Instrumenting Simple Risk Communication for Safer Browsing

Jacob Abbott
Indiana University
Bloomington, IN
jaeabbot@indiana.edu

Prashanth Rajivan
Carnegie Mellon University
Pittsburgh, PA
prajivan@andrew.cmu.edu

Zheng Dong
Microsoft
Redmond, WA
Zheng.Dong@microsoft.com

Siyu Chen
Amazon
Seattle, WA
siyuc@amazon.com

Jim Blythe
ISI, USC
Los Angeles, CA
blythe@isi.edu

L Jean Camp
Indiana University
Bloomington, IN
ljcamp@indiana.edu

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation: K.6.m Miscellaneous:Security K.4.4 Electronic Commerce: Security

Author Keywords
safety, security, risk perception, risk communication, usability, human behavior

ABSTRACT
The computer security community is in an ongoing struggle with usability, human risk perception, and behaviors. Similarly, people outside this community continue to struggle with computer security. For most people it is difficult to know if they are working with a secure remote system or facing an online threat. The popularity of fake antivirus is a testament to the role of simple confusion in high-risk online behavior. Education is a popular proposed solution, as is making the individual tools for secure computing more usable. Here we illustrate that it is feasible to communicate to people as they are, when they need it, in terms they understand, without additional education. As opposed to training or teaching people to recognize and deal with each risk as appropriate, we developed an interaction approach based simply on safety. We developed and tested a Java-based Web extension to engage users as partners in aggregate risk mitigation. Rather than focusing on specific decontextualized risks we provided a high level estimate that combined privacy and security risks and let participants choose to take those risks or not. Specifically, we tested a tool using simple cartoons that functioned both as communication and controls of security settings. The experimental group was safer on-line in that they ran fewer scripts, disable most iframes, blocked the vast majority of Flash, and trivially identified each webpage which they had previously not encountered as new. These participants did so with no security training, nor increased knowledge of security or technology. Nonetheless, these participants used the non-technical language of the interaction and expressed more awareness of risk. Conversely, those in the control group felt more safe despite their unprotected high-risk Internet browsing. Using simple images as the controller and the communication enabled participants to align their perceptions of risk with the actual on-line risk; they chose to be safer.

INTRODUCTION

Humans are often decried as being the weak link in security. It is common to hear of the "dancing pigs" problem, where "Given a choice between dancing pigs and security, users will pick dancing pigs every time." Our experiment illustrates that if the risk of pig dancing is simply communicated and easy to mitigate, they will reject the dancing pig for the safer quiet sty. Specifically, we have implemented a proof of concept and tested it in situ to test the potential for simple, aggregate communication and controls. Our results illustrate that when individuals understand that the image or video is a risky choice, they often choose security over the task at hand. Our experiment is grounded in the notion that people do not know they are choosing risk when they are choosing to watch a video (of cavorting animals or whatever content). When given the information that a particular media is risky, they may choose not to take the risk if that choice is easily actionable.

We created a browser extension that integrated warnings and control, so that people had to actively choose to take a risk to select multimedia content. Before building the extension we implemented a large-scale closed card sorting experiment to determine which of the various mental models of security risk previously observed in the literature were communicated with our specific cartoons [27]. We then used interviews to determine to what degree these communicated risk in a web extension configuration. We combined multiple vectors of online risk just as individuals combine multiple offline vectors of risk to consider their safety. For example, every driver is concerned with the potential of an accident. However, few

express concern about angular momentum when driving on a curved road as opposed to inadequate frictional coefficients on the wet straight away. People simply want to avoid accidents. When we do risk accidents by speeding or driving in inclement conditions, we are aware we are doing so. Our goal is to bring this safety awareness to web browsing.

There are a wide range of tools to assist in managing online risk, allowing people to implement highly refined risk calculus at the per-connection model. However, these often require that the individual know the risk exists, and require that people micromanage their risks. If they happen to identify the moment of risk, tools to mitigate the risk may be difficult to correlate with the actual risk for the given threat, and may be unusable once identified.

Security warnings and indicators have been the focus of considerable research effort. Early warnings were textual, often lengthy, and written at a high level of literacy. Since then, there has been significant improvement with icons, cartoons, and various indicators.

The goal of our work is to align the user perceptions with the actual user experience of risk. We seek to provide integrated risk communication and controls, so that the current perception of safety aligns with the actual level of willingness to take the risk. That is, our goal is not only to allow individuals to browse the web safely but also to allow them to knowingly take risk with the assumption that only the user can know the context and the urgency of any given task.

People engaging with this tool chose to take fewer risks, browsing primarily in a medium risk mode. As noted in the Federal Cybersecurity Research and Development Strategic Plan people "circumvent cybersecurity practices that they perceive as irrelevant, ineffective, inefficient, or overly burdensome." The month-long user test showed that individuals perceived the cartoon-based controls as relevant, effective, and acceptable while rejecting the click-through warnings.

There are two key observations with respect to this system. The first is low user involvement. People need security and most are aware, from experience or media, that computer security is important. Yet, most people lack the interest and competence to protect their systems. In particular, while users may take one-time actions, vigilant attention to security is less likely. The second key feature is highly scalable automated personalization of resources based on context. In other words, we use each person's history to better learn their own unique and homophilous selection of favorite or most-used websites.

Using safety and holistic rough measurements of risk we implemented an open source extension that takes a human-centered view of risk as any threat to online safety. This enables us to offer one-click risk management. We expect that the risk calculations will be rough and occasionally in error. By providing a simple override, participants were tolerant of initial errors and in fact each week fewer participants used the extension the minimal amount and more reported using it consistently.

Figure 1. The full extension is shown in the top image. The bottom image shows a zoomed in version of the low, medium, and high risk tolerance buttons. Interviews illustrated that users understood the association between risks and the setting, e.g., "I am the pig. That pig is dead." was a description of the high risk setting by one user.

Our extension uses very simple metaphorical cartoons to indicate low, medium, and high risk options. We began by iteratively designing cartoons to communicate the idea of low, medium, or high risk to non-technical users. These cartoons are instantiated as buttons that control the browser settings while communicating the level of risk for a given connection with the selected settings. In a four week experiment, we monitored participants' risk preferences, risk exposure, and interviewed them about their perceptions. During the in-situ experiment, the participants with the extension chose fewer online risks than the control group: scripts were blocked, certificates were analyzed, and iframes were not loaded. Participants in the experimental group expressed more awareness of risk without showing evidence of being more educated about the technologies themselves. Conversely, those in the control group felt more safe and less at risk despite their default (unprotected high-risk) browser settings. Rather than trying to educate all users, a potentially unachievable goal, we argue for simple controls aligned with mental models to simultaneously communicate and manage risk.

Rather than isolating and categorizing risks, our extension assumes that individuals are concerned by information and resource loss instead of the mechanisms behind the loss. So, rather than addressing specific risks, we leverage end user interests in avoiding risks of all types. In our four week study, we found that those with the extension took fewer risk and were more aware of online risk than the control group. The control group expressed confidence in their own safety while taking more risks. The level of knowledge, security education, and IT-related work experience were not significantly different between groups, before or after the experiment. As described below, risk-taking in this experiment is defined as running unknown scripts, running high-risk processes such as iframes and Flash, sending information over unencrypted connections, and accepting connections with unknown sites or sites with suspicious certificates. We sketch how these risks were measured; however, our focus is on user response to the risks, not on the exact measurement of the underlying risk itself.

RELATED WORK

In 1996 Zurko coined the phrase "user-centered security." [39] Providing clear actionable communication for non-technical busy people remains a challenge. Even in the presence of clear motivation, people often cannot translate their concern into action. To assist people in choosing to be safe, we offered the mental model of a safe space for risk communication. Figure 1

shows the three button panel that was the front end of the system. Participants could choose among 3 choices in terms of web safety while browsing: low risk (1), medium risk (2), and high risk with no safety (3). The high risk is how people normally browse the web, where everything works and all components load.

Recently, [18] surveyed the use of HTTPS indicators and icons, working with a tech-savvy population recruited via the Chrome Web Store, TechCrunch, omgchrome.com, and Reddit. They used an open question about urls with different indicators and found that the lock indicator does communicate information about the status of a connection to their participants. The lock seems to match their participants‘ perception of risk. Lin et al.captured mental models via crowdsourcing with a focus on mobile extensions and privacy risks [28]. These two studies targeted different populations, others have shown differences between experts and non-experts in the same experiment.

Kelley et al.'s work on eye tracking and authentication have illustrated that experts and non-experts have very different behaviors in analyzing webpages as being fraudulent or safe, both with and without different stress conditions [6]. A study at Carnegie Mellon University examined the distinction between advanced and novice users in diagnosing security warnings [16]. The results of that think-aloud exercise indicated that there are consistent differences in understanding warnings, leading to different diagnoses and responses to the warnings.

Asgharpour et al.showed differences in the mental models between experts and non-experts by using a card-sorting experiment where words were categorized into mental models [4]. This study validated that non-experts and experts have quite distinct mental models, with the differences being stronger as a more rigorous definition of expertise is applied. Non-experts typically categorized computer security risks with physical safety and criminal activity, whereas experts linked the risks to mental models associated with warfare and health. Her five mental models were chosen from Camp's categorization of naming in computer security literature [12].

Wash identified eight mental models, which he refers to as "folk models", using a snowball set of non-technical computer users. His models showed that home computer owners make security decisions they cannot delegate [32] using reasoning grounded in specific models of threats and threatening actors. Later work specifically constructed warnings using mental models, and found some validation for the use of simple mental models as opposed to more explicitly educational material [9]. Well-aligned mental models can be used to address the challenge of clear communication.

The differences between experts and non-experts is a challenge addressed by security researchers who have collaborated with cognitive science researchers in implementing mental models [5, 8, 31]. Implementing these models requires identifying the model of the specific user, which requires observing user choices and behaviors. Even the same person may vary their behavior based on their perception of the threats and potential harm [36, 30] or the inherent natures of the risks [20].

In 2010, Chen et al.noted a subtle difficulty with understanding cloud-computing threats arising from potentially inaccurate mental models of cloud computing as an always-available service [14]. This viewpoint can create a false sense of security, leading to inadequate security practices, such as a lack of data backups across multiple cloud providers and placing sensitive information on incorrect platforms. They also point out issues with " out-of-date" mental models - how current implementations differ from past concepts. The need for simple ways to communicate different risk levels is part of the motivation for this work and the driver for the highly simplified interaction.

In Zhang et al, the researchers used text, infographics, and a comic to educate participants on why updating anti-virus software is important [38]. Users expressed that they understood why it was important and while making decisions after the study, referenced the comic example for guidance. In the follow up interview, they also reported that they felt more confident in their decisions and decided to relay what they learned to friends and family. Further research also verified these results, indicating that the users better understood the comic which used a medical metaphor [37]. Garg explored the difference between the same script when presented as a video and presented as text in educating individuals on how to avoid being victimized by phishing [21]. He used the metaphor of a solicitor impersonating a banking investigator to leverage story-telling to educate older users. He provided them with a general strategy to avoid being victims of fraud. He also found that the video resulted in superior understanding and retention. The participants from both of these studies did not have technical expertise, but still understood the information. We used a different metaphor yet shared the common goal of simple metaphorical communication.

There has also been research specifically in novice users' views about security practices and awareness [2, 26, 24]. These qualitative investigations (interviews and field observations) enable a deep exploration of a narrow work context. Results, however, may not be generalizable to a larger population. Past research has also focused on exploring end user behaviors that effect the security posture of an organization and how expertise effects these [30]. The results of [30] showed that even with training, naïve risk-taking is a significant hazard in the organizations studied.

Other seminal works on user-centered security include Whitten and Tygar's "Why Johnny Can‘t Encrypt?" [34] which described a cognitive walkthrough and lab-based usability test examining the usability of PGP 5.0. Whitten and Tygar concluded that the usability of security software requires a different standard of usability than other software. Specifically, they suggest that it is necessary for users to be aware of the tasks they need to perform, are able to successfully perform said tasks without making dangerous errors, and are comfortable enough with a security interface to continue using it. The major issues Whitten and Tygar noted are lack of incentive alignment, lack of feedback, and inability to recover from errors, also known as the "barn door property," evoking the futility of lockdown after loss of information.

The "barn door property" captures the continued use and reuse of passwords after having exposed them to attackers through

insecure connections. Inglesant and Sasse found that while individuals do in fact care about security, password policies are too inflexible to match human capabilities [25]. A follow-up study illustrated that graphical passwords had similar difficulties [10].

Longer and unique password requirements are clearly good security hygiene, but have not always been found acceptable in practice e.g. [33, 15]. Leveraging graphical cues for passwords taps episodic memory, as opposed to the standard approach which uses only semantic memory. This was also shown to not be efficient [7], in no small part because the design of the system simply replaces a word with a picture. The time it takes a person to generate and recall graphical passwords is also much longer than just textual passwords [35]. Another cognitive approach tried to use opinions and facts, or, words to help people remember other words [11]. While this was found to be helpful, asking individuals to remember based on facts rather than opinions provided very weak security. In 1999, Adams and Sasse [1] argued that people are not careless with information protection (passwords in their study), but rather, they are rationally allocating their own resources. Security requirements that are antithetical to human capacities cannot be met (i.e., choose many passwords that are impossible to guess and are highly random, don't write them down, remember them, and change them often).

Similarly, policies that conflict with work procedures or prevent completion of tasks are sometimes rejected by users. Nearly concurrently with the emergence of the area of usable security, the interdiscipline of economics of security was also studying the issue of why we have so little security. The conclusion from that perspective is that there is incentive misalignment and inadequate information. Our design goal was to address the second i.e., lack of information.

Anderson [3] began two decades of work that examined the economics of security, explaining that incentive misalignment is a core problem in security technologies. Camp & Wolfram [13] illustrated that it is economically rational to under invest in security in the face of significant externalities imposed by other people's bad security practices. Across the network, imperfect information continues to significantly impact decision-making [22]. Even senior decision-makers experience the lemons market identified by Anderson a decade and a half previously.

The focus on safety and risk as concepts is often embedded in studies of warnings and indicators. For example, the browser icon design work in [18] evaluated user response with a question explicitly about safety, a question listing specific risks, and a question about likely behavior with a given icon. However, the difference between the risk people say they will accept and that risk people actually accept is so significant that the two are considered different measures: revealed preferences refer to behaviors while expressed preferences refers to individual descriptions of their behaviors. In this work we engaged in revealed preferences approach by providing our tool to forty people to use in-situ as they browsed.

These works, from 1999 to 2017 all converged on the same point: that if people do not feel that security will provide them utility (benefit), they will not strive to improve their security. Our goal was to communicate decreased risk and improved safety, in order to communicate the benefit of rejecting some unfamiliar or untrustworthy content. In addition, much of the content that was blocked was advertisements, which was perceived as a benefit.

Two fundamental challenges of security and privacy from a human-centered perspective are that individuals must be both motivated to and capable of adopting the technology. In terms of motivation, we focused on communicating to users that they were at risk, or showing that they were avoiding risk. Only the individual can bring to the table the value of an interaction, individual trust in the context, and decision to take the risk. For example, an insecure wireless link in the rural home of a friend is very different than an insecure link in an urban coffee shop in one of the global centers of e-crime. To the computer, these can appear to have the same risk profile. The goal of keeping the user central to security choices while automating as much as possible was core to our design.

EXTENSION GOALS

Our extension focuses on ease of use and transparency of the technology while also providing risk communication. The goal is to allow users to take security risk only while knowing that they are taking the risk, remember the context in which risks are acceptable, and minimizing risk in others. We provided three choices: low risk, safe (1), medium risk (2), and high risk with no safety By including both interviews and data compilations in the experiment, we hoped to provide measurable reductions of risk exposure while addressing core usability concerns.

The only warning that is ever activated under the high risk mode is when a participant entered a password over an unencrypted connection. High risk mode only blocked blacklisted content. The low risk tolerance level loads only white-listed resources, meaning it loads no scripts, images, specialized fonts, and iframes except for a few whitelisted site. The medium setting uses blacklists, stops scripts that autorun, disables known trackers, disables unknown iframes, and high thresholds for blocking functionality. Warnings are implemented at a less stringent threshold.

It is difficult to compare a multipurpose tool to a tool with a single function. That there are so many single function tools is a condemnation of security, in that it means we expect people to udnerstand ads, cookies, certificates, and other threats as individual vectors. A sample of two of the most popular extensions, NoScript and Ghostery, are shown in Figures 2 3. These allow highly refined control over which scripts to enable. Ghostery allows blocking categories of scripts, such as trackers, beacons, or widgets.

NoScript provides highly refined control, and in the opinion of this team, provides the best cross-site scripting protection available. Settings must be created for each particular site to allow purchases, or commenting. NoScript has a particularly high learning curve compared to Ghostery, due to the categorizations in Ghostery. Compared to our simple one click

interface, both of these extension are much more complicated.

These extensions give strong and refined control, our goal was to create a more simple yet more comprehensive controller. The combination of multiple security technologies using different targeted extensions or apps means that the users must determine which settings to change where there is a loss in functionality. The integration of these security tools into one can ensure that if there is something desirable being blocked, there is a single interface that will allow functionality.

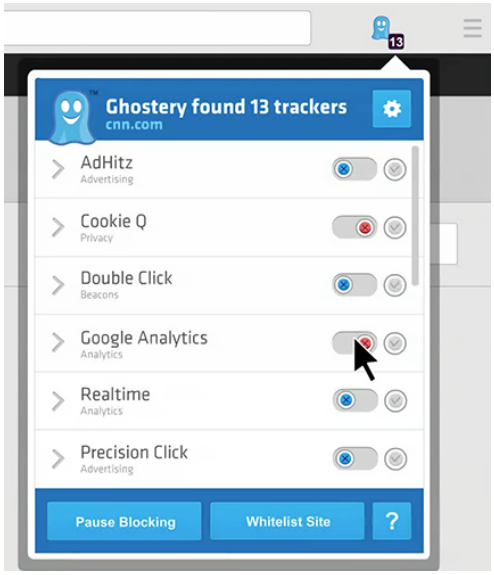

Figure 2. Screenshot of Ghostery

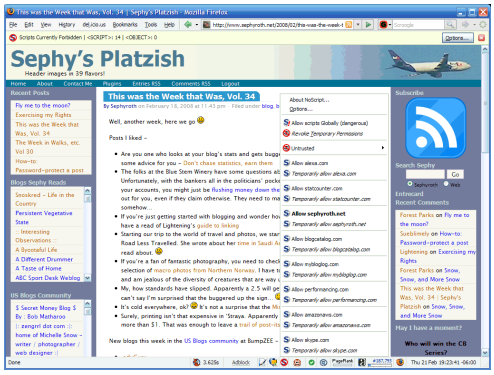

Figure 3. Screenshot of NoScript

The core theoretical grounding of this extension is user-centered design that brings information to people in a format they can understand, that can be translated to action, and without requiring that they dismiss their own interests and engage with us on the topic of computer security.

Password reuse and insecure use of passwords present major risks. We addressed these by instrumenting warnings about password reuse. Such warnings were more or less frequent based on the level of risk acceptance. These warnings could be dismissed and disabled by the user.

Insecure password use was a serious concern and this warning could not be disabled. This was driven by three factors. First, transmitting a password without encryption is risky and encryption is invisible to the user. In contrast, people know when they are reusing passwords. Second, the majority of phishing sites do not use HTTPS. Entering a password in an insecure site is a high risk act [29]. Third, the machine learning module for certificates in our extension had a high degree of accuracy in identifying HTTPS-enabled phishing sites (over 99% [17]). Adding to that the fact that phishing sites were added to the blacklist as soon as identified, users would know they were at high risk when visiting HTTPS phishing sites. Since it is likely a first visit to any given phishing domain, the default of medium risk would prevent loading without a warning. Interacting with a known phishing site or a site likely to be phishing required the user change their setting to high risk. Like the standard warnings we have today, that change of state attempts to undermine phishing as such social engineering attacks rely on the user not knowing they are at risk.

Our goal is to make four contributions. We instrumented and integrated discrete risk modules to meet these goals but all of which learned user preference over time. Second, we implemented end-to-end risk modeling at a per-connection level. Third, we sought to align user risk perception with actual exposure to risk on this per-connection level. Finally, we experimentally measured changes in both risk exposure and perception to test that alignment.

In terms of security settings the tool addresses loading scripts, video, images, evaluating domains as familiar or not, and detecting suspicious certificates. In a four week experiment, we monitored participants' behaviors as well as self-reported perceptions of their behaviors. Participants in the experimental group chose fewer online risks than those in the control group: scripts were blocked, passwords were not entered on unencrypted networks, and Flash was disabled.

We had three high-level concepts of threats in the architecture. These concepts and the underlying threat model are not unique. Each has been subject to individual publication. However, each of these now are unique vectors of risk. Any tool that addresses one risk is unlikely to impinge another.

Web context: Web context was a combination of domain name, certificate, and page elements, particularly scripts. Domain names were evaluated based on individual history with an initial default of the top million trusted. Domains became trusted one week after the first visit or upon explicit user action. That one week window in grounded in reported take-down times from private conversations in the AntiPhishing Working Group. Certificates were evaluated using machine learning as detailed in [17]. Scripts were based on familiarity using personal history, checks for common vectors for malware (i.e., Flash, iFrames), and any script that had an indication of cross site scripting. *Network context*: Network context evaluated the network policy and existence of encryption during transmission. This also included evaluation of familiarity of SSIDs and familiarity of the IDs of devices connected to the same SSID for wireless. *User context*: The likelihood of warnings was grounded in the risk setting chosen by the user. The default

was medium risk.

The sample code below shows the settings at a medium risk level.

```
if (riskTolerance == Risk.MED)
{docShell.allowAuth = false;
docShell.allowImages = true;
docShell.allowJavascript = true;
docShell.allowMetaRedirects = true;
docShell.allowPlugins = false;
docShell.allowSubframes = false;
docShell.allowWindowControl = true;
docShell.allowMedia = false;}
```

Our extension provided both active warnings and status indicators. The active warnings were pop-ups with messages for repeated password use, insecure login (no TLS/SSL), or dangerous websites. An example warning is shown in Figure 6.

EXTENSION USE

For our experiment, we gathered 82 participants by posting fliers at the university and various places of worship. The outreach to places of worship was grounded in team social connections and could arguably be considered snowball sampling. The goal of this outreach was to have a diverse sample and to avoid participants with computing expertise. All stages and work were reviewed and approved by the university IRB.

Users began by participating in an initial interview and survey that consisted of basic demographics and technical knowledge questions. This interview was done by the qualitative team members from the College of Arts & Sciences. We specifically sought non-technical users. Fifty-three of the original participants were invited to the longer study. The remaining 29 participants were deemed to have too much computer and security knowledge to continue the experiment. Those invited were arbitrarily divided into two groups: experimental and control.

Both groups were interviewed as they handed their computers over to the technical team members. The technical team members assisted in installation or updating of Mozilla Firefox. They also installed the extension from our team members. The extension included a local database, and due to resource constraints we did not build an installer. Installation and diffusion were beyond the scope of the study, we wanted to examine use. No user instructions were given, excluding a brief 70 second overview video. Both groups were given the extension with risk calculation and tracking capabilities. However, for the control group this ran in the background and was set not to interfere. That is, they were browsing "high risk" which is the equivalent of not having the extensions. It existed for them only as data compilation as a baseline.

The experimental group was given the full extension. The default setting for each website was set at medium for the experimental group. The participants were instructed to do as much of their Internet browsing as possible with Mozilla, and asked to not use any other extensions.

The participants returned once a week for four weeks. They were paid $20 for each session. These sessions consisted of the participant being interviewed in one room while their data were extracted by a technical team in another room. At the end of the four weeks, there was an exit interview and survey. We had 44 total participants complete the entire experiment. We had the same completion rate for both control and experimental groups.

RESULTS

Interview data and a computer log was collected every week for four weeks from all participants. The audio files were transcribed by crowd workers at TranscribeMe! The online qualitative data analysis service Dedoose was used to code the data and provide a first pass at the analysis. A team of researchers developed the original codes by examining the transcribed responses to the most relevant questions for this study. These were placed into Dedoose with the transcripts and demographic information. Two researchers coded small sections of transcripts until they achieved an inter-rater reliability score above 0.80 and then proceeded to code the remaining 200 transcripts.

Participants were asked to use Firefox with the tool enabled for at least six hours per week. Users reported time with the tool fluctuated over the course of the study, with 35% reporting that they used the tool for 0-9 hours in the first week. By the third week 33% reported minimal tool use., i.e. 0-9 hours. By week 4, 26% reported using the tool 0-9 hours; 44% used it 10-14 hours, and 22% used it more.

For the control group the extension only logged their browsing activity, and calculated the degree of risk for a given page. It was natural for the majority of the control group to respond that the tool gathers/tracks Internet browsing data. Only five people said otherwise, either believing that the tool was designed to track advertisements or that the tool was the same either anti-virus or malware protection. Three people reported that the tool was designed to change the computer speed, as some people reported issues with their computer operating noticeably slower.

The extensions most visible activity was blocking scripts that could contain malicious content. If participants clicked on the image of the pigs in the brick house then the tool blocked large sections of advertisements, images, and videos. If they clicked on the image of the pigs in the straw house then the tool blocked only items on the blacklist. In practice, this meant that the high risk, straw house, rating blocked almost nothing.

Individual participants' answers to "Based on your interaction with tool last week, what do you think the tool does?" ranged from accurate to erroneous, even in a single session. At some point in the 4 weeks 88% of all participants reported accurately that the "tool blocks (removes/hides) things based on the security settings." Over half of this group also reported that the tool provided anti-virus protection.

Participants expressed their perceptions on convenience versus security and efficiency versus security, as well as wanting particular content and realizing there was a security issue. "I felt like the piggy in the brick wall. My computer was safer thanks to the tool, but there's a battle going on between security and convenience." stated one participant. The same participant then said about the high risk setting, "The one

Figure 4. A chart of perceptions of increased security for participants in the experimental group.

it's currently on is its easiest setting and allows the website to work very efficiently." It is hard to judge perceptions on 'efficiency' except that the page would appear normal to them. Two users did report switching to the lowest setting to speed up their computer. No participant singled out security versus privacy.

Overall, 83% of participant responses indicated that they felt the pictures were effective as a tool for communicating computer security. Only two people said that they would have preferred words to pictures, one in the lowest expertise range and the other in the third. One of these two felt it was too simple, but indicated that it would work for others: "I think it's good. I think I'm a pretty savvy internet user, it's a big part of my job and so... um, it's very easy and it makes it very quick to notice and I kept thinking this would probably be really good for like, my mom, who doesn't know as much." A more detailed breakdown of the participants' responses are shown in Figure 4.

The primary objection to the tool was that it included warnings, particularly password reuse warnings. Every participant reused a password at least once and encountered the appropriate warning. Participants associated with the university were found to have no instances of password reuse associated with their university credentials, despite reuse of passwords elsewhere. The password warning for unsafe use was the only difficult to disable warning. Every other warning allowed individuals to reset their risk setting before moving forward. Every other warning indicated that the person could go forward, as shown in Figure 6. There is not a technical solution at the browser for sending a password over an unencrypted link over an unprotected wireless connection. Thus no such mitigation could be offered, unless Tor or a VPN were integrated with the extension.

Understanding the Tool

Participants largely understood the meaning of the pictures that conveyed their level of exposure to potential threats on webpages as a function of their own manipulated tool settings. There was some confusion between risk and protection as the lower security level represented higher risk. The example below portrays a typical response where confusion is evident, however the participant is more correct than they realize:

> Interviewer: This is Picture B. Can you tell me what this means?
>
> Participant: Big bad wolf. This is the medium setting. Again, with what I originally thought the software would do and these pictures... what they are, what they represent don't really line up to me. Cuz it's not like an anti-virus software. These pictures to me, make me think, it's going to moderately protect my computer against certain websites that could be dangerous. But that's not really what it does. It just tells me whether it's safe or not and it blocks some pictures. From what I can discern, ascertain. I don't know.

The descriptions were sometimes a bit vivid, as with this participant:

> Interviewer: So, just kind of tell me what things you notice in these images, starting with this one. So, what kind of things do you notice and what does it make you think of?
>
> Participant: Well, the pig is scared and the wolf is blowing down the twig house. If the pig's not careful, he'll die.
>
> Interviewer: All right, and what do you see in this one, and what does it make you think of?
>
> Participant: The pig is very contented and safe in the brick house, and no security threats can reach him. The wolf can't reach him.
>
> Interviewer: All right. And the last one?
>
> Participant: The pig's going to die, there's no protection.

One participant was worried about the pig in the high risk case, with the strongest word used being "upsetting." Recall from the previous section that another said, "I am the pig. That pig is dead." but it was said with humor not despair.

Others indicated less vivid but equally correct perceptions. The perceptions from the medium setting follow.

> Interviewer: Okay, thank you. What do you think these pictures from the tool mean?
>
> Participant: Yeah, thanks [laughter]. This is picture A, it's a picture from the tool. Yeah. Well, I think it shows that you are unsafe environment while you are on the internet, and there are potential risks around you. You might find out that you may not be able to accept but the risk is that it exists there.
>
> Interviewer: Picture B?
>
> Participant: It seems protection for your computer is required, and you can feel pretty safe if you have some protection.
>
> Interviewer: Picture C?
>
> Participant: If you don't have it you might be, you know, risk yourself and loss of data. Something like that.

They also connected the tool to safety. In response to the query, "Did the tool make you feel more or less safe while online?" the responses were usually positive (Definitely more, More safe absolutely, I felt safer). Awareness was mentioned by two participants, and "knowing" about risks by seven.

> Participant: It was blocking unwanted or unsafe material, in my opinion. I'm not sure that's what it was doing. That's what it seems like it was doing. It was making sure that my viewing experience or surfing experience was controlled which is like a parent, which is great. And that just makes me feel safer.

Another responded to the question about safety by discussing awareness.

> Participant: I wouldn't say safe, but I would say just aware. So it made me feel more that the sites that I didn't think that I probably shouldn't be on, it would tell me, "Your passwords are at stake here" I'd be like,"Well, maybe I shouldn't be on this site," and go off of it. But I feel it did give me that security to know, so it helped me be more aware of what I should be on or not.

Changing Tool Risk Levels

Ten of the twenty- five experimental participants reported keeping the security setting on the lowest level the entire time. Like the control group, the experimental group perceived their risk as lower than it was, as the graph of time spent at each level illustrates.

Twenty of the 25 experimental users reported reducing the settings at some point during the study period. Five reported it only once, two in the first week and three in the third. Reports of reducing the settings were consistent throughout the study. Participants generally wanted to see all of the content on the website or needed to reduce the settings in order to get the functionality from the site that they desired. There were more changes in risk level than reported. By the final week some participants reported not having to change the setting. The design goal was to make the tool highly usable. Therefore part of the customization was storing the participant choice for a site, so it was not necessary to change settings on return visits.

Participants offered various reasons for changing the risk setting. One decreased security when the default was placed on medium for trusted sites, expressing this as, "Uh I turned it on no security whenever it automatically bumped itself up to medium." A second also explained that decreasing risk was needed to access content, "Most of the time I would keep it on medium setting. That's always good. But if there's something like, if I needed to watch a video, I was like– I would just go on SportsCenter and if I wanted to watch a video I would have to put it on the low setting to watch some of the videos." A third participant explained, "On a site, like Reddit or a news– any type of site where if I click something and it takes me somewhere else - a site that redirects you - I would tend to maybe put it on medium more because I don't think I'm staying in the same place that I know is safe."

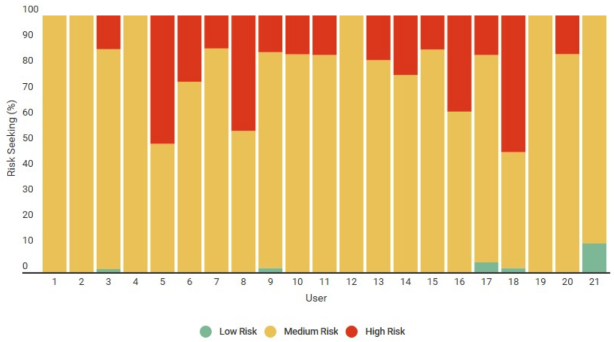

Figure 5. A graph of the level of risk that each user chose in the last week.

Eight people reported increasing the security of the tool, sometimes to hide the advertisements (2 people) but more often to play with the tool and see the changes in the webpage display. Only three people reported wanting to increase their security with the tool. Two of the three were in the lowest expertise score range. A total of 13 people reported simply playing with the tool. Many were pleased with the ad-blocking functionality.

In addition to the perceptions of changes, we examined how often there were changes. We evaluated how often a participant's browsing switched between high, medium, low risk settings across different websites. This is shown for the last week in Figure 5. This graph is only for the participants that continued the experiment through the fourth week. While there were some users that chose to be in high risk, most users spent a majority of the time in medium risk. We also noticed that users chose higher risk setting when surfing social media sites, most likely because the tool blocks most of the information on such sites.

Recall that we had a total of twenty-five participants in the experimental group. Figure 5 shows the percentage of time each participant spent in medium or low risk for each week. Similarly if there is no bar either the participant spent no time at lower risk, or did not continue to participate. When people ceased participating, we assume that they return to their high-risk default browser behaviors.

The extension defaulted to the medium level of risk whenever a user visited a new website, thus introducing protection from potentially malicious scripts and allowing the user to opt for increased protection or less. Not shockingly, defaults are powerful even when easy to change. One way of evaluating the graph above is that participants embraced the defaults setting most of the time.

Our instrumentation could only measure when Firefox was in use. If the participants changed browsers (although none of them reported doing so) then the data would not be included.

Warnings

The following quotes represent how one user felt about password notifications. The findings in this study point to the fact

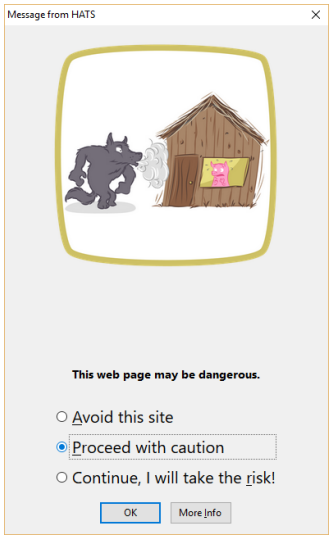

Figure 6. A message from the extension about an insecure web page.

that people not only won't change their passwords, but find the notifications about password security to be an inconvenience.

> Participant Week 1: With the warnings about the passwords, there's no option to shut those notifications off. As it is with almost every security thing it's like, "Well, if you want to keep being reminded of this every time then."

The other warnings were click-through and allowed risk level changes. The password was specifically mentioned as problematic.

> Participant Week 2: So, when it gives you the, "You've used this password before," there's got to be a check box for, "Stop reminding of this." So, that made it worse. That's pretty much it.

None of the warnings could be disabled, but the other warnings were not subject to complaint.

DISCUSSION

We attempted to design a system that approaches users as people making complex risk decisions, requiring simple communication. Instead of a plethora of add-ins, add-ons, and ever expanding vocabulary of attacks and defenses, each individual is provided with a single narrative with a consistent metaphor about the context, and a path to risk mitigation. These narratives are embedded in messages that (1) leverage mental models to describe the dangers, (2) describe particular risk levels that the user may be exposed to, and (3) are delivered close in time before the danger may actually be realized.

The design is grounded in the answers to core security questions:

> 1) What risks do users care about?
>
> 2) What risks should they care about?
>
> 3) How can interactions design enable users to manage these risks?

Currently the plethora of add-ins provide limited risk information via their interactions. They require users to have an understanding of the risk in order to select the correct add-on or protection. Our goal was specifically to respect the cognitive budgets of busy people. Individuals found the interaction easy to understand and easy to use.

We believed users should care about sending passwords over unprotected links to sites unprotected by encryption. Our participants disagreed. The password warning was the one consistently rejected component of the interaction. Google is currently struggling with the same issue, with the extension Password Alert. With Password Alert, the company requires that people change their password after they enter it over an insecure link [23].

This is an open question for the engineering and design communities. Certainly no one wants to listen to any warning. Password alerts may join seatbelt audible warnings as one that engineers agree is critical. Right now, there is not actionable ethical guidance to designers. It is clear that such warnings are not wanted.

Other risks were blocked, and the blocking was primarily seen as an asset. From the participant perspective, it sometimes appeared to be only blocking advertisements. The extension evaluates risks by examining past attacks including phishing, rogue certificates, cross site scripting, and network attacks such as surveillance and man-in-the-middle attacks. To non-technical individuals, these attacks are not a common worry. The role of the expert in this design was to automate the identification of online risk and mitigate these to the extent possible. The role of the participant was to understand their own willingness to take risks.

By coordinating the user communication and security settings of the system interaction, our goal was to provide responsive interaction to empower the user to distinguish and protect him or herself appropriately. As noted in the Federal Cybersecurity Research and Development Strategic Plan people "circumvent cybersecurity practices that they perceive as irrelevant, ineffective, inefficient, or overly burdensome." [19] The month-long user test showed evidence that the participants perceived the tool as relevant and acceptable.

The results of the four week test showed that people will change their risk exposure if it is simple to do so. Significant changes in risk exposure online at the individual level, aggregated over users, creates a decrease in exposure. It also illustrated that people did not necessarily feel that they were changing their behaviors. Although the changes in risk level continued over four weeks, the reported changes in risk level decreased. Our optimistic perspective is that this implied that changing the risk level became significantly easy as not to be remembered.

If this experimental system were to be used in the wild, the measures of risk would need to be more exacting. The domain names and certificate results were highly reliable, as shown in the publications on those modules. Yet a primary source of risk is scripts. The difficulty of measuring the risks of scripts is the reason we used whitelists and blacklists, as the effort for

the entire project could have focused only on malicious script identification.

CONCLUSION

Individuals are expected to learn about different threats and select individual tools to defeat individual threats. Highly customizable software exists to defeat viruses, limit scripting access, prevent tracking, confirm the legitimacy of certificates, and other specific threats. Machine learning, advanced user models, and interdisciplinary theories (such as macroeconomics) are being applied to meet the challenges of identification of threats. The threats are increasingly subtle and multifaceted, and the targeted technically complex defenses offered to users are increasingly inadequate in human terms.

This research builds on decades of findings surrounding how people perceive security communications, estimate risk, and interact with technology. Due to the extent of uncertainty in all of these variables the importance of defaults and automated settings has long been recognized. The users in this study had the ability to determine what level of risk they were comfortable with given the various contexts as they traversed the internet. Our team endeavored to measure these influences and the effects of our extension's communication methods, both active and passive, on user behavior.

As threat detection and security technology becomes more complex, non-technical people who are already overwhelmed cannot be expected to manage this complexity. These two trends – increasingly complex security and increasing user technical heterogeneity – have been treated as if they exist in opposition. In this work we have shown that these can be well-aligned by combining risk communication, usable computing, and complex, learning security technology to use the underlying complexity of the technology to simplify the interaction. We designed a holistic, trivial to use, technologically comprehensive tool which allows individuals to browse safely, with limited risk, or at highest risk. Given the option to mitigate risk, the experimental participants often choose to do so.

ACKNOWLEDGMENTS

This research is sponsored by DHS N66001-12-C-0137, Cisco Research 591000, and Google Privacy & Security Focused Research. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies or views, either expressed or implied, of the DHS, ARL, Google, Cisco, IU, or the US Government. We also want to acknowledge contributors: Mike D'Arcy (ISI), Krishna C. Bathina, Shakthi Gopavaram, and Jill Minor of Indiana University.

REFERENCES

1. A. Adams and M.A. Sasse. 1999. Users are not the enemy. *Commun. ACM* 42, 12 (1999), 40–46.
2. Eirik Albrechtsen. 2007. A qualitative study of users' view on information security. *Computers & security* 26, 4 (2007), 276–289.
3. R. Anderson. 2001. Why Information Security is Hard - an Economic Perspective. In *Computer Security Applications Conference, (ACSAC)*. IEEE.
4. Farzaneh Asgharpour, Debin Liu, and L Jean Camp. 2007. Mental models of security risks. In *Financial Cryptography and Data Security*. Springer, 367–377.
5. Steffen Bartsch, Melanie Volkamer, and TU CASED. 2013. Effectively Communicate Risks for Diverse Users: A Mental-Models Approach for Individualized Security Interventions.. In *GI-Jahrestagung*. 1971–1984.
6. Bennett Bertenthal. 2015. Tracking Risky Behavior on the Web: Distinguishing Between What Users\. In *2015 AAAS Annual Meeting (12-16 February 2015)*. aaas.
7. Robert Biddle, Sonia Chiasson, and Paul C Van Oorschot. 2012. Graphical passwords: Learning from the first twelve years. *ACM Computing Surveys (CSUR)* 44, 4 (2012), 19.
8. Jim Blythe and L Jean Camp. 2012. Implementing mental models. In *Security and Privacy Workshops (SPW), 2012 IEEE Symposium on*. IEEE, 86–90.
9. Cristian Bravo-Lillo, Lorrie Faith Cranor, Julie Downs, and Saranga Komanduri. 2010. Bridging the gap in computer security warnings: A mental model approach. *IEEE Security & Privacy* 2 (2010), 18–26.
10. Sacha Brostoff, Philip Inglesant, and M Angela Sasse. 2010. Evaluating the usability and security of a graphical one-time PIN system. In *Proceedings of the 24th BCS Interaction Specialist Group Conference*. British Computer Society, 88–97.
11. Julie Bunnell, John Podd, Ron Henderson, Renee Napier, and James Kennedy-Moffat. 1997. Cognitive, associative and conventional passwords: Recall and guessing rates. *Computers & Security* 16, 7 (1997), 629–641.
12. L Jean Camp. 2006. Mental models of privacy and security. *Available at SSRN 922735* (2006).
13. L Jean Camp and Catherine Wolfram. 2004. Pricing Security. In *Economics of Information Security*. Springer, 17–34.
14. Yanpei Chen, Vern Paxson, and Randy H Katz. 2010. What's new about cloud computing security. *University of California, Berkeley Report No. UCB/EECS-2010-5 January* 20, 2010 (2010), 2010–5.
15. Sonia Chiasson, Paul C van Oorschot, and Robert Biddle. 2006. A Usability Study and Critique of Two Password Managers.. In *Usenix Security*, Vol. 6.
16. Lorrie Faith Cranor and Simson Garfinkel. 2004. Guest Editors' Introduction: Secure or Usable? *Security & Privacy, IEEE* 2, 5 (2004), 16–18.
17. Zheng Dong, Apu Kapadia, Jim Blythe, and L Jean Camp. 2015. Beyond the lock icon: real-time detection of phishing websites using public key certificates. In *2015 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 1–12.
18. Adrienne Porter Felt, Robert W. Reeder, Alex Ainslie, Helen Harris, Max Walker, Christopher Thompson,

Mustafa Embre Acer, Elisabeth Morant, and Sunny Consolvo. 2016. Rethinking Connection Security Indicators. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. USENIX Association, Denver, CO, 1–14. https://www.usenix.org/conference/soups2016/technical-sessions/presentation/porter-felt

19. National Coordination Office for Networking, Information Technology Research, and Development. 2016. 2016 Federal Cybersecurity Research and Development Strategic Plan. (2016). https://www.nitrd.gov/Publications/PublicationDetail.aspx?pubid=61
20. Vaibhav Garg and Jean Camp. 2012. End user perception of online risk under uncertainty. In *System Science (HICSS), 2012 45th Hawaii International Conference on*. IEEE, 3278–3287.
21. Vaibhav Garg, L Jean Camp, Katherine Connelly, and Lesa Lorenzen-Huber. 2012. Risk communication design: video vs. text. In *Privacy Enhancing Technologies*. Springer, 279–298.
22. James T Graves, Alessandro Acquisti, and Nicolas Christin. 2016. Big Data and Bad Data: On the Sensitivity of Security Policy to Imperfect Information. *The University of Chicago Law Review* (2016), 117–137.
23. Andy Greenberg. 2015. Chrome Can Now Warn Users Who Type Gmail Passwords in Dumb Places. Wired. https://www.wired.com/2015/04/google-chrome-password-alert/
24. Tejaswini Herath and H Raghav Rao. 2009. Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. *Decision Support Systems* 47, 2 (2009), 154–165.
25. Philip G Inglesant and M Angela Sasse. 2010. The true cost of unusable password policies: password use in the wild. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 383–392.
26. Iulia Ion, Rob Reeder, and Sunny Consolvo. 2015. âĂIJ... no one can hack my mind: Comparing Expert and Non-Expert Security Practices. In *Symposium on Usable Privacy and Security (SOUPS)*.
27. Karrington Lewis, Krishna C. Bathina, Prashanth Rajivan, and L Jean Camp. 2016. Getting the Message Through: Risk Communication with Cartoons and Mental Models. (March 2016).
28. Jialiu Lin, Shahriyar Amini, Jason I Hong, Norman Sadeh, Janne Lindqvist, and Joy Zhang. 2012. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 501–510.
29. Emily Schechte. 2016. Moving towards a more secure web. *Google Security Blog* (8 9 2016). https://security.googleblog.com/2016/09/moving-towards-more-secure-web.html
30. Jeffrey M Stanton, Kathryn R Stam, Paul Mastrangelo, and Jeffrey Jolton. 2005. Analysis of end user security behaviors. *Computers & Security* 24, 2 (2005), 124–133.
31. Melanie Volkamer and Karen Renaud. 2013. Mental models–general introduction and review of their application to human-centred security. In *Number Theory and Cryptography*. Springer, 255–280.
32. Rick Wash. 2010. Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*. ACM, 11.
33. Brent R Waters, Dirk Balfanz, Glenn Durfee, and Diana K Smetters. 2004. Building an Encrypted and Searchable Audit Log.. In *NDSS*, Vol. 4. 5–6.
34. Alma Whitten and J Doug Tygar. 1999. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0.. In *Usenix Security*, Vol. 1999.
35. Susan Wiedenbeck, Jim Waters, Jean-Camille Birget, Alex Brodskiy, and Nasir Memon. 2005. PassPoints: Design and longitudinal evaluation of a graphical password system. *International Journal of Human-Computer Studies* 63, 1 (2005), 102–127.
36. Michael Workman, William H Bommer, and Detmar Straub. 2008. Security lapses and the omission of information security measures: A threat control model and empirical test. *Computers in Human Behavior* 24, 6 (2008), 2799–2816.
37. Leah Zhang-Kennedy and Sonia Chiasson. 2014. *Using Comics to Teach Users About Mobile Online Privacy*. Technical Report. Technical Report TR-14-02, School of Computer Science, Carleton University, Ottawa, Canada.
38. Leah Zhang-Kennedy, Sonia Chiasson, and Robert Biddle. 2014. Stop clicking on âĂIJupdate laterâĂİ: Persuading users they need up-to-date antivirus protection. In *Persuasive Technology*. Springer, 302–322.
39. Mary Ellen Zurko and Richard T Simon. 1996. User-centered security. In *Proceedings of the 1996 workshop on New security paradigms*. ACM, 27–33.