Contextualized Web warnings, and how they cause distrust

Steffen Bartsch[1], Melanie Volkamer[1], Heike Theuerling[2], and Fatih Karayumak[3]

[1] CASED, TU Darmstadt
Hochschulstraße 10, 64289 Darmstadt, Germany
{steffen.bartsch,melanie.volkamer}@cased.de
[2] IAD, TU Darmstadt
Petersenstr. 30, 64287 Darmstadt, Germany
h.theuerling@iad.tu-darmstadt.de
[3] Cyber Security Institute, TUBITAK BILGEM
41470 Gebze / Kocaeli, Turkey
fatih.karayumak@tubitak.gov.tr

Abstract. Current warnings in Web browsers are difficult to understand for lay users. We address this problem through more concrete warning content by contextualizing the warning – for example, taking the user's current intention into account in order to name concrete consequences. To explore the practical value of contextualization and potential obstacles, we conduct a behavioral study with 36 participants who we either confront with contextualized or with standard warning content while they solve Web browsing tasks. We also collect exploratory data in a posterior card-sorting exercise and interview. We deduce a higher understanding of the risks of proceeding from the exploratory data. Moreover, we identify conflicting effects from contextualization, including distrust in the content, and formulate recommendations for effective contextualized warning content.

1 Introduction

Warnings in Web browsing are an example of how difficult it is to craft effective security interventions. A plethora of studies (e.g. on certificate warnings: Sunshine et al. [19]) have shown that current warnings are ineffective at influencing the behavior of users for two main reasons: First, because of habituation effects from the frequent unhelpful warnings in non-critical situations [2]. Second, because of the technical language that prevents users from understanding the risks of proceeding – that is, how likely it is that an adverse event occurs and what the personal consequences are [6, 8, 13]. We thus not only need to prevent the occurrence of warnings in uncritical situations, but also make the warnings understandable so that the infrequent warnings will enable users to take informed decisions about proceeding based on the actual risks involved.

One proposal to solve the problem with the understanding of the risks is to move away from traditional approaches to warnings as described by Wogalter

[20]: generic hazard warnings with static texts and symbols for a broad audience. Instead, we follow earlier proposals to *contextualize* security interventions and thereby increase their *concreteness* [7, 4]. The idea is to employ additional information on the context (e.g. user intention) so as to generate more concrete warnings – for example, by mentioning specific consequences, such as credit-card fraud in case of online shopping – and therefore make it easier for users to relate to and understand the risk of proceeding.

Since contextualization has been primarily studied technically for warnings up to now – for example, on how to acquire the available context information [7] –, we address the practical value of contextualization in this paper. The goal of this work is to test whether contextualization is more effective in increasing the understanding of the risks and in influencing behavior than traditional content, and to explore how to craft effective contextualized warning content. We developed contextualized warning content based on a pre-study with lay and expert users. We then conducted a between-subject study with 36 participants who were confronted with warnings either showing the contextualized content or content from existing warnings while solving realistic tasks in a lab environment. In addition to the participants' reaction to the warnings, we also collected qualitative data from a posterior card sorting of the warnings and a posterior interview. Our main contributions are:

1. We show a positive effect from contextualization on how concretely participants assess the risks of proceeding;
2. We demonstrate how confounding factors, such as visual stimuli that imply severity, can dominate the effect of contextualization in real-world settings;
3. We identify complexities related to contextualization, including distrust in the warning content due to its concreteness;
4. We derive recommendations of how to craft effective contextualized content.

2 Prior research on the content of Web browser warnings

Bravo-Lillo et al. [6] showed empirically that warnings are not understood – for example, due to technical terminology. Improved warning content may help, though: Biddle et al. [5] found that their reformulated warnings made users more responsive to different levels of connection security. More specifically, Downs et al. [8] showed that phishing warnings are more often ignored if the threats and consequences are unknown. Furthermore, Kauer et al. [13] found that individuals are more likely to heed warnings if they perceive personal consequences. However, when Krol et al. [14] confronted users with either a very generic warning or one with more specific consequences, they found no significant difference in behavior.

To warn in an adequate form and achieve the necessary impact, De Keukelaere et al. [7] proposed to adapt the intervention to the context; they found improvements from considering the security risk and prior actions of the user. In this paper, we follow a related approach, the Framework for Contextualized Interventions (FOCI), which supports the systematic development of contextualized security interventions [4]. The framework targets two aspects: first, whether,

when, and in which form the intervention appears (intervention strategy, e.g. active as a warning or passive as a symbol), and, second, what content it conveys (e.g. technical threats or personal consequences). This paper focuses on the content aspect.

3 Pre-study: How expert and lay users assess Web risks

From prior work, it remains unclear, *what* contextualized content helps users in understanding the risks of proceeding. To guide our choice of content in the main study of this paper, we explored what is missing for users to understand the risks. Prior literature showed that expert and lay users differ in how they assess risks and that experts are more likely to have a sufficient understanding [3]. Thus, we analyzed the difference between expert and lay users in how they assess risks of Web browsing.

3.1 Study design

We recruited seven lay and seven expert users from personal contacts for a card-sorting exercise. Their task was to sort Web site screenshots into stacks of similarly perceived consequences if their personal account was compromised. Our goal was to motivate participants to talk about factors that influence their categorization. We asked expert and lay users to imagine that they have user accounts at 67 Web sites (selected from the Alexa.com Top-500 most-visited Web sites in Germany for diversity), which were presented to them as the cards to be sorted in the form of printed DIN-A5 screenshots of the pages ("picture sorting": giving visual clues [17, p. 83]). Expert users (age avg. 37 yrs., min 28, max 52) covered a broad span of participants professionally related to security, including at least two each of system administrators, security researchers, and security consultants. Lay users (age avg. 23 yrs., min 22, max 25) were without professional relation to security, but covered a broad span of self-assessed PC expertise from receiving help with computer problems to providing help, even with difficult problems.

3.2 Analysis

We qualitatively analyzed the transcribed recordings of the card-sorting exercise. We inductively developed codes for the risk concepts that participants used to assess risks. These concepts differed between arguments based on the type or function of the page (e.g. activity "Shopping", institution "Bank", content "Information") and risk-related factors (affected data "Contacts", Consequence "Financial loss", adversary activity "Hacker accesses my account"). We also found a difference in how concrete these arguments were (e.g. for consequences "I'll lose money from my account" vs. "This somehow affects my finances").

Table 1. Primary concepts used in the categorization of Web sites

Argument	Examples	Lay	Expert	$p < 0.05$
Type of page	Activity "Shopping"	225 **58%**	134 **38%**	Yes
Risk factor	Consequence "Financial loss"	172 **45%**	236 **67%**	Yes
Total		385	354	

3.3 Experts focus more on consequences and adversary activities

Expert and lay users significantly differ in their argumentation as shown in Table 1[4]. Experts more frequently used the risk-factor arguments, particularly the specific consequence and the adversary activity, than lay users. Lay users, in contrast, more often relied on the Type-of-page factors of a Web site without explicitly considering risk factors – for example, only the possible activities ("Eventim, that's where one may buy, order tickets"). Table 2 shows how the risk factors break down into different risk concepts. When lay users discussed risks, they less often mentioned consequences and adversary activities. Our hypothesis for warning content thus is to emphasize these factors for lay users to help them to better understand the risks of proceeding.

3.4 Experts are more concrete

Not only did lay users less often discuss risk factors than experts; when they did, they did so less concretely. Experts rather formulated concrete adversary activities ("modifies my preferences") and named the concrete consequence or affected personal data ("bank account data put there"), and the concrete evaluation of specific risk factors ("I will find out quickly"), instead of only mentioning solely a general risk level such as "I'd classify it as comparatively bad" when categorizing Web sites (cf. concreteness in Table 2).

[4] We applied a Welch Two Sample t-test on the individuals' proportions and noted in the last column for which proportion the differences between expert and lay users are significant, i.e. the null hypothesis was rejected because of $p < 0.05$.

Table 2. Frequency of different risk concepts and their concreteness

Risk concept	Lay	Expert	$p < 0.05$
Data-related	101 59%	126 53%	No
Concrete data	41 24%	86 36%	No
Consequence	63 37%	148 63%	Yes
Concrete consequence	34 20%	112 47%	Yes
Adversary activity	75 44%	148 63%	Yes
Concrete activity	22 13%	114 48%	Yes
Further risk factors	90 52%	142 60%	No
Concrete risk factor	3 2%	65 28%	Yes

4 Research hypotheses

The findings from the pre-study indicate that it is helpful for lay users if we emphasize adversary activities and consequences, and we are thereby more concrete with respect to the current situation. This is further supported by literature on risk communication: According to Rothman and Kiviniemi [16] concrete risks are more successful in creating awareness and influencing behavior in health risk communication: Consequences (symptoms) that are easier to picture increase the awareness, as do testimonials of affected individuals when there is an identification with those. Cognitive psychology indicates that it is important that people are able to "simulate" or imagine the antecedents and consequences of risks [12]. As previously noted, Kauer et al. [13] found that individuals are less likely to ignore warnings when they perceive personal risks, corresponding to the experience from medical risk communication. Overall, as depicted in Figure 1, we thus expect that *contextualization* of the content and thereby including *concrete risks* according to the situation will increase the *understanding of risks* and thus the *motivation to behave securely.*

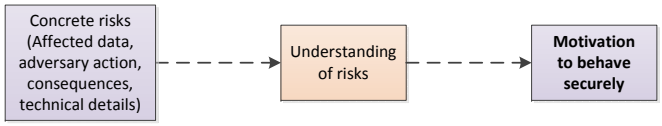

Fig. 1. Model underlying the research hypothesis

In this paper, we apply this model to study the behavior of participants when confronted with different warnings, that is, whether they follow the recommendation of the warning and leave the Web site (comply) or whether they proceed with their task on the Web site. While prior studies [14] have found that the habituation effect dominates the effect of different content, we assume that our more intensively improved content should influence the behavior of the participants. Accordingly, our first hypothesis is:

> *H1 The participants who are confronted with the contextualized content more frequently comply with warnings than those with standard content*

When the change in behavior is due to better understanding of the risks, we expect that this change in whether to comply (the warning effect) occurs differently depending on the objective risk of the individual situation [13], despite potential confounds, such as additional visual stimuli:

> *H2 The relation between the* warning effect *and the objective risk is stronger for warnings with contextualized content than for standard content*

Moreover, the difference in understanding should not only show in the behavior, but also when asked to consciously assess the criticality of the situation (warning perception):

> *H3 The relation between the* warning perception *and the objective risk is stronger for warnings with contextualized content than for standard content*

Lastly, since we hypothesize that better understanding is related to perceiving risks concretely, we expect participants to also emphasize concrete aspects in their risk assessment depending on the type of warning:

> *H4 Participants who are confronted with the contextualized content assess the risks of proceeding more concretely than those with standard content*

5 Research method

Our study has two goals: first, testing the effectiveness of contextualized content in warnings in behavior (H1–2) and in conscious assessment (H3–4), and, second, exploring how to optimally contextualize content. To generate realistic behavioral results, we confront 36 participants either with warnings with contextualized or standard content while they solve twelve realistic Web-browsing tasks. Moreover, we collected and analyzed posterior qualitative data.

5.1 Study design overview

The between-subjects study on warnings with contextualized or standard content consisted of two main parts. In the first, behavioral, part, participants solved twelve tasks and were interrupted with warnings in five of these, representing situations of different levels of objective risks. Due to the technical complexity of integrating the different warnings in the Web browsing tasks, the study was conducted in our usability lab on a study laptop. In the second, explanatory, part, participants conducted a card-sorting exercise of screenshots of the warning scenarios, explaining their reasoning, and were interviewed on the risks of proceeding in each situation.

No IRB consent was required as all university criteria for studies without explicit IRB consent were met. For privacy reasons, the screening data that included personal identification (name was optional, but an email address was required for experiment logistics) was separated from the screening data used later for demographics. After the end of the first part of the study, the participants were informed about the actual goal of the study.

5.2 Instruments: Warnings with contextualized and standard content

We created prototypes of warnings with contextualized and standard content for five scenarios of different objective risk levels for the study. We redesigned both

Table 3. Scenarios with technical threat, estimated likelihood of attack (L), and the estimated severity of likely consequences (S)

Scenario	Activity	Technical threat	Data at risk	Highlighted properties	L	S
Bank	Log in to online banking	Self-signed certificate	Banking credentials	Identity	High	High
Shop	Pay with credit card	Unprotected connection	Payment credentials	Identity, confidentiality	Med	High
OSN	Register for OSN	Unprotected connection, negative reputation	Personal data	Identity, confidentiality, trustworthiness	Med	Med
Insurance	Request quote for insurance	Self-signed certificate	Health data	Identity, confidentiality	Med	Med
Information	Find flight cost	Negative reputation	Travel destination	Trustworthiness	Med	Low

types of warnings to have the same "newness" effect for both types of warnings [18]. For standard content, we reused and adapted the content from warnings from Mozilla Firefox 3 and Web of trust 1.4.8. For the contextualized content, we followed the insights of how lay and expert users differ in risk assessment (cf. Section 3). Since we recruited only lay users, we included concrete information on the risks of proceeding.

We crafted the scenarios with warnings to represent a wide range of objective risks to enable a within-subject comparison of participants' behavior regarding different levels of risks. The scenarios, listed in Table 3, include self-signed certificates, unencrypted connections, and negative reputation for the activities banking, shopping, social networking (OSN), requesting an insurance quote, and information seeking for flights.

A translated version of the warning with contextualized content for the banking scenario is shown in Figure 2. The warning with the contextualized content was developed in an iterative process that included eliciting the concrete risk aspects to mention, expert consultations, and user feedback on the warning design and content. The version employed in the study included:

1. the user intention – "entering account number and PIN" in the banking scenario;
2. a warning headline with an indication of the attack probability – "probably an attack";
3. the potential personal consequences from proceeding – "attackers could plunder your account". From the potential consequences to name, we selected those appropriate for the situation that were most often mentioned in the pre-study;
4. boxes with concrete and transparent indications whether and how three main security properties of the situation (identity of Web site provider, confiden-

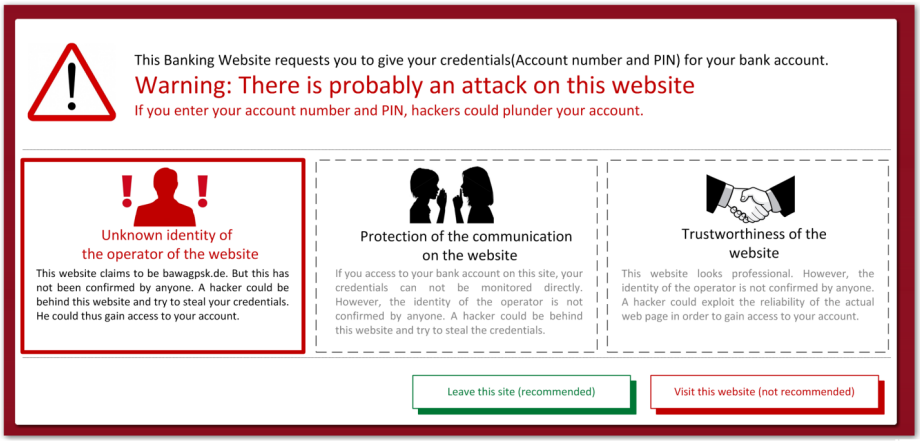

Fig. 2. Translated example warning with contextualized content

tiality of connection, trustworthiness of Web site provider)[5] are upheld. Each box included a short description of how the security property affects the user when proceeding. The boxes with threatened properties are highlighted (shown in Table 3) as a confound to explore how such visual stimuli interact with the effects of contextualization.

5.3 Procedure

After an initial introduction that included the priming as a Web-browsing usability study – to counter a potential unrealistic focus on the warnings –, the experimenter informed the participants that they would need to complete twelve tasks (cf. Section 5.2). To counter the effect that participants may feel an unrealistic urge to complete the given task in the lab setting [18], we offered an alternative: filling out a "usability problems" form for the study, which required the participants to enter a carefully selected amount (3 lines) of information on a separate sheet of paper. In this way, participants would not perceive the alternative as an easy way to get around the tasks.

Each task described a problem related to the overarching theme of travelling and gave instructions, including an address of a Web site, to solve it. Where it was necessary to enter data, such as credentials, the instructions also included these. To reduce the confounding effects of using a stranger's laptop and personal data, the experimenter presented himself as student whose personal credentials and laptop were used in the study. As part of completing the task, each task either caused a warning to appear or not (warning or dummy task, respectively). To prevent the participant from noticing the actual intent of the study early on,

[5] We identified these properties by analyzing an extensive list of threats in Web browsing and how these can be addressed through security properties. This approach to content presentation follows Biddle et al. [5].

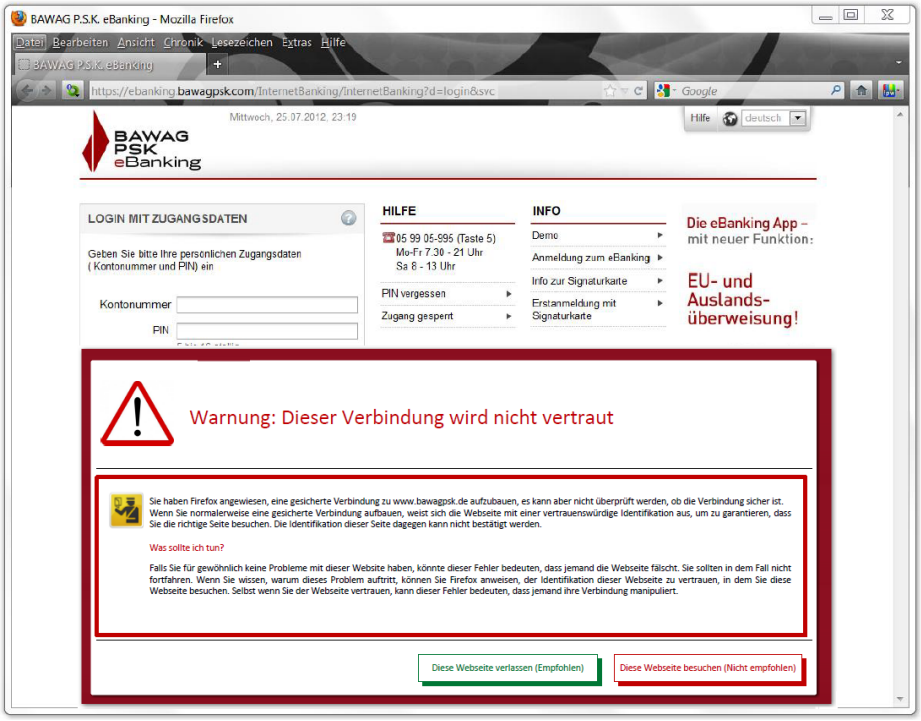

Fig. 3. Example warning with standard content as screenshot (original German content)

one to three dummy tasks occurred between the warning tasks. We organized the warning tasks in one of two fixed orders to cancel out effects from the order, either starting with the most or least critical scenario, banking or flight information, respectively.

In the second part of the study, the experimenter revealed the actual goal of the study to the participants and instructed them to read the warnings again. To further explore their perception of the risks in the warning tasks, participants were asked to carry out a card-sorting exercise with printouts of the warning scenarios (Web site screenshot with warning overlayed, as shown in Figure 3), sorting them by criticality and commenting on their reasoning. The experimenter further asked the participants to explain for each warning what they thought why the warning appeared and what the potential consequences of proceeding would have been.

The audio was recorded for the entire study.

5.4 Participant recruitment

We targeted lay users with the warning content so that we excluded participants with security-related professional or study background. We advertised for the

Table 4. Participant demographics

Group	Contextualized	Standard
Female	11	11
Male	7	7
Mean age (stddev)	26.3 (4.1)	24.8 (2.6)
Mean PC knowledge	63.9 (20.0)	51.0 (22.5)

study as one on usability problems with Web browsing using posters at public places (local supermarkets, bus stops), direct dissemination of a similar flyer to people on and off-campus and through email to non-technical students' mailing lists. We offered EUR 10 compensation for participation. Potential participants had to complete an online screening survey, including demographics, their professional/study background, and PC skills. From those, we selected participants and randomly assigned them to the two groups, but arranged for gender balance. The demographics of the two groups are shown in Table 4.

5.5 Data collection and analysis

To test the hypotheses, we collected quantitative and qualitative data from the study. Quantitative data consisted of:

1. Which warnings participants complied with from the experimenter's notes (for H1–2)
2. The order of the warnings from the card-sorting exercise (H3)

Qualitative data was collected through the audio recordings, which were transcribed for analysis. In particular, we analyzed the qualitative data for

1. How participants reasoned about risks while conducting the card-sorting exercise and while answering the interview questions (H4)
2. Further comments on the appearance and content of the warnings

For both aspects, we coded the qualitative data, a method that has been successfully employed in HCI research [1]. We inductively developed codes by first applying "open coding", then "selective coding" from Grounded Theory [11]. To analyze the participants' reasoning about risks, we identified different risk concepts that participants used – for example, whether they referred to the affected data, consequences, technical threats, adversary activities, or abstractly as "this is a dangerous situation". For comments on the warning, we identified the categories design, content, understanding, and doubts. One researcher assigned a total of 823 codes (625 on risks, 198 on warnings) to 733 quotes in the transcripts. For coding reliability, a second researcher independently coded six of the transcripts as suggested by [15], showing a good overlap.

Table 5. Overview of average compliance with warnings relative to all warnings for both groups

Group	Contextualized		Standard	
	n		n	
Average compliance	18	**46%**	18	**17%**
Female	11	35%	11	15%
Male	7	63%	7	20%
Low PC knowledge	2	40%	4	30%
Med. PC knowledge	10	50%	9	18%
High PC knowledge	6	40%	5	4%

6 Results

6.1 H1–3: The effectiveness of contextualization

We recorded the compliance of the participants with each warning while completing the tasks to test H1:

> *H1 The participants who are confronted with the contextualized content more frequently comply with warnings than those with standard content*

H1 was confirmed, since the participants with contextualized content significantly (Fisher's exact test for the distribution of compliance count, $p = 0.04$) more often complied with the warnings than the group with the standard warning (shown in Table 5 as overall relative compliance). We saw similar trends for different demographic groups. Since the self-reported PC knowledge should represent the self-confidence of participants with respect to interacting with PCs and people feeling insecure tend to comply with warnings [14], it is not surprising that lower knowledge scores seem to correlate with higher compliance, particularly for the standard warnings.

We further hypothesized that participants can better differentiate between the different risk levels by measuring their compliance to the warning as the warning effect:

> *H2 The relation between the* warning effect *and the objective risk is stronger for warnings with contextualized content than for standard content*

This hypothesis cannot be confirmed by our results. We even see a contrary effect as shown in Figure 4: The participants with the standard warnings, who needed to deduce the risk level from the scenario and the technical threat, showed a general trend that corresponds to the objective risk level (supporting the findings from Kauer et al. [13]). However, this was not the case for the contextual-warning group. If the group with the contextualized warnings had better understood the situation, the trend should have been more pronounced. Instead, the shop and OSN scenarios caused more compliance than expected

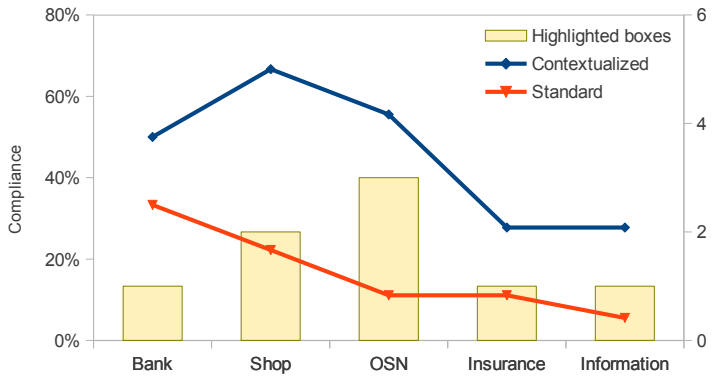

Fig. 4. Average compliance by scenario for both groups, with the number of highlighted boxes in contextualized warnings

from the relative risk level. A likely explanation is that our implanted confound, the number of highlighted boxes, strongly influenced the decision to comply. This result shows that content can be easily dominated by other factors, in line with the results of Krol et al. [14].

We not only expected the behavior to more closely correspond to the objective risk levels, but also tested how participants perceived the warnings when instructed to read them carefully. We conducted the posterior card-sorting exercise for this hypothesis:

> *H3 The relation between the* warning perception *and the objective risk is stronger for warnings with contextualized content than for standard content*

In the card-sorting exercise, the order of the contextual group corresponded only slightly better with the objective risk than the control group (particularly for the bank and flight scenarios; see Table 6 that shows the mean sort order). This is supported by the lower standard deviation (in brackets in the table) for the most and least critical scenarios; the contextual group produced less spread in the sorting than the standard-content group. Moreover, the bump from the highlighted threats is not present in the card-sorting results, where users were instructed to actually read the warning, further supporting the notion that the bump in the behavior was caused by the implanted confound.

6.2 H4: Participants' assessment of the risks

We instructed the participants to think aloud while sorting the warnings after completing the tasks, and, in addition, asked them to state the reasons for each warning's occurrence and what could have been the consequences of proceeding in each situation.

Table 6. Average sorting position for the warning scenarios, 1 being most, 5 least risky (with standard deviation)

Group	Contextualized	Standard
Bank	**1.3 (1.1)**	**1.7 (1.8)**
Shop	2.1 **(0.9)**	2.1 **(1.9)**
Insurance	3.2 (1.9)	2.9 (1.8)
OSN	3.8 (1.7)	3.9 (1.6)
Information	4.6 (1.8)	4.3 (2.3)

H4 Participants who are confronted with the contextualized content assess the risks of proceeding more concretely than those with standard content

We coded how participants mentioned or reasoned about risk in the transcripts, differentiating between different concepts of risks. In Table 7[6], we report the occurrence of the different concepts relative to the total quantity of risk-related codes for the two groups in the study. While the contextualized and the standard-content groups similarly often mentioned the affected data as a risk consideration, the context group more often mentioned consequences (in particular, concrete consequences, such as property-related, like losing money) and adversary activities, such as how an adversary would access their account[7]. In contrast, the standard-content group more often resorted to problematic consequence concepts, such as abstract "something bad will happen"; more technical aspects, such as the missing encryption; and more abstract reasoning, such as "this is a dangerous situation"[8]. As elaborated in Section 4, we expect that more concrete concepts are more "natural" and thus more understandable for lay users that we recruited the participants for. Accordingly, we conclude from the reported frequencies of risk concepts that the contextual warnings were more understandable. We will verify this aspect in future work.

6.3 Further findings on the contextualization

The participants mentioned further aspects on the warnings that relate to the content of the warning and its contextualization.

"Too much text" Five participants who were confronted with the contextualized warnings mentioned in the posterior interview that there was too much

[6] p values of a Welch Two Sample t-test on the participants' proportions for each risk concept are noted in the last column.

[7] We also checked whether participants only directly reproduced (reading aloud) the content of the warning. This was not the case. Due to the interview situation, all participants formulated their own statements. Moreover, the majority at least paraphrased the content – for example, for property-related consequences, participants used different terms in 77% of the cases.

[8] In contrast to abstract consequences, abstract risk reasoning does not point to any consequences at all.

Table 7. Risk concepts mentioned by participants relative to the total number of mentioned risks, including different types of consequences mentioned

Group		Contextualized		Standard		p
	Example	n		n		
Risk		354		271		
Data	Payment credentials	81	23%	61	23%	0.98
Adversary activity	"Accesses account"	79	**22%**	25	**9%**	< 0.001
Consequences	Financial loss	120	34%	75	28%	0.065
Mitigation	Enter fake data	9	3%	8	3%	0.77
Technical	"Missing encryption"	33	**9%**	65	**24%**	< 0.001
Context	"Unknown site"	9	3%	12	4%	0.28
Abstract	"Seems dangerous"	10	**3%**	24	**9%**	< 0.01
Other		13	4%	1	0%	
Consequences		120	34%	75	28%	
Annoyance	Spam	5	1%	11	4%	0.049
Property	Loose money	78	**22%**	28	**10%**	< 0.001
Problematic	Unknown, misconception...	11	**3%**	32	**12%**	< 0.001
Other		26	7%	4	1%	

content or too small text in the warning. However, several also stated that all the information given was necessary.

Prior partial knowledge and experiences Due to our recruitment strategy, none of the participants was a security expert. However, eleven participants referred to their prior partial knowledge on risks or prior adverse experiences at some point in the risk assessment. While this knowledge helped in the assessment of the risks, its absence in the majority of cases also demonstrated the lack of reliability of warnings if their understanding requires prior knowledge to deduce consequences. Moreover, prior general knowledge also caused the speculation on and misconceptions of consequences as seen in the above analysis of mentioned risk concepts. One effect was that the availability heuristic led participants to assume less severe consequences.

Risk attitudes Participants differed in what consequences they considered relevant for them. For example, one participant mentioned that it would be more interesting to mention that pictures from the OSN account would be reused than id theft.

Trust in the warning In several cases, statements of the participants revealed distrust in the warning, particularly for the contextualized warnings. For example:

> "But I found this strange because an employer must not access my data, really... because everything would need to be passed on and registered and that cannot be true!" (T2)

The distrust either related to whether the described attack could take place, as in this quote (4 cases); to the stated consequences (8); or to the basis of the risk assessment, such as user ratings (13). All of these aspects were originally included in the warning content to increase the warning's concreteness.

7 Discussion

7.1 Challenges of drawing attention to warning content

While we show that the contextualized warnings significantly more often caused participants to comply with the warning, our findings also support the notion that it is difficult to draw people's attention to the content of warnings in real-world scenarios (cf. [14]). We assume that habituation, lack of helpful information, and time pressure provide strong incentives for people not to expend enough cognitive effort on a warning to completely grasp its content. Having a combined behavioral/explanatory study allowed us to underscore the previously reported discrepancies between near-practice situations and offline consideration of warnings [18], furthering the point that warnings always need to be tested in carefully crafted, realistic study designs as ours.

7.2 Contextualization helps in risk assessment – and in understanding the risk

However, independent of the problems with creating attention for the warning content and with motivating users to consider the content, the content needs to be optimally understandable. Concerning this goal, we find, based on the qualitative data, that participants reasoned about risks more concretely and less technically or abstractly than the control group (cf. Section 6.2). In particular, the reasoning depended to a lesser degree on the prior knowledge about threats or prior personal experiences of adverse events. Since experts in our pre-study were similarly more concrete in their risk assessments than lay users, we see the changed reasoning as an indication that contextualized content caused a better understanding of the risk of proceeding. Our findings thus extend prior research – for example, by Kauer et al. [13] – that showed personal consequences as more effective in warnings.

7.3 Building trust in contextualized warnings

Moreover, we identified problems that participants encountered due to the contextualization of the content in Section 6.3, particularly related to trust in the warning. The complex interrelation between user characteristics (such as expertise), the concreteness of content, understanding, and trust in the warning warrants a closer look at the problems and how we can address them. Focusing on the results from the explanatory part of the study (card sorting and interview), we need to extend the model from Section 4 that our hypotheses were

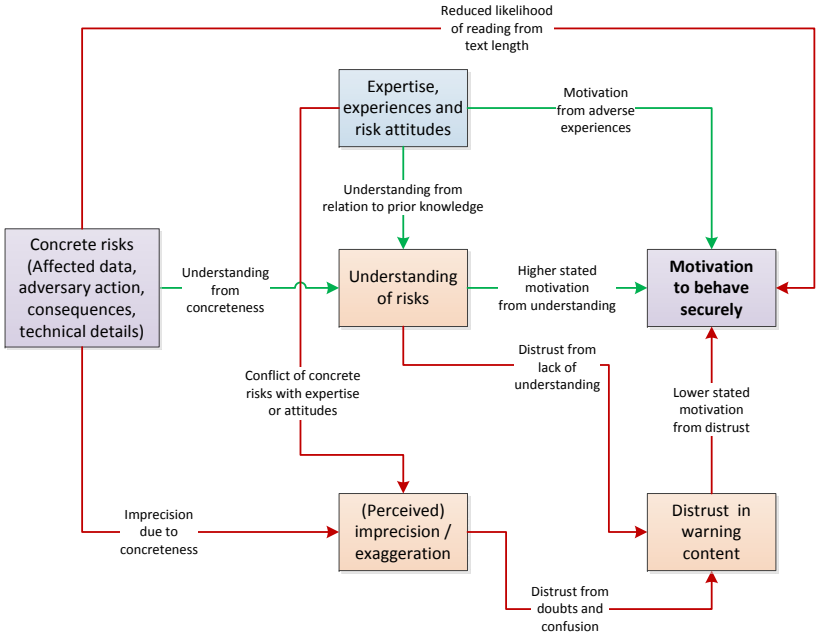

Fig. 5. The effects of concreteness and contextualization (edges represent effects; green for positive, red negative)

based upon. The extension of the model in Figure 5 shows how the concreteness of the warning content has – for some participants – negative effects. One such effect is distrust as shown in the quote in Section 6.3.

Specifically, the extended model still describes how mentioning concrete risks (affected data, consequences, ...) leads to a higher level of *understanding of the risks* from proceeding and thereby motivate users to follow the recommendation laid out in the warning content (*Motivation to behave securely*). Our results indicate that the *expertise, prior experiences, and risk attitudes* of the user play an important role in the understanding and the motivation. However, the extended model now also shows that concreteness leads to *distrust of the warning content* if the more concrete information is not understood (*Understanding of risks*) or if *(perceived) imprecision or exaggeration* in the content raises doubts. For example, depending on which risks are considered problematic by the participant (*risk attitudes*; e.g. only financial consequences, not so much social-privacy consequences), mentioned consequences were considered exaggerated. Conflicts of the content with the user's expertise can have similar effects. From the participant's comments in the study, we expect that distrust will also reduce the motivation to follow the recommendation from the warning content.

Thus, our results indicate several negative side effects from the content's concreteness. We conclude that to realize the positive effects of increased con-

creteness without compromising on other factors (e.g. the trust in the warning content), individualization for the user is necessary: For instance, people with higher expertise need different content – for example, less concrete consequences, so as to not raise doubts about the given information – than people with lower expertise.

While it has been found before that trust plays an important role in the behavior of users when confronted with security-critical situations, prior research has focused on the trust in the Web site [10, 21, 9]. Krol et al. [14] also mentioned as one conclusion of their study that the trust in the warning needs to be restored, but they addressed the habituation effects from over-frequent warnings in non-critical situations. Our results and the derived model go beyond those findings by addressing the trust in the warning as affected by the warning content.

The extended model is foremost based on the qualitative and subjective data from a relatively small sample of 36 participants. Therefore, the extended model should primarily serve as a hypothesis for further studies on the contextualization of content with larger and more representative samples that we are planning as future work, particularly on the individualization of warnings between lay users.

Acknowledgments

The work presented in this paper is supported by funds of the Federal Ministry of Food, Agriculture and Consumer Protection (BMELV) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support programme.

We thank Michaela Kauer and Christoph Seikel for their support on designing and conducting the pre-study.

References

1. Adams, A., Lunt, P., Cairns, P.: A qualitative approach to HCI research. Cambridge Univ. Press, Cambridge (2008)
2. Amer, T., Maris, J.: Signal Words and Signal Icons in Application Control and Information Technology Exception Messages – Hazard Matching and Habituation Effects. Tech. Rep. 06-05, Nothern Arizona University (2006)
3. Asgharpour, F., Liu, D., Camp, L.J.: Mental Models of Computer Security Risks. In: WEIS '07: Workshop on the Economics of Information Security (2007)
4. Bartsch, S., Volkamer, M.: Towards the Systematic Development of Contextualised Security Interventions. In: Proceedings of Designing Interactive Secure Systems, BCS HCI 2012. BLIC (2012)
5. Biddle, R., van Oorschot, P.C., Patrick, A.S., Sobey, J., Whalen, T.: Browser interfaces and extended validation SSL certificates: an empirical study. In: Proceedings of the 2009 ACM workshop on Cloud computing security. pp. 19–30. CCSW '09, ACM, New York, NY, USA (2009)
6. Bravo-Lillo, C., Cranor, L.F., Downs, J., Komanduri, S., Sleeper, M.: Improving Computer Security Dialogs. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) Human-Computer Interaction – INTERACT

2011, vol. 6949, pp. 18–35. Springer Berlin Heidelberg, Berlin, Heidelberg (2011), http://www.springerlink.com/content/q551210n08h16970
7. De Keukelaere, F., Yoshihama, S., Trent, S., Zhang, Y., Luo, L., Zurko, M.: Adaptive Security Dialogs for Improved Security Behavior of Users. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R., Winckler, M. (eds.) Human-Computer Interaction – INTERACT 2009, Lecture Notes in Computer Science, vol. 5726, pp. 510–523. Springer Berlin / Heidelberg (2009)
8. Downs, J.S., Holbrook, M.B., Cranor, L.F.: Decision strategies and susceptibility to phishing. In: SOUPS '06: Proceedings of the second symposium on Usable privacy and security. pp. 79–90. ACM, New York, NY, USA (2006)
9. Egelman, S., Cranor, L.F., Hong, J.: You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In: CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems
10. Fogg, B.J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., Treinen, M.: What makes Web sites credible?: a report on a large quantitative study. CHI '01, ACM, New York, NY, USA (2001)
11. Glaser, B.G., Strauss, A.L.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine Transaction (1967)
12. Kahneman, D., Tversky, A.: The simulation heuristic. Cambridge University Press, Cambridge, MA, USA (1982)
13. Kauer, M., Pfeiffer, T., Volkamer, M., Theuerling, H., Bruder, R.: It is not about the design – it is about the content! Making warnings more efficient by communicating risks appropriately. In: GI SICHERHEIT 2012 Sicherheit – Schutz und Zuverlässigkeit (2012)
14. Krol, K., Moroz, M., Sasse, M.: Don't work. Can't work? Why it's time to rethink security warnings. In: 7th International Conference on Risk and Security of Internet and Systems (CRiSIS). pp. 1 –8 (Oct 2012)
15. Lazar, J., Feng, J.H., Hochheiser, H.: Research methods in human-computer interaction. Wiley (2010)
16. Rothman, A.J., Kiviniemi, M.T.: Treating People With Information: an Analysis and Review of Approaches to Communicating Health Risk Information. J Natl Cancer Inst Monogr (25) (1999)
17. Rugg, G., McGeorge, P.: The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. Expert Systems 14(2), 80–93 (1997)
18. Sotirakopoulos, A., Hawkey, K., Beznosov, K.: On the challenges in usable security lab studies: lessons learned from replicating a study on SSL warnings. In: SOUPS '11: Proceedings of the 7th Symposium on Usable Privacy and Security. ACM, New York, NY, USA (2011)
19. Sunshine, J., Egelman, S., Almuhimedi, H., Atri, N., Cranor, L.F.: Crying Wolf: An Empirical Study of SSL Warning Effectiveness. In: USENIX Security 2009 (2009)
20. Wogalter, M.S.: Handbook of warnings. Routledge (2006)
21. Wu, M., Miller, R.C., Garfinkel, S.L.: Do security toolbars actually prevent phishing attacks? In: CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems. pp. 601–610. ACM, New York, NY, USA (2006)