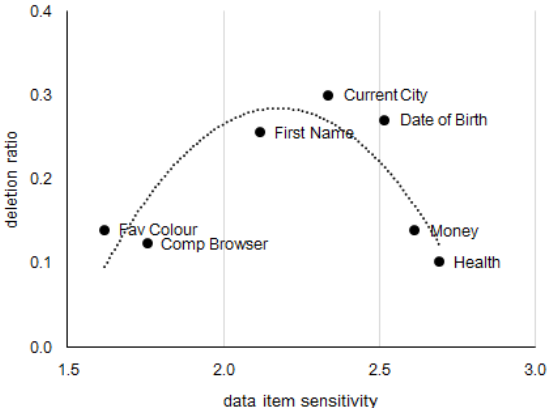


**Figure 6: A parabolic (U-shaped) relationship for the relationship between item sensitivity and deletion ratio.**

## H3b: Options minimising effort

We hypothesised that the options that minimise user effort and save time will be preferred. From the warnings that varied the options on the buttons, the one with the highest click-through rate was 'highlight optional' (53%), followed by 'clear optional' (43%) and 'submit only mandatory' (34%). This came as a surprise since we hypothesised that the warning that would minimise user effort would be preferred. H3b is therefore rejected.

## H4a: Computer literacy

We hypothesised that participants will differ in their deletion/altering behaviour depending on their levels of computer literacy. Computer literacy was measured with an 8-item question battery (reliability measured by Cronbach's alpha: $\alpha = 0.69$). Participants were asked whether they had ever performed tasks indicative of elevated computer literacy. This included changes to computer configurations, some which were security-related (various browser settings, firewall configuration), registered a domain or designed a website, or unscrewed anything on their laptop. Participants generally showed a high level of computer literacy, with the lowest skill being ever having written some computer programme at 33%. As many as 95% respectively 92% of participants had changed the homepage or the default search engine in their browser. In the context of the study (web-form completion behaviour and usage of browser-provided privacy-enhancing technologies), these are very high values. Fewer than one per-cent of participants did not know whether they had done so, also indicating high literacy about these computer concepts. Deletion behaviour was distributed uniformly over varying levels of computer literacy as presented in (Fig. 7) and there was no statistically significant relationship between participants' deletion behaviour and their computer literacy. Hypothesis H4a is therefore rejected.

## H4b: Cyberthreat exposure

We hypothesised that participants will differ in their deletion/altering behaviour depending on their cyberthreat exposure. The term 'cyberthreat exposure' was adopted from Sunshine et al. [21], but we expanded their original item battery which only contained four questions. A 14-item battery of yes/no questions (reliability measured by Cronbach's alpha: $\alpha = 0.70$) was used. Questions included: whether participants had ever been victims of cybercrime, Internet-mediated fraud (e.g., phishing, spam), banking fraud (e.g., fraudulent transactions) and privacy crime in particular (e.g., data breaches, identity theft). As with computer literacy, there was a check question here to ensure that participants were paying attention. The individual scores were calculated for each participant by giving each affirmative answer the score of 1 and summing these. We mapped these scores on deletion behaviour and there was no clear trend there (Fig. 8). Hypothesis H4b is therefore rejected.

## 5.3 Coded free-text feedback from participants

### 5.3.1 *Reasons for disclosure*

Based on their disclosure behaviour in the experiment, participants were asked in the follow-up questionnaire for the reasons why they had provided or had not provided their date of birth. Since these were free-text responses, participants were able to provide as many reasons as they wanted (or none at all) and there was no length restriction for their answers. The data was coded using thematic analysis [4].

**Reasons for disclosing DOB.** Amongst those who provided their DOB, the most common reason for why they did so was that they did not mind (308 out of 1,612 responses, 19%). P76 explained: *"I did not mind revealing that information."* This was followed by participants saying they did not see any harm in providing it (297, 18%). In addition, 17% stated they did so to help research in some way. Also the sensitivity of DOB was thematised by participants, 15% of participants stated DOB was not a sensitive or personal piece of information whereas 16% it was a sensitive piece of information but they often gave another reason why it was fine to disclose it this time. For example, P15 stated: *"While its sensitive information, I don't feel too concerned about sharing it."* Also, 10% stated they thought providing their DOB would bring some benefits with 87 (5%) emphasising a benefit to themselves, P397 stated: *"If i give you my date of birth, You may use it for helping me"*. Interestingly, 159 (10%) participants filled in their DOB to satisfy their completionist instinct, P91 explained: *"It would have bothered me more to have everything filled in but that one."* Further, 130 (8%) participants argued that they their DOB is already out there so not providing it on the form would not have made much difference: *"Date of birth is easy to obtain through a simple google search, more information is typically available on the average users facebook page"* (P879). P118 also assessed the probability of something bad happening as low: *"My date of birth it public information that can be easily found. The likelihood of a negative outcome from giving this information was low."* Participants also reported on their coping strategies, 5% of participants stated they provided an incomplete DOB (e.g., only year or only day and month). Also time and effort were thematised, one answer was particularly interesting as it shows that users are attempting to economise their effort, P318 said: *"it took about as long to enter it as it would have to think about whether or not i wanted to enter it."* Finally, 43 (3%) participants stated it has become their habit to enter their DOB.

**Reasons for not disclosing DOB.** When asked why participants did not provide their date of birth, two re-

sponses dominated: 34% (693 out of 2,060) of participants stressed their DOB was a sensitive or personal piece of information and 32% emphasised DOB was an optional form field. Further, 11% participants said that DOB can be used to identify them. Interestingly, 4% of participants mentioned that DOB can help identify them when combined with other information they provided on the form or that someone could find on the Internet, P96 explained: *"I thought you might be able to use that, along with my first name, and my city to search other databases and get more information about me."* Also, 5% of participants stressed providing DOB can harm them in some way; for example, P3585 stated: *"It can be used to steal my identity"*. Interestingly, 169 (8%) participants stated they heeded the warning that told them to limit disclosure. Also, 6% stressed they did not see the reason why their DOB was needed in this situation and 2% said they would have had no problem providing their age but not their full DOB.

#### 5.3.2 Warnings have a bad reputation

There were several participants who expressed gratitude for being able to participate in the study saying it taught them a lesson: "*its to make sure to let you know that you don't have to put in all you info. i wish more websites had notifications like that*" (P2658). However, some participants expressed negative attitudes towards warnings and were disillusioned about the reasons why the industry used them. Twelve participants in the study interpreted the warning as a means to transfer liability for keeping the data secure from the requester or mTurk onto the user. P719 stated the purpose of the warning was *"[t]o try to protect the liability of the company or website it was displayed on"* and P2090 thought the warning was *"[a] disclaimer that the information [they] provided will be at [their] own risk"*.

There were three participants for whom the control condition, the orthography warning, did not work. P1525 described what they thought the purpose of the warning was by saying: *"It was a verification request to make certain I wanted to submit the form in the previous survey. It was slightly funny in that it pointed out the reasoning was because of spelling errors and not because of security issues."* This shows that pop-up warnings have a strong connotation with security advice. Similarly, P1735 stated: *"I had never seen this before on turk. I thought it was strange. I felt like it meant that I hadn't filled out all of the fields, and so I should go back and check them."*

#### 5.3.3 Possibility of tapping into existing knowledge

Some participants also read things into the warning that were not included there. 354 (9%) participants mentioned some type of crime that could happen as a consequence of over-disclosure, the most common was identity theft, P116 stated: *"any information you share online could be used against you; even information that is vaguely identifying could lead to potential theft of your identity"*. The fact that US population seems to be particularly aware of and responsive to identity theft could be used in the future to illustrate possible consequences of over-disclosure in the warning text.

## 6. DISCUSSION

The study found that warnings can be effective in reducing users' privacy exposure resulting from excessive disclosure of personal details on webforms. We acknowledge that from a user experience perspective, warnings suffer from more fundamental flaws (Sec. 2.3 and 6.1.3), which would nonetheless not influence the research question at hand.

Our study found that warnings that evoke a privacy or security threat make users take privacy-protective measures significantly more often than an arbitrary warning not related to the threats of disclosing personal data. When comparing the conditions where users were warned about over-disclosure with the condition notifying users about spelling mistakes, the former resulted in users going back to the form and removing previously entered details with much higher prevalence. Therefore, the fact that users disclosed fewer optional data items is not due to just any warning, but specifically due to the privacy and security warnings containing an explanation.

The results show that it made no difference whether the threat evoked negative consequences in the privacy or the security domain. Participants were nearly equally likely to delete previously entered information after having seen either of the warnings. However, providing an explanation increased the deletion ratio. This corroborates the findings by Egelman et al. [5] who found that participants were more tolerant of security delays if the threat was explained to them. Similarly here, when there was no explanation provided, participants rated the warning as less severe and less usable. The explanation is an aid for participants to make an informed decision about the real threat. They are thus more able to choose an action that better reflects their privacy preferences. One can argue such explanations are important for users even when the privacy indicator would not be deployed as an active pop-up warning.

Data sensitivity is a traditional operationalisation of privacy threats. The law enumerates a list of particularly sensitive data items and it is corporate practice that certain items require protection above and beyond the safeguards implemented for other items of personally identifiable information. However, data sensitivity does not seem to be a decisive factor for users' behaviour regarding disclosing or withholding their personal details; its influence is less significant than that of perceived relevance and fairness [13]. At the same time, disclosure rates were previously identified as a way to measure privacy concerns and those varied by data item sensitivity [14].

Users' deletion behaviours were found to vary significantly by data item sensitivity, but the relationship is not monotonous. On the one hand, users do not delete low sensitivity items they had previously entered, as the privacy threat may not exceed their desire to disclose. On the other hand, high-sensitivity items cannot be deleted because users left those fields blank in the first place. Consequently, medium sensitivity items are deleted most often.

### 6.1 Contributions of this study

#### 6.1.1 Material results

This study explores the uncharted territory of over-disclosure indicators and finds that warnings can help users manage the disclosure of optional details through webforms. The main lessons learned include: the mention of privacy/security threats makes a privacy warning effective and an explanation of the threats associated with disclosure further heighten users' propensity to remove previously entered personal details. Actionable warnings are characterised by keeping the

user in control as to which items they would like to remove from a webform: optimisation for reduced manual effort leads to lower acceptance than giving users the choice about which optional fields to submit and which ones to erase.

### 6.1.2 Methodological novelty

The study stands out from previous research into privacy by (i) being a true experiment (ii) in a naturalistic environment, (iii) tested on a large number of participants (iv) using fully functional warnings. First, there have been numerous studies in both security and privacy showing that there is a discrepancy between users' reported and actual privacy and security choices (e.g., [2, 11]). Here, we studied users' actual reactions: we did not ask participants about their appreciation of privacy-enhancing tool support but trialled the scheme under scrutiny and adopted objective metrics to assess its effectiveness. Second, participants took part in the study in their homes as opposed to in a lab. If a study is conducted in a lab, participants might feel less comfortable and there might be a trust relationship between the experimenters and the participants. An online study at least partially removes such bias (there is still a potential impact of the requester's mTurk reputation etc.). Third, each warning was tested on hundreds of participants which would not have been feasible if the study was lab-based. Fourth, in exploratory studies, the warnings or designs used are sometimes mock-ups which can give the participants the impression the interaction is not real. In the study here, the warnings were fully functional, they did what they said they would. For instance, if the 'clear' or the 'highlight' button were pressed, form fields were actually erased or flagged.

### 6.1.3 But aren't warnings bad?

Previous research has shown that warnings are effortful and users are conditioned to ignore them. Feedback data from the study confirmed that users have negative associations with warnings. For these reasons, an over-disclosure warning should ideally be implemented as a browser plug-in that would highlight mandatory and optional fields for the user while a webform loads on a website. In this way, one could achieve visible but non-disruptive warnings—users would not need to click them away to continue with their primary task. However, unlike previous passive indicators discussed in Section 2.1, they would not be placed on the periphery of the screen but in the centre of user's visual attention. The plug-in could also be configurable and the user could predefine what kind of information they are comfortable sharing with what type of website. This would pave the way for a context-aware auto-fill feature. The results of the study show that a more paternalistic approach, such as suppressing optional form fields, needs to be assessed carefully given users' quest for control.

Even implemented in the way they were used in the study, privacy warnings would not significantly contribute to warning fatigue. They would typically apply to webforms found in sign-up procedures for a service for example. When the user subsequently uses the service, they log in to the website with their username and password, both mandatory items. An over-disclosure warning would only appear during the initial sign-up, but not for the following usage transactions. Pop-up frequencies would be much lower than for download warnings [11]. Such a warning would also feature a possibility to opt out of future notifications (as shown in Fig. 1).

## 6.2 Limitations

### 6.2.1 Sample

The sample of participants who tested our warnings might not be representative of the general population. Although studies have shown that mTurk workers are diverse across several demographic dimensions such as gender, age and income [18], more recent research demonstrated that they might be more concerned about their privacy [10].

### 6.2.2 Weak primary task

Since security and privacy are secondary tasks that users encounter while performing a primary task online (e.g., shopping, emailing), a study has to mimic this to produce ecologically valid results [12]. Although we did not provide an explicit cover story for collecting the data, it is likely that the participants considered the study to be a pre-screen for future tasks. One could argue that the lack of an explicit primary task could have made our participants more suspicious or on the contrary, more inclined to volunteer their details. The stated limitations open up future research opportunities which we discuss in Section 6.3.

### 6.2.3 Motivation to provide correct information

It might be argued that the participants in our study had no motivation to provide correct information and this could invalidate the results of our study. We can think of at least three reasons why this is not a limitation but rather a reflection of what happens in the real world. First of all, our participants were randomly assigned to the experimental treatments and the proportion of participants who might have provided fake data should be evenly distributed across the experimental treatments. Second, a study of this type was seen by some participants as a pre-screen for future tasks. In the exit questionnaire, we asked participants what they thought the purpose for collecting this information was and 287 (7%) answered it might be to pre-screen them for future tasks. Thus, the more accurate the information, the more likely they would be invited to the correct tasks. Third, we wonder why participants would delete information that was not correct in the first place. We assume there was a proportion of participants who falsified the data and a proportion of participants who provided accurate data. If the data was fake from the beginning, they did not need to delete it. In the real world, the warning would have been of no use to those falsifying information because they have a different coping strategy for dealing with being asked to provide information they do not want to provide.

### 6.2.4 Habituation

One might argue that users are habituated to warnings and did not pay attention to the ones in our study. It is true that users are habituated to things that they see in familiar context but this was a novel situation. Similarly, Krol et al. [11] found in their study that users did ignore the advice of the warning and proceeded, but only after having read it which was supported by eye-tracking data and the fact that participants correctly identified the warning after the session. While Krol et al.'s [11] study looked at warnings similar to existing ones, the warnings trialled in this study were new and presented in a novel situation. Therefore, chances are high that participants did notice our warnings.

### 6.2.5 A one-off reaction

Our experiment looked at only one reaction to one instance of a warning and this might not have been representative of the user's general behaviour in varying situations. A longitudinal study to examine repeated reactions to the same warnings is earmarked for future research.

## 6.3 Further research

Future research could test privacy indicators over a longer period of time to measure how they impact users' actions. Only such panel studies can assess habituation effects, counterbalance novelty biases and observe reactions across contexts. A long-running study would allow privacy indicators to become part of the Web browsing routine.

# 7. CONCLUSIONS

We conducted a study with 4,620 participants that aimed to help design better privacy warnings for those completing webforms. Since our previous study showed that users overdisclose at the expense of time, cognitive and physical effort, this study looked into ways how to minimise effort and inform disclosure on webforms. We mapped out and explored the design space of different warning parameters. The results indicate that users preferred warnings that provided an explanation and preferred options that left them in control of what information they disclose even at the expense of effort. We found that effective privacy warnings should feature: (i) the diagnosis and explanation of the privacy threat originating in voluntary disclosure and (ii) the possibility for the user to easily, but selectively remove optional data items.

## Acknowledgments

We would like to thank Angela Sasse for her feedback on the design of the study. Many thanks to Simon Parkin and Ingolf Becker for their help in the preparation of this paper. This research was funded by EPSRC's grant to the Security Science Doctoral Training Centre, grant number: EP/G037264/1.

# 8. REFERENCES

[1] Akhawe, D., and Felt, A. P. Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness. In *USENIX Security* (2013), pp. 257–272.

[2] Berendt, B., Günther, O., and Spiekermann, S. Privacy in e-commerce: stated preferences vs. actual behavior. *Comm. of the ACM 48*, 4 (2005), 101–106.

[3] Berners-Lee, T., and Connolly, D. Hypertext Markup Language - 2.0. http://tools.ietf.org/html/rfc1866section-8, 1995.

[4] Braun, V., and Clarke, V. Using thematic analysis in psychology. *Qualitative research in psychology 3*, 2 (2006), 77–101.

[5] Egelman, S., Acquisti, A., Molnar, D., Herley, C., Christin, N., and Krishnamurthi, S. Please Continue to Hold: An empirical study on user tolerance of security delays. In *Workshop on the Economics of Information Security (WEIS'10)* (2010).

[6] Egelman, S., Cranor, L. F., and Hong, J. You've been warned: An empirical study of the effectiveness of web browser phishing warnings. In *SIGCHI Conference on Human Factors in Computing Systems (CHI'08)* (2008), pp. 1065–1074.

[7] Ghostery. Knowledge + Control = Privacy. http://www.ghostery.com/, 2016.

[8] Herley, C. So long, and no thanks for the externalities: The rational rejection of security advice by users. In *New Security Paradigms Workshop (NSPW'09)* (2009), ACM, pp. 133–144.

[9] Herley, C. More is not the answer. *IEEE Security & Privacy 12*, 1 (2014), 14–19.

[10] Kang, R., Brown, S., Dabbish, L., and Kiesler, S. B. Privacy Attitudes of Mechanical Turk Workers and the US Public. In *Symposium on Usable Privacy and Security (SOUPS'14)* (2014), pp. 37–49.

[11] Krol, K., Moroz, M., and Sasse, M. A. Don't work. Can't work? Why it's time to rethink security warnings. In *International Conference on Risk and Security of Internet and Systems (CRiSIS'12)* (2012), IEEE, pp. 1–8.

[12] Krol, K., Spring, J. M., Parkin, S., and Sasse, M. A. Towards robust experimental design for user studies in security and privacy. In *Learning from Authoritative Security Experiment Results (LASER) Workshop* (2016), IEEE.

[13] Malheiros, M., Preibusch, S., and Sasse, M. A. "Fairly Truthful": The Impact of Perceived Effort, Fairness, Relevance, and Sensitivity on Personal Data Disclosure. In *Trust and Trustworthy Computing*, M. Huth, N. Asokan, S. Čapkun, I. Fléchais, and L. Coles-Kemp, Eds., vol. LNCS 7904. Springer Berlin Heidelberg, 2013, pp. 250–266.

[14] Preibusch, S. Guide to measuring privacy concern: Review of survey and observational instruments. *International Journal of Human-Computer Studies 71*, 12 (2013), 1133–1143.

[15] Preibusch, S., and Bonneau, J. The privacy landscape: Product differentiation on data collection. In *Workshop on the Economics of Information Security (WEIS'11)* (2011).

[16] Preibusch, S., Krol, K., and Beresford, A. R. The privacy economics of voluntary over-disclosure in Web forms. In *Workshop on the Economics of Information Security (WEIS'12)* (2012).

[17] Privacy Bird®. Privacy Bird® – Tour. http://www.privacybird.org/tour/1_3_beta/tour.html, 2016.

[18] Ross, J., Zaldivar, A., Irani, L., and Tomlinson, B. Who are the turkers? Worker demographics in Amazon Mechanical Turk. *Department of Informatics, University of California, Irvine, USA, Technical Report* (2009).

[19] Schechter, S. E., Dhamija, R., Ozment, A., and Fischer, I. The Emperor's New Security Indicators. In *IEEE Symposium on Security and Privacy (S&P 2007)* (2007), pp. 51–65.

[20] Staw, B. M. Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action. *Organizational behavior and human performance 16*, 1 (1976), 27–44.

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

[22] W3C, Ian Hickson. HTML5. A vocabulary and associated APIs for HTML and XHTML, Section 4.10 Forms, 2012. W3C Working Draft – 25 May 2011.

[23] Wu, M., Miller, R. C., and Garfinkel, S. L. Do security toolbars actually prevent phishing attacks? In *SIGCHI Conference on Human Factors in Computing Systems (CHI'06)* (2006), pp. 601–610.

# 9. APPENDIX

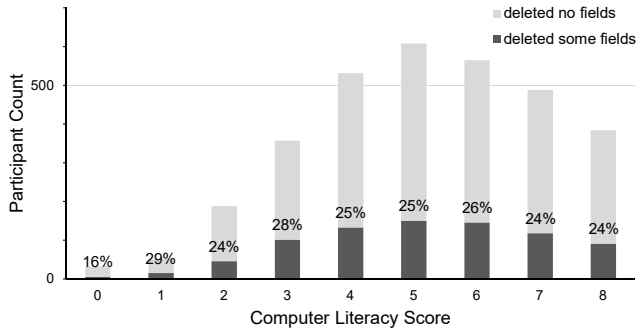


**Figure 7: Deletion prevalence by computer literacy score.**

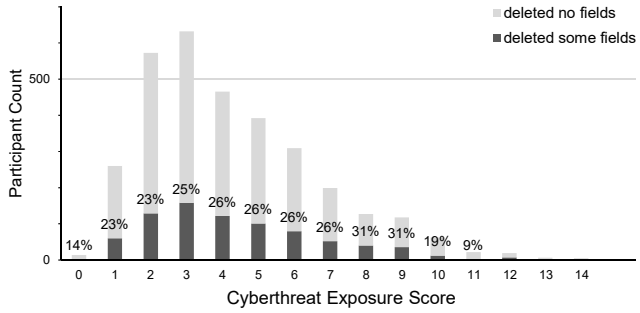


**Figure 8: Deletion prevalence by cyberthreat exposure score.**