Don't Work. Can't Work? Why It's Time to Rethink Security Warnings

Kat Krol
Department of Computer Science
Security and Crime Science Doctoral Research Training Centre (SECReT)
University College London
Email: k.krol@cs.ucl.ac.uk

Matthew Moroz
Department of Computer Science
Security and Crime Science Doctoral Research Training Centre (SECReT)
University College London
Email: m.moroz@cs.ucl.ac.uk

M. Angela Sasse
Department of Computer Science
University College London
Email: a.sasse@cs.ucl.ac.uk

***Abstract*—As the number of Internet users has grown, so have the security threats that they face online. Security warnings are one key strategy for trying to warn users about those threats; but recently, it has been questioned whether they are effective. We conducted a study in which 120 participants brought their own laptops to a usability test of a new academic article summary tool. They encountered a PDF download warning for one of the papers. All participants noticed the warning, but 98 (81.7%) downloaded the PDF file that triggered it. There was no significant difference between responses to a brief generic warning, and a longer specific one. The participants who heeded the warning were overwhelmingly female, and either had previous experience with viruses or lower levels of computing skills. Our analysis of the reasons for ignoring warnings shows that participants have become desensitised by frequent exposure and false alarms, and think they can recognise security risks. At the same time, their answers revealed some misunderstandings about security threats: for instance, they rely on anti-virus software to protect them from a wide range of threats, and do not believe that PDF files can infect their machine with viruses. We conclude that security warnings in their current forms are largely ineffective, and will remain so, unless the number of false positives can be reduced.**

I. Introduction

People who spend time online routinely encounter security warnings. While trying to achieve their work or personal goals, users are frequently disrupted by prompts, adverts, chat windows, pop-ups and alerts – a negative user experience.

As the number of Internet users has grown, so has the number of security threats and the sophistication of attacks [15]. The protection against those comes in various guises, but security warnings are a common way of alerting users to threats, and preventing them from unsafe acts. Thus, security warnings have become part of the large number of disruptive alerts users encounter today – but do they pay attention to them, and even if they do so – do they follow their advice?

We conducted a study to determine if users heed security warnings – and if not, why not. Our participants were presented with PDF download warnings in a naturalistic context – they were invited to bring their own laptops to a usability test of a new academic article summary tool.

In the recent years, PDF (Portable Document Format) files have become one of the most dangerous file types [9]. To warn about this threat, browsers, notably Google Chrome, introduced warning messages into download dialog boxes (Fig. 1).

Fig. 1. Google Chrome's PDF download warning.

But Google Chrome users registered their frustration in forums: "I am not going to 'cope with the warning'. I am switching back to FF [Firefox Mozilla]. Working in publishing, I need to download many PDFs every day from trusted sources. I can't add extra steps without affecting profitability" [12]. Subsequently, Google Chrome version 10 changed the algorithm making a warning show only if there was an indication that this might be a drive-by download not initiated by the user. Google Chrome's policy of frequent warnings provided inspiration for our research – the number of warnings has been steadily increasing, and given that users dislike disruptions to their activities, it is questionable whether this is effective.

We present the results of a laboratory experiment with 120 participants, who responded to a recruitment ad for an evaluation of an academic article summary tool, named 'Summarix'. Participants were asked to test the tool on their own laptops. The actual purpose of the study was to see participants' reaction to a PDF download warning they encountered when downloading a paper in one of the two test cases. We maximised ecological validity by making security the secondary task as it is in users' everyday interactions, and by having them put their own laptops at risk.

Our results show that – whilst nearly all participants noticed the download warning, the vast majority (81.7%) ignored its advice. The level of detail provided in the warning did not make a difference. Female participants were significantly more cautious than male. The post-experiment interviews revealed that our participants mainly ignored the warnings because they have been desensitised by the high number of security warnings, most of which turn out to be false positives.

The paper is organised as follows. First, we review previous studies on warnings. Second, we describe the design of the experimental study, and the debriefing interviews that followed. Third, we present and discuss the findings and limitations of our study. In the final section, we make suggestions on how to overcome desensitisation and make warnings

work. We suggest that algorithms triggering warnings need to become more accurate, to reduce the number of warnings users experience. Warnings also need to communicate the threat and consequences more accurately, to correct current misconceptions.

II. Related Work

A. Dismissal due to False Positives

Previous research on security indicators and warnings shows that users often ignore them and choose to rely on their own assessment of online risks instead. Wu et al. [33] found that participants disregarded toolbar warnings even when they were explicitly asked to pay attention to them. Users distrust a toolbar once they have experienced false positives, for example, if the tool warns them against a Website they know to be legitimate. Fogg [10] describes this as an earned lack of credibility.

B. Phrasing of Warnings

When it comes to the text displayed in warnings, there seems to be a trade-off between providing enough detail but being brief in order not to discourage users from reading. Zeltser [34] advises warning designers to include enough detail to help users make an informed decision, but at the same time, to be brief and avoid technical jargon. Egelman et al. [8] showed that users are more likely to ignore security warnings that use technical jargon that is hard to understand.

C. Immersion in the Primary Task

Sharek et al. [27] conducted research on whether users can distinguish fake pop-up warnings from real ones. Many participants reacted wrongly to the stimulus pop-up and 42% later explained that they just wanted to "get rid of the message" and complete their task. Similarly, in a study on phishing, 45% of participants stated that they ignored security warnings as they were absorbed by their task and "had to take some risks to get the job done" [33]. Sasse et al. [24] point out that user behaviour is driven by their primary goal. If the users perceive likely losses due to security issues to be less than losses they would incur by not having completed the task, they will ignore security advice to complete their task.

In many laboratory studies of security tools, participants knew that the experiment was observing their security-related behaviour. For instance, Wu et al. [33] made their participants focus their attention on completing a task, nevertheless, they told them in the briefing that "the purpose of the study was to test web browser security indicators that detect fake web pages that look like pages from well-known legitimate websites". This seriously threatens the validity of the results [8]. A study cannot be ecologically valid if participants do not behave as they would in the real world [25].

D. Real Risk

Sharek et al. [27] acknowledged the limitation that in their study the security of the laboratory computer was at stake and not the participants' own computers. Sheng et al. [28] mention a similar limitation to their study on phishing. They state that participants might have been more willing to take risk as they might have felt immune to any adverse consequences of their actions. This is a serious limitation since decisions participants take in an experiment are only generalisable when they are led to believe they are really at risk [8].

In their study on phishing, Schechter et al. [25] showed that those participants who used their own credentials were far more security-vigilant than those in a role-playing condition who were given login details from the experimenters and therefore experienced no direct personal threat from their actions.

III. Research Hypotheses

A. Specific vs. Generic Warning

In this study, we investigate to what extent users do pay attention to warnings, and whether not paying attention is due to warnings being too generic. Sunshine et al. [29] have argued that current messages do not convey in sufficient detail what can happen in a security risk, and how likely it is to happen, and suggest that user behaviour can be improved through warnings in which this information is given. We tested this suggestion to see whether increasing the seriousness of a warning by naming possible consequences and mentioning the probability would influence the likelihood of a PDF being downloaded. We hypothesise that participants will be largely desensitised towards the generic warning because of every-day exposure and will be more likely to ignore it ($H\,I$).

B. Demographic Characteristics

We also wanted to monitor if different types of people react differently to the warnings – based on their gender, browser use, computer literacy and virus exposure. There is a large body of research exploring risk appetite in men and women (e.g., [3]), and results show that women are more risk-averse than men. Thus, we hypothesise that female participants in our experiment will be less likely to download an article with a warning than males ($H\,II$). We hypothesise that existing users of Google Chrome will be more likely to ignore the warning than users of Firefox Mozilla, due to previous exposure ($H\,III$). We also control for characteristics relating to computer experience, looking at whether those with varying computing skills and virus experience differ in their reactions to a warning. We hypothesise that those with a high level of computer expertise will be more likely to recognise the warning as being a false positive, and download the article ($H\,IV$). We hypothesise that those who have had negative experience with viruses, scams and fraud will be less likely to download an article with a warning ($H\,V$).

IV. Study Description

A. Participant Recruitment and Demographics

Participants were recruited through the UCL Psychology Subject Pool. We advertised for participants in a usability study focussing on an academic article summary tool. There were four requirements for participation. (1) Participants

needed to be at least 18 years old. (2) They were required to bring their own laptop to the study. (3) The laptop needed to run Windows OS (since UNIX and Mac OS differ in security and HCI). (4) Participants had to be fluent in English (they must have lived in an English-speaking country for at least a year). Requirement 2 was explained as being necessary because the study wanted to check tool display and performance on different types of laptops. Requirement 4 had to do with the assessment of the quality of the summary produced by the tool. The test took place in a usability laboratory in the UCL Department of Computer Science. Participants received a fee of £14 for their participation in the study lasting for about 50 minutes. 125 participants took part, 5 were later excluded because of incomplete data. 53.3% (64) participants were female and 46.6% (56) male. Age ranged from 19 to 52 years (M=25.7, SD=6.1, variance=37). As their main browser, 44 (36.7%) participants reported using Google Chrome, 40 (33.3%) Firefox Mozilla, 33 (27.5%) Internet Explorer and three used other browsers. Participants' length of computer use ranged from 5 to 30 years (M=13.6, SD=4.6, variance=21.1).

B. Design

The experiment was a between-subjects design. The participants were randomly assigned to two conditions, they were either shown a generic or a specific warning (Fig. 2 and 3). While the text of the generic warning was vague (directly taken from the Google Chrome download dialog box), the specific warning mentioned that some malicious elements had been detected within the PDF, that it could cause massive damage to the participant's computer and assessed the probability of it happening as high.

Fig. 2. Generic warning.

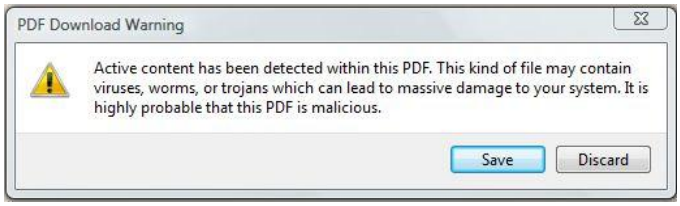

Fig. 3. Specific warning.

C. Materials and Apparatus

During the recruitment process, participants were required to fill out an online form asking them for demographic details (gender, age), their reading habits (to keep up the appearance of a study on an academic article summary tool) and their computing skills. Upon arrival, participants were given an information sheet and an informed consent form.

In the semi-structured interview, participants were asked three questions: (1) Whether they could identify which warning they saw from a set of four (two original warnings and two similar distractors). (2) What they would have done if they were confronted with any of the remaining three. (3) Whether they have had any experience with scams, viruses or fraud.

An eye-tracker Tobii x50 was used to gather data on participants' eye movement during the computer-based task and to produce a screenrecording which the experimenters and the participant watched afterwards. The main aim of having an eye-tracker was to establish how much attention participants paid to the warning and whether they read it. Tobii Studio 2.0.4 software was used for data analysis. The laboratory computer ran Windows XP, the screen size was 22" with a resolution of 1920×1080 pixels. The reading task placed in a browser simulator was designed using the Python programming language.

V. Procedure

Upon arrival, the participant was asked to read through an information sheet and sign a consent form. The experimental set-up was explained to the participant and they handed over their laptop to the first experimenter. The first experimenter connected it to wireless Internet, established an UltraVNC connection with the lab computer and placed the Summarix icon on the participant's desktop. It was not possible to conduct the entire study directly on participants' computers due to the eye-tracker which was installed on the laboratory's computer. A briefing sheet was then given to the participant and they read it. Finally, the task was explained to the participant using a task list with bullet points.

For the task, the participant sat at the lab computer and they could see the desktop of their laptop on the screen of the lab computer. The experimenters pointed to the Summarix icon placed there. Then, the eye-tracker was calibrated. Upon starting of the eye-tracking, the task appeared in the browser simulator. At this stage, the participant was reminded of the time limit and the experimenters left the room.

Once the participant had finished, the experimenters re-entered the room. Together with the participant, they watched the screenrecording of the task. The experimenters asked usability questions relating to Summarix, consistent with the description of the study on a summary tool. When the warning appeared, the second experimenter asked: "Oh, what was that?" and looked at the participant waiting for their interpretation of what happened. Then the experimenter asked: "So what did you do?" to elicit some more explanation. Afterwards, they continued watching the screenrecording, and the experimenters asked further questions relating to the tool.

After the screenrecording, a semi-structured debriefing interview followed which was audio-recorded. Each participant was asked the same questions; however, if they were willing to share any other computer security-related experience, they were welcome to do so. Once the interview was finished,

the participant was fully debriefed in line with UCL's ethics guidelines. The experimenters stressed that the participant's computer was not harmed in any way. The participant could ask any questions. Finally, the participant received payment.

VI. Results

A. H I: Reaction to Warnings

In both conditions, 98 (81.7%) participants downloaded an article with a warning, and only 22 (18.3%) participants refused to do so. For generic warning, 52 (86.7%) participants decided to download the article and 8 (13.3%) refused to do so. For specific warning, 46 (76.7%) participants download the article and 14 (23.3%) refused to download it (Table I). Despite fewer downloads for the specific warning, this difference is not statistically significant (p=0.24).

Warning	Refusal	Download	Total
Generic	8	52	60
Specific	14	46	60
Total	22	98	120
	$\chi^2 = 1.4$, $p = 0.24$, $df = 1$		

TABLE I
Reactions by warning type: download vs. refusal to download

B. H II: Gender

Regardless of warning type, of those who refused to download an article with a warning, 16 were female and 6 were male. This difference was found to be statistically significant (p=0.048). When split by warning type, the difference was not statistically significant – not surprising given the small counts.

C. H III: Browser Use

Among those 22 participants who did not download an article with a warning, 8 used Google Chrome as their main browser, 7 Mozilla Firefox and 7 Internet Explorer. These numbers are proportional to the numbers of participants who reported using the browsers and there is no statistically significant relationship between browser use and reaction to warning.

D. H IV: Computing Skills

We explored whether participants with different levels of computer literacy differed in their reaction to a warning. A score for computing skills was computed based on five questions. Two of them were direct: (i) "How many years have you been using computers?" (ii) "How would you rate your level of computing skills?" and three were proxy-questions: (i) "Have you ever designed a Website?" (ii) "Have you ever registered a domain name?" (iii) "Have you ever configured a firewall?" For the length of computer use, the data was split into quartiles based on the distribution. Four categories were formed of those who have been using computers for (1) five to ten years, (2) eleven to thirteen years, (3) fourteen to fifteen years and (4) sixteen to thirty years. Corresponding to this, participants were given scores from 1 to 4. The self-rating of computing skills was on a five-point scale from 'very low' to 'very high'. No participant stated that their skills were 'very low', thus, the levels 'low' and 'average' were grouped together and 'high' and 'very high'. A score of 0 was given for the first and 1 for the second. In the three proxy-questions, a score of 2 was given for each 'yes' and 0 for each 'no'.

The scores for computer literacy ranged from 1 to 11. The mean score for computing skills for those who downloaded an article was 4.7 and for a person who refused to download it 3.2. This difference was found to be statistically significant in a t-test (p=0.005). The higher the level of computer literacy, the more likely were the participants to download an article with a warning.

E. H V: Virus, Scam and Fraud Experience

Based on interview data, each participant received a score between 0 and 3 for their virus, scam and fraud experience (the higher the score, the more significant the experience). The score was assigned by two researchers independently and in the case of any discrepancies, a structured discussion helped them to identify the appropriate score. To avoid bias, the two researchers listened to interview recordings and did their rating not knowing how any of the participants reacted to the warnings. Participants who refused to download an article with a warning had a higher mean value for virus, scam and fraud experience (2.32) than those who downloaded the article (1.78). This difference was found to be statistically significant in a t-test (p=0.037).

F. Hypothetical vs. Actual Response

In the beginning of the interview, we asked participants to identify the warning they saw in the study from a set of four. After they had identified it, we asked them what they would have done with the remaining three warnings. Based on their answer, a comparison was made between participants' actual behaviour and what they claimed they would have done. In the case of the generic warning, 19 (32%) of participants who were not confronted with this type of warning in the task said they would not have opened the file. However, out of those who were confronted with that specific warning, only 8 (13%) refused to download the file. The difference between self-reported and observed behaviour was even more pronounced for the specific warning, here 47 (77%) participants who did not get this kind of warning in the experiment stated they would not have downloaded the file. In the experiment, only 14 (23%) participants in the generic warning condition refused to download the file.

Discrepancies between self-reported and observed behaviour have received considerable coverage in privacy research [4], [14]. In our study, we attribute them to the social desirability bias due the fact that in the interview participants started to realise that the study was on security rather than a summary tool. When asked directly how they would react to a warning, the participant may be giving an answer they believe is 'correct' and expected from them. Of course, there are other factors that made participants say they are more

attentive to warnings than they actually are, one of them is the fact that they were directly asked for what their reaction would be by the experimenters. Nevertheless, it shows that without the deception element in the study the results might have been different.

G. Attention Paid to Warning

Eye-tracking data was analysed to establish the total time of how long a participant focussed on the first warning. Two t-tests were conducted to establish whether the differences between warning and reaction types were significant. Regardless of whether the participant decided to open or discard the chosen article, participants fixated their gaze for an average of 6.13 seconds for generic warnings, and 6.33 for specific warnings (p=0.9). When considering reaction type, participants fixated their gaze on the warning (specific or generic) for an average of 6.94 seconds if they ultimately refused to download the article, and for 5.63 if they downloaded it (p=0.39).

The long fixation times which indicate that participants read the text of the warning are supported by results from the semi-structured interviews. Here, 107 participants correctly identified the warning they had seen within the task, suggesting that participants did pay attention to the warnings, regardless of warning type and their subsequent reaction.

H. Reasons for Reaction

In the interviews, we asked participants why they downloaded the article despite a warning. The answers were transcribed and analysed. For each participant, at least one reason was identified through a thematic analysis [2], although in many cases the participants mentioned more than one. They were free to do so and all reasons were taken into account. The following categories of rationale were identified:

1) Desensitisation: 55 of our 120 participants mentioned desensitisation to warnings as a reason for disregarding them. Participant 115, for instance, said: *"I thought the warning was just a routine thing"*. 12 participants explicitly stated that they realise that most warnings are just false positives: *"I decided to open it because the pop-up came for all of the articles. So I thought it just must be a like a generic thing. Sometimes the nature of the PDF itself will produce it even if there is nothing wrong"* (P005).

2) Dismissal of a Pop-up: 12 participants stressed that they did not take the pop-up warning seriously because they associate pop-ups with annoying adverts and updates: *"I didn't think it was a serious problem, it was just like a pop-up, something you get all the time"* (P003).

3) Trusted Source: 29 participants mentioned that they perceived the laboratory setting as a trusted environment. They said that similarly, they would have trusted their friends or lecturers that they are not letting them download anything that is malicious. Participant 017 stressed: *"It depends on what the source was, if I was getting it from a dodgy Website, I probably wouldn't download it. But if something was sent to me by a friend or a lecturer or I was downloading it from a library catalogue, I would have opened it anyway"*.

4) Focus on the Task: 20 participants said that being focussed on their reading task made them want the warning to disappear: *"I just wanted to get to the task"* (P021).

5) Trust towards Anti-virus: 18 participants mentioned feeling protected by their anti-virus software: *"I trusted that the anti-virus on my computer would pick anything up"* (P026).

6) Trust towards PDF: 15 participants said they trusted PDF as secure file format and dismissed the warning: *"I don't think PDF files can have this kind of harm in them"* (P087). Two reasons were mentioned for the trust – the omnipresence of the files and their appealing aesthetics. Again participants judged trustworthiness by cues that are salient in the real world but not necessarily in the online world [16].

VII. Discussion

81.7% of participants opened the file regardless of warning type. From the observed behaviour and responses in the debriefing interviews, we can infer that the warning text does not matter, because most of our participants had been conditioned just to 'swat away' warnings. Additional reasons given for ignoring the warnings were trust in the researchers who had set up the experiment, reliance on anti-virus tools, and lack of awareness of the risks associated with PDFs. We found that those who did not download a PDF with a warning were mostly women, participants with lower levels of computing skills, and those who had experienced viruses, scams or fraud.

A. Problems Associated with Warnings

Pop-up warnings of the kind we used in our experiment ignore basic principles of usability: as Cooper [5] would put it, pop-up warnings in dialog boxes are impolite, disrupting users' primary task. It is basic HCI knowledge that warnings should be reserved for genuine exceptions, and only ask users to make decisions they are capable of making.

Current warnings are also hardly distinguishable from other pop-up windows which users are conditioned to 'swat away' such as adverts and updates. Figure 4 shows a download dialog box from Internet Explorer and Figure 5 shows Internet Explorer suggesting the user disables some add-ons to speed up the browsing. These pop-ups have different functions and the urgency with which they need to be dealt with differs, nevertheless, they have exactly the same design. They both look very elegant, nevertheless, they might be quite confusing to the user. As West [32] stresses, security messages should be instantly distinguishable from other dialogs and this principle is clearly violated here. Furthermore, both pop-ups appear at the bottom of the screen and this is likely to cause confusion in the beginning and increase user effort since the gaze does not fall there [20]. Dhamija et al. [7] warn that indicators placed outside of the user's focus of attention are likely to be ignored.

Messages like "This type of file can harm your computer. Are you sure you want to download?" shift the focus from the software being unable to detect potential malicious elements to putting pressure on the user. However, Google Chrome is not alone here, the wording of the download warning in Opera is problematic too, it states: "It is not known whether this type of

Fig. 4. Internet Explorer's download dialog box.

Fig. 5. Internet Explorer suggesting disabling add-ons to speed up browsing.

file is safe. Are you sure you want to download it?" (Fig. 6). By using the passive rather than the active voice ("it is not known" vs. "we don't know"), it takes away the focus from the browser not being able to check the file for malicious elements to the user who has to be sure about the download.

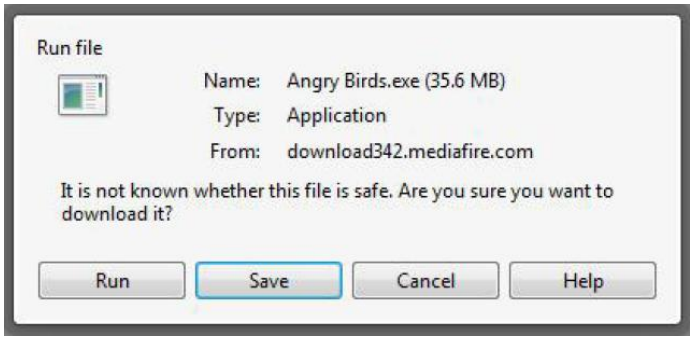

Fig. 6. Opera's warning text in a download dialog box.

VIII. Folk Models of Internet Security

In our experiment, we found that those with high level of computer expertise were less likely to heed a warning. One might think that they rejected the warning as they knew it was a generic one. In the post-experiment interview however, the participants revealed a number of misconceptions.

Based on Wash [30], we identified the following folk models (inaccurate mental representations of the real world which may lead to erroneous decision-making [6]) of Internet security in our interview data.

A. The Anti-virus Will Save the Day

18 participants downloaded an article with a warning because they trusted their anti-virus to deal with any problems. Participants did not realise that anti-virus software may not be able to cope with all kinds of threats, such as zero-day attacks and sophisticated malware (e.g., rootkit).

B. Viruses Have an Instant Effect

We also saw outdated knowledge regarding viruses, with participants expecting an immediate and visible effect if their computer contracted a virus: *"I saved it and then nothing bad happened. If it was really bad and I downloaded a virus to my computer, I would have expected like the 'blue screen of death' "*. In the past, some hackers left images or animations on the victims' computers in order to cause unrest. Today, the longer a hacker can stay undetected with their malware or spyware, the longer they can extract the victim's resources.

C. Trust towards the Source

29 participants said they decided to download the article because they trusted the researchers not to make them download malicious files. They mentioned that they would trust friends and lecturers in a similar way. There are two kinds of problems here. First, participants did not realise that even the people they trust might not themselves be aware of the malicious elements in their files. Second, they did not realise the risks of impersonation: email accounts can be hacked, especially if users have poor password practices and the industry has low standards [1], [21]. Third, it is possible to create a plausible email address and send out messages to people the targeted user knows using information from social networking sites.

D. PDFs are Safe

15 participants expressed a great deal of trust towards PDF files, particularly due to their omnipresence and professional look – we found users are largely unaware of the risks associated with this widely used filetype. The reason for this might be that the mass use of PDFs as attack vectors is a fairly recent phenomenon.

E. "Not Me, not this Time"

Seven participants stressed that they did not believe something bad would happen to them or this time. This is a psychological mechanism already observed by West [32]. Users don't believe that they are a target, because they have nothing valuable on their computer or only little money in their bank account. This is an erroneous belief first identified by Weirich and Sasse [31], our findings show that this misconception of attacks being personally targeted persists. In the case of an untargeted attack, criminals are not after some valuable data but just after computing power, they want new machines for their botnet to launch large scale attacks. In the case of financial fraud, the attackers are usually not after big sums, but want to steal a little from many people as small transactions are less likely to draw attention [19].

IX. Conclusions

A. Why Do Users Ignore Warnings?

81.7% of participants in our study ignored warnings, and most of those who did justified doing so with previous

experience of online security warnings being false alarms. In this situation, users' dismissal of a download warning is what Herley [13] describes as rational rejection of security advice: the effort of reading a warning and the cost associated with not downloading a file are much greater than the perceived likelihood of harm from a virus. Additionally, the probability of an attack happening is relatively low. Schneier [26] explained how experience of false alarms desensitises defenders and enables attacks; the way in which the vast majority of our participants responded provides yet another example.

Those with higher level of computing skills ignore the warnings more often than those who are not confident in their computing skills – meaning users with more experience of current security warnings pay less attention to them. This is hardly a ringing endorsement for the effectiveness of current warnings. But our interviews show that the confidence of experienced users is not matched by knowledge.

As the results of our experiment indicate, participants rely on their own judgment, rather than a security warning. This aligns with previous results which showed that participants judged the legitimacy of a Website by how it 'looked and felt' [11], [33] and interpreted signs as trust signals that are not valid in the online environment [16], [23] rather than used other more salient cues (e.g., URL).

B. How to Make Warnings Work?

As the responsibility for ignoring warnings was identified on both sides, the users' and the industry's, attempts to overcome desensitisation and protect users from malicious PDF files need to be made in two directions.

C. Restore Trust in Warnings

Users need to be re-sensitised. As long as there is a high number of false positives, users will not heed warnings. For that reason, it is necessary to significantly limit the appearance of warnings. To achieve this, browsers should give up on the idea of warning about possible harm to the user's computer each time they want to download a file. Sophisticated and intelligent algorithms need to be developed to diminish the number of false positives. This might be time and resource-consuming, and needing frequent updates; nevertheless, it might turn to the browser's advantage over competitors.

The Venn diagram in Figure 7 shows when warnings should be used. At any other time they are redundant.

Case 1. *Genuine concern of significant danger, with certainty of maliciousness.* Here the warning is redundant as the action must simply be blocked without warning to protect the user.

Case 2. *No certainty of maliciousness, and no genuine concern of significant danger.* Here the warning is redundant and if shown is simply an expression that "We have no real clue, but we will interrupt your work flow anyway".

Case 3. *The intersection.* Here is the only situation whereby a warning can be justified. The case where genuine danger has been detected within, but there is still a chance that the file in question is not malicious. Rather than simply block, there must be an option to still download it.

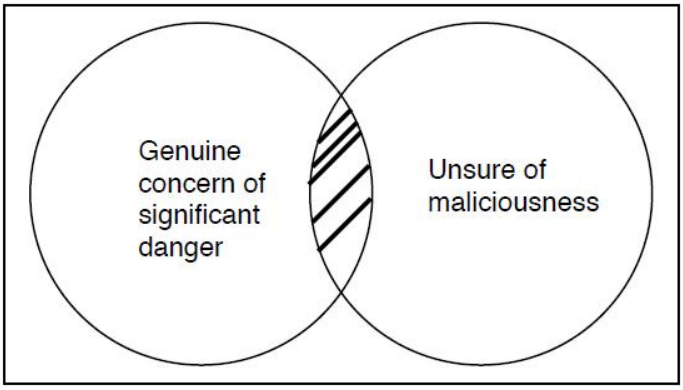

Fig. 7. A Venn diagram illustrating when users should receive a warning message.

Only displaying warnings when the intersection is true would cut down the number of warnings faced and would in turn make them less likely to be swatted away and instead read and contemplated. Assuming we could get to this point, where the sophistication of detection improves enough to support such genuine warnings, the question remains as to what the user should do next and whether they at this point have enough information to make the correct decision. If we consider the fact that even the most experienced and diligent users are still fallible and that for every thousand correct decision it only takes one wrong one to ruin all the good work, we must conclude that the user is still ill-equipped to deal with these decisions with consistent correctness.

We conclude that in general, the user should not be the one left responsible. In a corporate situation security specialists could instead be in charge of making these decisions, the company's security being their primary task, whereas other users in the company are instead fixated on their goals. Elsewhere ISPs must do more to take the burden away from the user. What is of utmost importance is getting as far away from the current warning culture as possible. The number of false positives means no added security, and hands a perfect platform to attackers. Users will either ignore warnings as they do not trust them or if able to they will disable the warning systems completely.

D. Provide Education and Training

The poor state of participants' knowledge on online security is worrying. Education is urgently needed; first, the users need to be made aware of the fact that PDF files can carry malicious elements and second, that they can lose a lot if they behave carelessly online. However, the place for education should not be in the dialog box when the users are in the middle of their critical activities, but elsewhere. Several participants stressed that they would welcome some guidance or even training from the ISPs or the state. As shown in studies on phishing, education can have a significant impact on the participants being able to recognise a fake Website from a real one [17], [18], [28]. However, while for phishing users can perform some checks themselves (looking at the URL, SSL etc.), it

is hard for them to do the same for PDF files. Therefore, research is needed into what kind of security checks can be performed and which ones would be the least time-consuming and disruptive ones for the users.

E. Limitations

There are a number of limitations to our study that need to be acknowledged. First, most participants were students and rarely older than 35. For that reason, the findings might not be generalisable to different age groups in the wider population. Second, there was a problem of the participants trusting their computer would not be harmed in an experiment carried out at a university. Third, the experiment focussed only on one single reaction of a participant which might not be representative of their general behaviour.

X. Future Work

The limitations of our study open space for future work. Future research should aim to study users' reactions to browser warnings in a more comprehensive way, considering security and privacy threats alike. Unobtrusive, privacy-aware monitoring of participants' computers could record their behaviours towards warnings over a longer period of time. This measure would ensure the reactions are elicited in a natural setting, in different contexts, and record reactions to a much wider variety of warnings and alerts. As a cost-efficient alternative, crowd-sourcing could be used as a testbed [22]. Field experiments have the additional benefit of age and occupational variety, which could influence how people react to warnings. Parallel studies could be conducted in different environments or contexts, to see whether the reactions would be different than they were in a university setting.

Acknowledgement

We would like to thank Nigel Harvey, David Clark, Sacha Brostoff, Amin Amiri and John Atkinson for their help in the study. The study was funded through EPSRC's grant to the Security Science Doctoral Training Centre (EPSRC Grant no: EP/G037264/1).

References

[1] Bonneau, J., and Preibusch, S. The password thicket: Technical and market failures in human authentication on the web. In *WEIS* (2010).

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

[3] Byrnes, J., Miller, D., and Schafer, W. Gender differences in risk taking: A meta-analysis. *Psychological Bulletin 125*, 3 (1999), 367.

[4] Connelly, K., Khalil, A., and Liu, Y. Do I do what I say?: Observed versus stated privacy preferences. In *INTERACT* (2007), pp. 620–623.

[5] Cooper, A. 14 Principles of Polite Apps. *VBPJ* (1999), 62–66.

[6] D'Andrade, R. *The development of cognitive anthropology*. Cambridge University Press, 1995.

[7] Dhamija, R., Tygar, J. D., and Hearst, M. Why phishing works. In *CHI* (2006), pp. 581–590.

[8] 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 *WEIS* (2010).

[9] F-Secure. PDF most common file type in targeted attacks. http://www.f-secure.com/weblog/archives/00001676.html, 2009. Accessed 04.02.2011.

[10] Fogg, B. *Persuasive Technology: Using Computers to Change What We Think and Do (Interactive Technologies)*. Morgan Kaufmann, 2002.

[11] Fogg, B., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., et al. What makes Web sites credible?: A report on a large quantitative study. In *CHI* (2001), pp. 61–68.

[12] Google Chrome Forums. How do I stop Chrome from warning me about saving a PDF? I don't want to have to click OK just to save a PDF. http://www.google.com/support/forum/p/Chrome/thread?tid=773b3155e267c660hl, 2010. Accessed 03.04.2011.

[13] Herley, C. So Long, And No Thanks for the Externalities: The Rational Rejection of Security Advice by Users. In *NSPW* (2009).

[14] Jensen, C., Potts, C., and Jensen, C. Privacy practices of internet users: self-reports versus observed behavior. *International Journal of Human-Computer Studies 63*, 1 (2005), 203–227.

[15] Kaspersky Labs. Number of online attacks soar in 2010. http://www.kaspersky.com/about/news/virus/2011/Number_of_online_attacks_soar_in_2010. Accessed 25.08.2011.

[16] Kirlappos, I., and Sasse, M. Security education against phishing: A modest proposal for a major re-think. *IEEE Security and Privacy 10*, 2 (2011), 24–32.

[17] Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L., Hong, J., and Nunge, E. Protecting people from phishing: the design and evaluation of an embedded training email system. In *CHI* (2007), pp. 905–914.

[18] Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L., and Hong, J. Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology (TOIT) 10*, 2 (2010), 7.

[19] Lovet, G. Dirty money on the wires: the business models of cyber criminals. In *Proceedings of the 16th Virus Bulletin International Conference* (2006).

[20] McCarthy, J., Sasse, M., and Riegelsberger, J. The geometry of web search. *People and Computers XVIII-Design for Life* (2005), 249–262.

[21] Preibusch, S., and Bonneau, J. The password game: Negative externalities from weak password practices. In *Decision and Game Theory for Security*, T. Alpcan, L. Buttyan, and J. Baras, Eds., vol. 6442 of *Lecture Notes in Computer Science*. Springer, 2010, pp. 192–207.

[22] Preibusch, S., Krol, K., and Beresford, A. The privacy economics of voluntary over-disclosure in Web forms. In *WEIS* (2012).

[23] Riegelsberger, J., Sasse, M., and McCarthy, J. Do people trust their eyes more than ears?: media bias in detecting cues of expertise. In *CHI* (2005), pp. 1745–1748.

[24] Sasse, M., Brostoff, S., and Weirich, D. Transforming the 'weakest link': a human/computer interaction approach to usable and effective security. *BT Technology Journal 19*, 3 (2001), 122–131.

[25] Schechter, S. E., Dhamija, R., Ozment, A., and Fischer, I. The emperor's new security indicators. In *Proceedings of the 2007 IEEE Symposium on Security and Privacy* (2007), pp. 51–65.

[26] Schneier, B. *Secrets and Lies: Digital Security in a Networked World*. John Wiley & Sons, 2000.

[27] Sharek, D., Swofford, C., and Wogalter, M. Failure to recognize fake internet popup warning messages. In *Proceedings of Human Factors and Ergonomics Society* (2008), pp. 557–560.

[28] Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L., Hong, J., and Nunge, E. Anti-phishing phil: The design and evaluation of a game that teaches people not to fall for phish. In *SOUPS* (2007), pp. 88–99.

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

[30] Wash, R. Folk models of home computer security. *(SOUPS)* (2010).

[31] Weirich, D., and Sasse, M. Pretty good persuasion: A first step towards effective password security in the real world. In *NSPW* (2001), pp. 137–143.

[32] West, R. The Psychology of Security. *Communications of the ACM 51* (2008), 34–40.

[33] Wu, M., Miller, R., and Garfinkel, S. Do security toolbars actually prevent phishing attacks? In *CHI* (2006), pp. 601–610.

[34] Zeltser, L. How to Design Security Warning Messages to Protect Users. http://blog.zeltser.com/post/3638747689/designing-security-warnings, 2011. Accessed 29.05.2011.