Scaling the Security Wall

Developing a Security Behavior Intentions Scale (SeBIS)

Serge Egelman
International Computer Science Institute
Berkeley, CA, USA
egelman@cs.berkeley.edu

Eyal Peer
Bar-Ilan University
Ramat Gan, Israel
eyal.peer@biu.ac.il

ABSTRACT

Despite the plethora of security advice and online education materials offered to end-users, there exists no standard measurement tool for end-user security behaviors. We present the creation of such a tool. We surveyed the most common computer security advice that experts offer to end-users in order to construct a set of Likert scale questions to probe the extent to which respondents claim to follow this advice. Using these questions, we iteratively surveyed a pool of 3,619 computer users to refine our question set such that each question was applicable to a large percentage of the population, exhibited adequate variance between respondents, and had high reliability (i.e., desirable psychometric properties). After performing both exploratory and confirmatory factor analysis, we identified a 16-item scale consisting of four sub-scales that measures attitudes towards choosing passwords, device securement, staying up-to-date, and proactive awareness.

Author Keywords

Psychometrics; Individual differences; Security behavior

ACM Classification Keywords

H.1.2. Models and Principles: User/Machine Systems; K.6.5. Management of Computing and Information Systems: Security and Protection; J.4. Social and Behavioral Sciences: Psychology

INTRODUCTION

Each year, billions of dollars are spent on computer security education programs. Many employers require employees to undergo training regimens, prior to granting network or computer access, or on an ongoing basis. The U.S. National Institute of Standards and Technology (NIST) offers guidelines for these programs [46], and even the U.S. Department of Homeland Security has a "National Cyber Security Awareness Month" [42]. The NIST guidelines state that "formal evaluation and feedback mechanisms are critical components of any security awareness, training, and education program" [46], yet provide no evaluation metrics. Measuring attack rates is a suboptimal strategy: fewer successful attacks may simply be due to fewer attempts; increased attack success may be due to new vulnerabilities, increased sophistication, or other factors that are independent of user behavior.

In addition to training and education programs, a plethora of organizations offer tips and advice to end-users on how they can stay safe online. However, due to the wide range in specific actionable items offered, most end-users are likely to be confused about which advice to follow and which to ignore. For example, the U.S. Computer Emergency Readiness Team (US-CERT) offers over 500 actionable items [41], whereas Verizon, a large U.S. cellular phone provider and ISP, offers only 4 items [43]. If there are a core set of general behaviors underlying most of this computer security advice, the actionable items offered to end-users can be drastically reduced to be made more memorable. Similarly, in certain cases, reliable self-reporting of these behaviors could replace the need to perform costly and time-consuming observational studies.

When studying human behavior, researchers across many different fields will often use scales as proxies for observation. Scales "are intended to measure elusive phenomena that cannot be observed directly" [11]. Scales are therefore highly useful when behavioral or observational experiments are either too costly, complex, or simply not possible. A scale can also measure how attitudes or behaviors change over time. For instance, the Westin privacy index is used to segment the population based on attitudes towards online privacy, and while there is substantial debate about whether it is predictive of actual privacy behaviors [48], it is still a useful tool that researchers use to examine how privacy attitudes evolve over time [22]. While carefully constructed scales are a core tool when studying human behavior in other domains, they have yet to be applied to the study of computer security behaviors.

We developed a new scale for assessing the computer security behaviors of end-users: the Security Behavior Intentions Scale (SeBIS). In short, our scale measures users' self-reported adherence to computer security advice. We employed the 4-step scale development procedure offered by Netemeyer *et al.* [29], wherein we used common security advice offered to end-users as our content domain, generated and refined the individual questions, performed Exploratory Factor Analysis (EFA) to explore underlying constructs and develop sub-scales, and then performed Confirmatory Factor Analysis (CFA) to finalize our scale and measure its reliability. Our resulting security behaviors scale was evaluated on 3,619 respondents, and consists of 16 items that measure 4 underlying constructs: device securement, password generation, proactive awareness, and updating behaviors.

CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea.

ACM 978-1-4503-3145-6/15/04...$15.00.
http://dx.doi.org/10.1145/2702123.2702249

RELATED WORK

Several models have been created to analyze how humans make security decisions (e.g., [47, 9, 30]), which researchers have used to offer recommendations. For instance, Egelman *et al.* showed that when less-frequent high-risk warnings appear similar to frequent low-risk security warnings, users ignore both [12]. Felt *et al.* showed that smartphone permission warnings are overlooked because they occur too frequently and with poor timing [16]. In this vein, usable security research has greatly improved the security interventions to which users are exposed. We posit that further gains can be made by differentiating users according to individual traits and offering different intervention designs based on those traits. However, for this to be possible, we first need a reliable tool for measuring users' traits, such as their adherence to common computer security practices.

Scales are most often used in psychology, as a means of measuring underlying psychological constructs. One scale that might be relevant to our work is the Domain-Specific Risk-Taking (DoSpeRT) scale, which measures a person's self-reported propensity to engage in risky behaviors across four dimensions: ethical, financial, health/safety, recreational, and social [4]. We hypothesize that an individual's propensity to take safety risks may predict their computer security behaviors. Other psychometrics may also apply: for instance, high impulsivity (e.g., [31]) may predict an individual's willingness to visit questionable URLs, low concern for future consequences (e.g., [21]) may predict an individual's disinclination for running security software, and so forth.

While we are not aware of any studies to develop a scale to measure end-user security behaviors, researchers have raised questions about the psychological underpinnings of computer crime for over three decades. Copes *et al.* created a survey instrument to examine correlations between victimization and demographic factors [8]. Sherizen suggested that there may exist certain "markers" that could be used to detect the propensity for committing online crimes [38]. Only recently have others started to explore this concept. For instance, self-control is a psychological construct with corresponding scales that can be used to measure its prevalence in people. Researchers have administered these scales to see if they correlate with either committing online crimes [27], or being the victim of online crimes [5, 26].

Measuring more general types of computer security behaviors has been limited to within organizations. Stanton *et al.* created a taxonomy of workplace computer security behaviors, and showed that they could be grouped by two dimensions: intentions and expertise [39]. Ng *et al.* [30] used factor analysis to create and evaluate a model of end-user security behavior. Their goal was to use this model to predict email security behaviors, whereas our goal is the development of a composite measurement scale.

Privacy is related to security and has several of its own scales that have been developed over the years and are actively used by researchers to better understand people's attitudes and intentions. The most well known privacy scale is the Westin index, which is used to classify consumers into three groups based on their privacy attitudes: fundamentalists, pragmatists, and the unconcerned [22]. While the Westin index is well-cited and frequently used by privacy researchers, it has been shown to be a poor predictor of privacy-preserving behavior (e.g., [48, 16]). The scale is therefore used to illustrate the gap between stated preferences (as measured by the Westin index) and observed behaviors (as measured through empirical studies)—the so-called "privacy paradox." Thus, the self-reported data may not be "wrong," but instead there may be many other factors at play that influence behavior [1, 2].

The Westin index is unidimensional and therefore only measures one aspect of privacy, whereas other scales measure privacy attitudes and behaviors across multiple dimensions. Malhotra *et al.* developed the Internet Users' Information Privacy Concerns (IUIPC) scale by performing careful factor analysis, and showed that privacy concerns can be measured across three dimensions: control over information, awareness of privacy practices, and attitudes about information collection [23]. Buchanan *et al.* developed three different unidimensional privacy scales, which together measure concerns, general behaviors, and use of technical solutions [6].

The Security Behavior Intentions Scale (SeBIS) is a new scale to measure end-users' intentions to comply with computer security advice. We followed the scale development procedure outlined by Netemeyer *et al.* [29]:

1. **Construct Definition and Content Domain:** Clearly defining the construct that the scale intends to measure.
2. **Generating and Judging Measurement Items:** Creating a pool of candidate questions and then evaluating the questions to remove ones that are invalid.
3. **Designing and Conducting Studies to Develop and Refine Scale:** Performing EFA to reduce the set of questions and explore latent constructs to build a model.
4. **Finalizing the Scale:** Performing CFA to confirm that the scale fits the intended model, as well as reliability analysis.

In the remainder of this paper, we describe how we applied each of these steps and their results. We then describe how we used our scale to explore correlations between self-reported security behaviors and existing psychometrics.

CONSTRUCT DEFINITION AND INITIAL ITEMS

The goal of our scale is to measure end-users' intentions to comply with common security advice. This created two constraints: limiting our metric to security behaviors applicable to most end-users (*vis-à-vis* advice for corporate users, etc.), and only consisting of "widely accepted" security behaviors. We examined the advice offered to end-users by nationwide U.S. ISPs, US-CERT, and industry consortia. Based on this advice, we enumerated a preliminary set of computer security behaviors, omitting advice that was too technical.

Our initial list consisted of 26 behaviors. Next, we met with a group of six computer security experts who offered additional items, increasing our set to 30 behaviors. We transformed these 30 behaviors into personal statements that could be answered based on degree of agreement (i.e., "strongly agree" to "strongly disagree"). To minimize acquiescence bias (i.e., the

#	Question	N/A		μ	σ
A1	*I apply software updates as soon as my computer prompts me.*	5	(1.0%)	3.20	1.221
A2	*I am happy to use an older version of a program, as long as it meets my needs.*	5	(1.0%)	^{r}1.99	1.000
A3	*Whenever I step away from my computer, I lock the screen.*	5	(1.0%)	2.50	1.306
A4	*Others can access my smartphone or tablet without needing a PIN or passcode.*	21	(4.4%)	^{r}3.34	1.545
A5	*When I discover a computer security problem at work, I'm likely to promptly report it to my employer.*	64	(13.4%)	4.08	0.995
A6	*It's important to use a WiFi password to prevent unauthorized people from using my home network.*	11	(2.3%)	4.66	0.690
A7	*I frequently click links in email messages to see what they are, regardless of who sent the message.*	5	(1.0%)	^{r}4.51	0.922
A8	*It's important to run anti-virus software on my computer.*	7	(1.5%)	4.35	0.941
A9	*When browsing websites, I frequently mouseover links to see where they go, before clicking them.*	4	(0.8%)	4.13	0.977
A10	*When using public WiFi, I visit the same websites that I would visit when using the Internet at home.*	20	(4.2%)	^{r}2.93	1.266
A11	*I usually do not pay attention to where I'm downloading software from.*	2	(0.4%)	^{r}4.38	0.900
A12	*I frequently backup my computer.*	5	(1.0%)	3.07	1.165
A13	*I frequently visit websites even when my web browser warns me against it.*	8	(1.7%)	^{r}3.98	1.028
A14	*I circumvent my employer's computer usage policies when they prevent me from completing a task.*	86	(18.0%)	^{r}3.54	1.184
A15	*I am careful to never share confidential documents stored on my home or work computers.*	15	(3.1%)	4.36	0.757
A16	*Frequently checking the access control settings on social networking websites isn't worth the time it takes.*	18	(3.8%)	^{r}3.56	1.165
A17	*I always write down my passwords to help me remember them.*	6	(1.3%)	^{r}3.60	1.313
A18	*Creating strong passwords is not usually worth the effort.*	6	(1.3%)	^{r}4.05	1.047
A19	*I frequently check my financial accounts for fraudulent charges.*	10	(2.1%)	4.11	0.914
A20	*If I receive a suspicious email from a company that I do business with, I'll phone the company to make sure the email is accurate.*	22	(4.8%)	3.52	1.236
A21	*I never give out passwords over the phone.*	7	(1.5%)	4.53	0.787
A22	*I frequently purchase things that I see advertised in unsolicited emails.*	4	(8.8%)	^{r}4.51	0.840
A23	*I tend to ignore computer security stories in the news because they don't impact me.*	4	(8.8%)	^{r}3.83	1.050
A24	*I use encryption software to secure files or email messages.*	10	(2.1%)	2.74	1.225
A25	*Once I create a password, I tend to never change it.*	5	(1.0%)	^{r}3.30	1.182
A26	*I try to create a unique password for every account I have.*	5	(1.0%)	3.21	1.284
A27	*Rather than logging out of websites, I usually just navigate elsewhere or close the window when I'm done.*	7	(1.5%)	^{r}3.06	1.299
A28	*I always make sure that I'm at a secure website (e.g., SSL, "https://", a lock icon) when transmitting information online.*	4	(0.8%)	3.80	1.173
A29	*I frequently use privacy software, "private browsing" or "incognito" mode when I'm online.*	6	(1.3%)	3.17	1.247
A30	*I frequently let others use my computing devices (e.g., smartphone, tablet, laptop).*	3	(0.7%)	^{r}3.79	1.172

Table 1. Initial set of security questions evaluated on a 5-point Likert scale (from "strongly disagree" to "strongly agree") by 479 participants. Depicted are the questions, the rate of "N/A" responses, and the average responses and standard deviations after recoding negatively-phrased questions (^{r}).

propensity for respondents to agree with every question) [34], we reworded half the questions to be negatively phrased (i.e., agreement with the statement indicates *not* following common security advice). Thus, our initial question pool consisted of the 30 questions listed in Table 1.

Method and Demographics

In August of 2014, we performed an initial evaluation of our questions. We recruited participants from Amazon's Mechanical Turk, restricting the survey to participants based in the U.S. who had an approval rate of 95% or greater. We paid each participant $1.00 to complete our survey, which consisted of the 30 aforementioned statements evaluated on a Likert scale ("strongly disagree (1)," "disagree (2)," "neither agree nor disagree (3)," "agree (4)," and "strongly agree (5)"). We gave participants a "N/A" option and the ability to simply not answer questions so that we could determine whether any questions were not applicable to our audience.

We were concerned that social desirability bias may have an impact on participants' responses [10]. That is, some participants may misreport their behaviors in order to make it appear as though they engage in more socially acceptable behaviors (e.g., participants may be ashamed to admit engaging in ill-advised behaviors). We took two steps to mitigate this concern. First, we advertised the experiment as a "computer behavior survey," so as to not specifically mention security during recruitment (thereby minimizing a potential selection bias). Second, we asked all participants to complete the 10-item Strahan-Gerbasi version of the Marlowe-Crowne Social Desirability Scale [40], which we then correlated with participants' responses to the security behavior questions.

As with all surveys, we were concerned with being able to deter and detect careless responses. Based on Meade and Craig's recommendation [25], we included two separate "instructed response items" in our survey. On the first page of the survey after the consent form, we included the following:

> *This study requires you to voice your opinion using the scales below. It is important that you take the time to read all instructions and that you read questions carefully before you answer them. Previous research on preferences has found that some people do not take the time to read everything that is displayed in the questionnaire. The questions below serve to test whether you actually take the time to do so. Therefore, if you read this, please answer 'three' on the first question, add three to that number and use the result as the answer on the second question. Thank you for participating and taking the time to read all instructions.*
>
> *I would prefer to live in a large city rather than a small city.* [Strongly disagree (1), (2), (3), (4), (5), (6), Strongly agree (7)]
>
> *I would prefer to live in a city with many cultural opportunities, even if the cost of living was higher.* [Strongly disagree (1), (2), (3), (4), (5), (6), Strongly agree (7)]

If participants failed to select 3 and 6, respectively, they were given a second opportunity featuring bolded instructions. If they failed a second time, they were prevented from completing the survey. In addition to this attention check task, we included a second question hidden amongst the true/false Strahan-Gerbasi social desirability statements: *I do not read the questions in surveys.* We excluded participants who answered "true" for this question from our analysis.

We received a total of 503 responses, though removed 24 (4.8%) who did not answer the second attention check question correctly. Our resulting sample of 479 participants consisted of 273 males (57.0%) and ages ranged from 18 to 69 ($\mu = 33.2$, $\sigma = 10.5$). Thirty-six percent of our participants (173 of 479) held bachelor's degrees, while an additional 12.9% had completed some graduate school, and median household income was between $35,000 and $50,000. We therefore consider our sample to be representative of the U.S. online population, albeit with a slight male skew.

Results

We observed that the internal reliability of our initial question set was quite high (Cronbach's $\alpha = 0.83$). We performed Pearson correlations between the Strahan-Gerbasi social desirability scale and each question, and only observed one to be statistically significant: A28 ($r = 0.160$, $p < 0.0005$). However, given the small effect size (i.e., 2.6% of variance is shared between this question and the social desirability scale), we chose to retain the question. This suggests that participants answered truthfully and consistently. At the same time, we discovered several issues with our items, which highlighted the need for question refinement: non-applicability, poor item-total correlation, and ceiling effects.

Non-Applicability

We measured how many participants skipped items or responded with a "N/A" option. Two questions stood out, *A5* and *A14*, which both had to do with workplace computer usage. Participants did not respond to these questions 13.4% and 18.0% of the time, respectively. In hindsight, possible confounds are obvious: for instance, not having a job involving computer use. Thus, we decided to remove the question about circumventing policies (*A14*) and reworded *A5* to focus on promptly fixing/reporting security problems in general.

Additionally, questions *A22* and *A23* were not answered by 8.8% of participants. The former concerned buying items advertised via unsolicited emails, which we decided is not a security concern *per se*. The latter had to do with ignoring computer security stories in the news, which we felt was ambiguous: it did not describe a specific behavior, and a negative response could indicate either not following the news or disagreeing with the impact of computer security stories. Instead, we reworded it to be about proactively changing passwords after publicized data breaches.

We also realized that file sharing (*A15*) and social networking (*A16*) might not be widely applicable. We removed *A17* because there is no longer a clear consensus that writing down passwords is a poor practice (especially if it leads to creating stronger passwords). Similarly, we removed *A30*, which concerned sharing devices with others, because research suggests that this behavior is quite nuanced [13], and therefore this question may have been ambiguous.

Poor Item-Total Correlation

Item-total correlation measures how well the responses to one item correlate with the other items in the scale (i.e., homogeneity). Questions are deemed to not represent the rest of the scale when their item-total correlations are below 0.2 [14]. This occurred in four questions: *A2* (0.117), *A3* (0.147), *A17* (0.113), and *A29* (0.196). We observed that these questions all used qualifiers (e.g., "I *frequently*..." or "I *always*..."). Given that the purpose of the scale was for participants to report how frequently they engaged in each activity, we realized that we needed to remove the qualifiers from each statement and change the response format so that, instead of measuring agreement with a certain behavior, our scale would measure the frequency of engaging in that behavior using the options of, "always," "often," "sometimes," "rarely," and "never."

Ceiling Effects

Because a scale is designed to quantify one respondent's response relative to others, it is important that individual questions exhibit adequate variance; if everyone responds in the same manner, the scale is of little utility. Ceiling and floor effects occur when questions fail to capture sufficient variance; responses lie at one end of the spectrum because the provided responses did not adequately capture the true range of opinions. We observed that this was the case with questions *A6*, *A7*, *A8*, *A11*, *A15*, *A21*, and *A22*. The average responses to these questions were close to the maximum (i.e., $4.0 < \mu < 5.0$), and had relatively low standard deviations (i.e., $\sigma < 1.0$). As a result, we explored alternate wordings.

Additional Testing

We evaluated the impact of wording changes on responses, both with regard to variance and applicability. These iterations involved 2,184 additional participants, who we recruited in the same manner as our initial sample. Additionally, we also screened out participants who had completed a previous iteration of the survey. We produced a refined set of 24 questions that exhibited adequate variance, were applicable to almost all participants, were free from ceiling effects, and showed high item-total correlations.

REFINING THE SCALE

We recruited an additional cohort of participants so that we could perform Exploratory Factor Analysis (EFA) on the resulting 24 items. We describe our methodology and results.

Methodology

We recruited 500 new participants to respond to a set of 24 security questions (Table 2). Additionally, we randomly assigned a fifth of our sample to complete the 16-item Privacy Concerns Scale [6]. We did this to assess the discriminant validity of our scale: whether it was measuring a new construct or something already captured by existing privacy scales (PCS correlates significantly with other established privacy scales, the Westin Index [22], and the IUIPC [23]). (We piloted established psychometric tests on the participants who did not receive PCS, which we describe later in this paper.)

#	Question	N/A		μ	σ
B1	*When I'm prompted about a software update, I install it right away.*	1	(0.2%)	3.23	0.993
B2	*When my computer wants me to reboot after applying an update or installing software, I put it off.*	0	(0.0%)	^{r}3.08	1.116
B3	*I try to make sure that the programs I use are up-to-date.*	1	(0.2%)	3.89	0.914
B4	*I manually lock my computer screen when I step away from it.*	5	(1.1%)	2.58	1.386
B5	*I set my computer screen to automatically lock if I don't use it for a prolonged period of time.*	3	(0.7%)	3.47	1.552
B6	*I log out of my computer, turn it off, put it to sleep, or lock the screen when I'm done using it.*	1	(0.2%)	3.87	1.283
B7	*I use a PIN or passcode to unlock my mobile phone.*	18	(3.9%)	3.27	1.798
B8	*I use a password/passcode to unlock my laptop or tablet.*	22	(4.8%)	3.99	1.525
B9	*If I discover a security problem, I continue what I was doing because I assume someone else will fix it.*	15	(3.3%)	^{r}3.98	0.972
B10	*When someone sends me a link, I open it without first verifying where it goes.*	0	(0.0%)	^{r}3.85	1.068
B11	*I verify that my anti-virus software has been regularly updating itself.*	9	(2.0%)	3.59	1.215
B12	*When browsing websites, I mouseover links to see where they go, before clicking them.*	0	(0.0%)	3.48	1.115
B13	*I know what website I'm visiting based on its look and feel, rather than by looking at the URL bar.*	3	(0.7%)	^{r}2.92	1.107
B14	*I backup my computer.*	2	(0.4%)	2.97	1.216
B15	*When I hear about websites that have been hacked, I wait to change my passwords until I have been personally notified.*	16	(3.5%)	^{r}3.41	1.232
B16	*I use some kind of encryption software to secure sensitive files or personal information.*	20	(4.4%)	2.56	1.401
B17	*I do not change my passwords, unless I have to.*	0	(0.0%)	^{r}2.40	1.103
B18	*I use different passwords for different accounts that I have.*	1	(0.2%)	3.59	1.084
B19	*I do not include special characters in my password if it's not required.*	2	(0.4%)	^{r}2.95	1.349
B20	*When I create a new online account, I try to use a password that goes beyond the site's minimum requirements.*	2	(0.4%)	3.37	1.193
B21	*When I'm done using a website that I'm logged into, I manually log out of it.*	0	(0.0%)	3.64	1.125
B22	*I submit information to websites without first verifying that it will be sent securely (e.g., SSL, "https://", a lock icon).*	1	(0.2%)	^{r}3.65	1.103
B23	*I use privacy software, "private browsing," or "incognito" mode when I'm browsing online.*	3	(0.7%)	2.55	1.119
B24	*I clear my web browsing history.*	0	(0.0%)	3.31	1.114

Table 2. Revised set of security questions evaluated on a 5-point Likert scale (from "never" to "always") by 456 participants. Depicted are the questions, the rate of "N/A" responses, and the average responses and standard deviations after recoding negatively-phrased questions (^{r}).

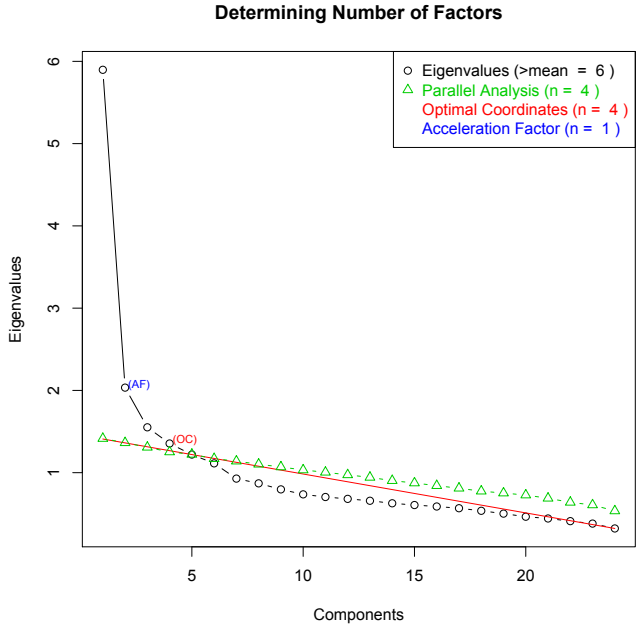

Figure 1. Scree plot showing eigenvalues from EFA, acceleration factor, optimal coordinates, and parallel analysis. We extracted 4 components.

Since we found that responses to the security questions did not correlate with social desirability, we removed the Strahan-Gerbasi scale, which also removed our second attention check (i.e., *I do not read the questions in surveys*). To compensate for this, we followed Meade and Craig's recommendation to screen out careless responses based on completion times [25]. Because reading speed is so varied and because there was no obvious threshold in our data, we removed participants whose completion times were in the top 10% (i.e., those finishing the survey in under $227s$), which left us with 456 valid responses.

Results

We examined individual questions in terms of means, standard deviations, and applicability. Table 2 shows that the ceiling effects in our initial experiments disappeared: all means were below 4.0 and standard deviations exceeded 0.9. Similarly, each question was applicable to over 95% of our participants. Thus, we proceeded to factor analysis and measured reliability. While we allowed participants to omit specific questions so that we could measure applicability, factor analysis and reliability metrics (e.g., Cronbach's alpha) require all scale items. As a result, we removed 80 participants who omitted responses to any questions, leaving us with a sample size of 376 for the remainder of our analysis. This sample size still exceeds the recommendation of Hair *et al.* to use a minimum of 5 participants per item [18].

Factor Analysis

Bartlett's test of sphericity ($\chi^2(276) = 2037.7$, $p < 0.0005$) and the Measure of Sampling Adequacy (0.869) showed that our data were correlated and measured common factors. We performed an exploratory Principal Component Analysis (PCA) and extracted six components based on the Kaiser criterion (i.e., retaining all components with eigenvalues > 1.0). Since the Kaiser criterion is considered to be overly inclusive [15], we also determined the optimal coordinates and performed parallel analysis (Figure 1). Determining optimal coordinates involves linear regression to fit a line to the smallest eigenvalues and then determining the point at which an eigenvalue diverges [33]. As can be seen in Figure 1, this test indicated that we should retain four components.

	Passwords	Securement	Awareness	Updating	
α	0.764	0.728	0.668	0.719	
IIC	0.449	0.407	0.288	0.469	ITC
B18	.742				0.512
B20	.742				0.629
B19	.735				0.580
B17	.692				0.543
B5		.760			0.526
B8		.724			0.513
B7		.720			0.502
B4		.714			0.547
B22			.727		0.496
B10			.697		0.488
B9			.628		0.402
B13			.552		0.326
B12			.540		0.403
B1				.797	0.474
B3				.773	0.589
B11				.726	0.580

Table 3. Factor loadings (PCA with Varimax rotation; loadings < 0.25 removed) from EFA. The first two rows depict reliability measures: Cronbach's α and average inter-item correlation. The last column depicts each item-total correlations.

Parallel analysis involves extracting eigenvalues from random data [19]. Any retained eigenvalues should be greater than those extracted due entirely to random sampling error. Thus, the eigenvalues generated by parallel analysis are compared to those generated by EFA, and when the i^{th} eigenvalue of the former is greater than the i^{th} eigenvalue of the latter, it is concluded that there are $i - 1$ underlying components. This corroborated our decision to extract four components.

We applied a Varimax rotation and considered an item to be loaded on a factor if its loading exceeded 0.5. We also applied Saucier's criterion of only considering an item to be loaded on a component if its loading on that component was more than twice as high as its loading on other components [36]. Based on these requirements, we removed the following seven items: *B6*, *B14*, *B15*, *B16*, *B21*, *B23*, and *B24*. Finally, we observed that removing *B2* increased Cronbach's alpha of its associated sub-scale from 0.707 to 0.719. Thus, we retained the remaining 16 items and reran PCA. The rotated factor loadings are depicted in Table 3.

The four components that we extracted predicted over 55.6% of variance. Based on each component's items, four distinct themes emerged: password generation (e.g., creating strong passwords, changing passwords), device securement (e.g., using a PIN on a smartphone, locking a desktop screen when stepping away), proactive awareness (e.g., checking links before clicking them), and updating (e.g., applying software updates in a timely manner). Each of these components accounted for 26.7%, 11.6%, 9.0%, and 8.3% of variance, respectively. Our scale thus consists of four sub-scales.

Reliability

We examined scale reliability in three parts. First we computed Cronbach's alpha for the full scale ($\alpha = 0.801$) and its sub-scales (see Table 3). We observed that our data complied with the metric used by McKinley *et al.* [24]: a multi-component scale is reliable if $\alpha > 0.6$ for all sub-scales and $\alpha > 0.7$ for a majority of sub-scales.

Next, we calculated the item-total correlation of each item, which is the Pearson correlation between the item and the average of all other items in its sub-scale. Everitt recommends a threshold of 0.2 [14], which our items all exceeded. Finally, we calculated the average inter-item correlation; Robinson *et al.* classify average inter-item correlations between 0.20 and 0.29 as "extensive" and above 0.30 as "exemplary" [35]. The lowest average inter-item correlation was found in the "awareness" component, though it was still 0.288 (i.e., "extensive"), whereas the reliability of the other three components was consistently above 0.4. Thus, by all three measures, we concluded that our sub-scales each had high reliability.

Discriminant Validity: Correlation with Privacy Concerns

We randomly selected one fifth of our participants[1] to complete an additional question set: the 16-item Privacy Concern Scale developed by Buchanan *et al.* [6]. The reason for using this scale in particular was that it is unidimensional and has high internal reliability; across the 68 participants (18% of 376) who completed it, $\alpha = 0.935$. We performed Pearson correlations between participants' average privacy concern score, and the average scores for each of our new scale's four sub-scales: *passwords* ($r = 0.151$, $p < 0.219$), *securement* ($r = 0.153$, $p < 0.213$), *awareness* ($r = 0.251$, $p < 0.039$),[2] and *updating* ($r = 0.164$, $p < 0.183$). The lack of significant correlations suggests that security behaviors are orthogonal to privacy concerns, and therefore our new scale is measuring something different.

FINALIZING THE SCALE

Based on the previous EFA, we created a final survey instrument that included only the 16 final questions (Table 4), in random order, with the N/A option removed (i.e., all questions were now required). We included additional questions from several well-studied psychometric tests in an attempt to correlate the self-reported security behaviors measured by our instrument with existing psychometrics. In this section we describe the Confirmatory Factor Analysis (CFA) that we performed and how it showed reasonable goodness of fit between our questions and the hypothesized latent factors. We show how the scale also has good reliability and how its underlying constructs relate to well-established psychometrics.

Factor Analysis

We recruited an additional cohort of 500 participants from Mechanical Turk to participate in a final experiment. We followed the recommendation of Peer *et al.* [32], who showed that restricting tasks to workers who have completed over 500 previous tasks and at least 95% approval rates negates the need for attention check questions. In addition to these two requirements, we restricted participation to those 18 or over in the U.S. who had not been exposed to one of our previous question sets. We included the initial instructed response question, but did not do any post hoc screening.

[1] As we noted earlier, the remaining participants completed other psychometrics, which we piloted prior to our CFA experiment, discussed in the next section.

[2] Upon applying the Bonferroni correction to account for multiple testing, we do not consider this to be statistically significant.

#	*Device Securement* (28.47% of variance explained; $\lambda = 4.555$)	μ	σ
F4	I set my computer screen to automatically lock if I don't use it for a prolonged period of time.	3.20	1.559
F6	I use a password/passcode to unlock my laptop or tablet.	3.78	1.525
F3	I manually lock my computer screen when I step away from it.	2.63	1.343
F5	I use a PIN or passcode to unlock my mobile phone.	3.21	1.733
#	*Password Generation* (12.95% of variance explained; $\lambda = 2.071$)	μ	σ
F12	I do not change my passwords, unless I have to.^{r}	2.65	1.091
F13	I use different passwords for different accounts that I have.	3.75	1.037
F15	When I create a new online account, I try to use a password that goes beyond the site's minimum requirements.	3.31	1.096
F14	I do not include special characters in my password if it's not required.^{r}	3.30	1.292
#	*Proactive Awareness* (8.36% of variance explained; $\lambda = 1.337$)	μ	σ
F8	When someone sends me a link, I open it without first verifying where it goes.^{r}	4.01	1.014
F11	I know what website I'm visiting based on its look and feel, rather than by looking at the URL bar.^{r}	3.17	1.077
F16	I submit information to websites without first verifying that it will be sent securely (e.g., SSL, "https://", a lock icon).^{r}	3.69	1.102
F10	When browsing websites, I mouseover links to see where they go, before clicking them.	3.69	1.027
F7	If I discover a security problem, I continue what I was doing because I assume someone else will fix it.^{r}	4.08	0.976
#	*Updating* (6.77% of variance explained; $\lambda = 1.082$)	μ	σ
F1	When I'm prompted about a software update, I install it right away.	3.07	1.035
F2	I try to make sure that the programs I use are up-to-date.	3.78	0.890
F9	I verify that my anti-virus software has been regularly updating itself.	3.55	1.228

Table 4. The final questions for the Security Behavior Intentions Scale and associated sub-scales. The means of reverse-scored questions (denoted by ^{r}) have been recoded. The final question numbers have been enumerated from F1-F16 in order to differentiate them from prior exploratory questions. Responses were reported on the following scale: *Never (1)*, *Rarely (2)*, *Sometimes (3)*, *Often (4)*, and *Always (5)*.

Participants' genders were split fairly evenly: 46.8% were female, 52.8% were male, and two declined to specify. Participants' ages ranged from 19 to 71 ($\mu = 34.3$, $\sigma = 10.78$), median income was between $35,000 and $50,000, and 37.8% of participants held bachelor's degrees (an additional 14% had completed some or more graduate school). We therefore believe our sample is comparable to our previous samples and is also representative of the U.S. online population.

We performed PCA using a Varimax rotation, extracted four components, and observed that all items were loaded on the same unique component that they were in the previous experiment. These four components predicted 56.6% of variance, and are shown in Table 4 along with the mean responses.

Next, we performed CFA by examining the goodness-of-fit of our data to the latent variable model. Multiple metrics showed that our data adhered to the model. We did not rely on the chi-square goodness-of-fit test, because of our large sample size [44]. Instead, we examined the relative chi-square [28], the Root Mean Square Error of Approximation (RMSEA), the Standardized Root Mean Square Residual (SRMR), the Comparative Fit Index (CFI), and the Tucker-Lewis Index (TLI).

Our relative chi-square statistic, χ^2/df, was approximately 2.71. While there is no firm consensus on the exact cutoff value to indicate a "good" fit, our statistic is below 3, which is the more stringent requirement recommended by Hair *et al.* [18]. Our analysis yielded a RMSEA of 0.058 and a SRMR of 0.050, which are both below recommended cutoff points; Hu and Bentler recommend a cutoff of 0.06 for RMSEA and 0.08 for SRMR [20]. Finally, our CFI and TLI were 0.920 and 0.902, respectively, which are above the 0.90 cutoff recommended by Netemeyer *et al.* [29].

Reliability

We calculated the composite reliability of the four sub-scales (Table 5) [17]. Since these values are all above the threshold of 0.6 recommended by Bagozzi and Yi [3]. Additionally, we conducted a test-retest procedure to make sure responses were stable over time. Eleven days after our 500 participants had completed our survey, we invited them to participate in a followup survey for an additional $1 payment. Of the 500 invited, 354 participated. We calculated the average measures intraclass correlation coefficient between their sub-scale average scores (Table 5), and found that all of them were above the 0.6 threshold recommended by Weir [45].

	Securement	Passwords	Awareness	Updating
CR	0.78	0.78	0.64	0.69
ICC	0.89	0.86	0.81	0.81

Table 5. Each sub-scale's composite reliability (CR) and average measures intraclass correlation coefficient (ICC) during test-retest.

Relation to Psychological Constructs

We explored whether certain psychological constructs may be correlated with an individual's security behaviors. Thus, participants completed several additional scales:

- **Domain-Specific Risk-Taking Scale (DoSpeRT)** [4], which measures propensity for risk-taking across five dimensions: ethical (1), financial (2), health/safety (3), recreational (4), and social (5).
- **General Decision-Making Style (GDMS)** [37], which measures how people approach decision-making with regard to five dimensions: rational (1), avoidant (2), dependent (3), intuitive (4), and spontaneous (5).
- **Need for Cognition (NFC)** [7], which is a unidimensional scale that measures the propensity to engage in "cognitive endeavors."
- **Barratt Impulsiveness Scale (BIS)** [31], which measures impulsivity across three dimensions: attention (1), motor (2), and non-planning (3).
- **Consideration for Future Consequences (CFC)** [21], which is a unidimensional scale measuring how much attention people pay to long-term consequences.

	Securement	Passwords	Awareness	Updating
$DoSpeRT_1$		**-.201	**-.226	**-.201
$DoSpeRT_2$				
$DoSpeRT_3$			**-.204	**-.164
$DoSpeRT_4$				
$DoSpeRT_5$		*.141		
$GDMS_1$		**.145	**.224	**.229
$GDMS_2$	*-.133	**-.220	**-.230	**-.247
$GDMS_3$			**-.157	
$GDMS_4$				
$GDMS_5$				*-.129
NFC	**.164	**.290	**.231	**.253
BIS_1		**-.243	*-.140	**-.218
BIS_2			**-.147	**-.145
BIS_3		**-.235	**-.171	**-.247
CFC	**.184	**.317	**.307	**.303

Table 6. Correlations between sub-scales and various psychometrics. $^*p < 0.005$**,** $^{**}p < 0.001$**.**

Because this work is exploratory, we performed Pearson correlations rather than building a regression model (Table 6). To counteract effects from repeated testing, we only report correlations significant at the $p < 0.005$ level. Across all four sub-scales, participants who were more inquisitive (as determined by the NFC scale) were more likely to report engaging in better security practices. We also noted that the highest correlation across all four sub-scales was with consideration for future consequences (CFC), suggesting that good security behaviors are tied to long-term thinking.

Our initial hypothesis was that willingness to take risks involving health or safety would be inversely correlated with computer security behaviors; we observed this was true, but only for behaviors involving proactive awareness and keeping software updated. Similarly, across all four sub-scales, people who engaged in "better" security behaviors were less likely to procrastinate (as determined by the GDMS avoidant sub-scale). People scoring low on GDMS dependence scored high on SeBIS awareness; being proactive about computer security means *not* relying on other people. Many of the security behaviors also correlated inversely with impulsivity: this suggests that adhering to security advice involves foresight.

DISCUSSION

The final Security Behavior Intentions Scale (SeBIS) consists of 16 items, which map uniquely onto four factors: device securement (locking devices using passwords, PINs, etc.), password generation (creating and using passwords), proactive awareness (noticing and taking into account environmental security cues), and updating (ensuring software is up-to-date). The scale showed high psychometric properties: all subscales were internally consistent, all items loaded highly on their respective component and did not load highly on other components, correlations between sub-scales were low, two administrations of the same test to the same sample showed high correlations, and some of the sub-scales were correlated with some other, well-established, psychological constructs.

Both researchers and practitioners can use this new scale for various ends. Researchers may use it to measure intended security behaviors and examine how these change between different populations, over time, or following educational interventions. They may also use the scale to determine users' involvement in security-related situations, which could mediate how users may respond to different security mitigations. For example, researchers who are interested in compliance with security warnings may use the scale to explore whether some types of warnings are more effective for one type of user or another. The scale may also be used to explore its predictive ability of real world behaviors (e.g., falling for phishing attacks) and estimate the effectiveness of various interventions (e.g., security education) by measuring changes in the scale following the interventions, and over time.

Practitioners may use this scale in various organizational settings. First, they may use it to estimate the security behaviors among their employees and to identify prevalent weaknesses which could then be remedied using targeted interventions. This may help reduce the cost, and increase the effectiveness, of organizational training aimed at improving end-user security behavior. Second, they may use the scale to examine the common behaviors among their employees in order to try to match the organization's security policy with employees' natural behavior. For example, if an organization realizes that many of its employees do not lock their computers when stepping away, the organization may instigate a policy of forcing all computers to be locked after a short period of idle time. Lastly, managers may use the scale to identify "extreme" individuals (both on the high and low ends of the scale) and take targeted actions toward these individuals (e.g., reward the highly secure ones and educate the lowly secure ones).

Another potential implication of this scale could be for public policy. As described in the beginning, current computer security recommendations targeted at end-users are highly scattered, inconclusive and inconsistent. This could be very confusing to end-users who are not highly educated on this topic. Instead, a different type of policy, or set of policies, could be formed if policy makers knew the current prevalent attitudes and behaviors of the general public. The scale we developed can help assess these attitudes and behaviors on a large scale, and inform policy makers on the specific domains and situations that pose a potentially higher risk due to users' lower adherence to security standards.

Future Work

While we established the reliability and content validity of SeBIS in this paper, there is still much work to be done. We showed that it correlates with several well-studied psychometrics. However, it is unclear how they interact with each other and how they could be used to predict security behaviors. In future work, we expect to perform studies to examine other psychometrics and build a predictive model. One of our ultimate goals is to be able to use existing psychometrics to infer a user's computer security attitudes and behaviors so that adequate mitigations can be tailored to their needs and abilities based on observations of other behaviors.

Some may be concerned that SeBIS targets both personal and business contexts at the same time. While the final scale contains questions that are likely to be *impacted* by workplace policies (e.g., screen locking, software updates), we believe

this is a minor concern: like the vast majority of psychometric tests, responses are self-reported, and self-reports are impacted by exogenous factors (i.e., not just personal factors). We do not presume to explain *why* people report behaving the way they do, just how secure they report behaving. In fact, if our measure is impacted by policies, it actually makes it more relevant for organizational experiments: it can show how a change in workplace policy affected (or failed to affect) employee behaviors. Still, we agree that this issue does need to be addressed empirically, which is why studies are needed to compare SeBIS results between corporate and home users.

Similarly, while SeBIS is reliable and stable over time, it is not clear how well it correlates with actual security behaviors, since it only measures intentions. We plan to perform experiments to examine how each of its component sub-scales correlate with behaviors, as well as how other psychometrics could be used to predict computer security behaviors. Given its demonstrated reliability and discriminant validity, our results show that SeBIS is still a useful tool for measuring *intentions* and how they might change over time or vary between individuals. However, if construct validity is also strong, then in many cases SeBIS may be a reasonable substitute for performing costly and time-consuming laboratory experiments.

ACKNOWLEDGMENTS

This work was supported by NSF under award CNS-1343433. We would also like to thank Alessandro Acquisti, Chris Hoofnagle, Diogo Marques, David Wagner, members of the Berkeley Laboratory for Usable and Experimental Security (BLUES), and Refjohürs Lykkewe.

REFERENCES

1. Acquisti, A., and Grossklags, J. Losses, gains, and hyperbolic discounting: An experimental approach to information security attitudes and behavior. In *Proceedings of The 2nd Annual Workshop on Economics and Information Security (WEIS '03)* (2003).
2. Acquisti, A., and Grossklags, J. Privacy and rationality in individual decision making. *IEEE Security & Privacy* (January/February 2005), 24–30.
3. Bagozzi, R. P., and Yi, Y. On the evaluation of structural equation models. *Journal of the academy of marketing science 16*, 1 (1988), 74–94.
4. Blais, A.-R., and Weber, E. U. A domain-specific risk-taking (dospert) scale for adult populations. *Judgment and Decision Making 1*, 1 (2006), 33–47.
5. Bossler, A. M., and Holt, T. J. The effect of self-control on victimization in the cyberworld. *Journal of Criminal Justice 38*, 3 (2010), 227–236.
6. Buchanan, T., Paine, C., Joinson, A. N., and Reips, U.-D. Development of measures of online privacy concern and protection for use on the internet. *Journal of the American Society for Information Science and Technology 58*, 2 (2007), 157–165.
7. Cacioppo, J. T., Petty, R. E., and Feng Kao, C. The efficient assessment of need for cognition. *Journal of personality assessment 48*, 3 (1984), 306–307.
8. Copes, H., Kerley, K. R., Huff, R., and Kane, J. Differentiating identity theft: An exploratory study of victims using a national victimization survey. *Journal of Criminal Justice 38*, 5 (2010), 1045–1052.
9. Cranor, L. F. A framework for reasoning about the human in the loop. In *Proceedings of the 1st Conference on Usability, Psychology, and Security*, USENIX Association (Berkeley, CA, 2008).
10. Crowne, D. P., and Marlowe, D. A new scale of social desirability independent of psychopathology. *Journal of consulting psychology 24*, 4 (1960), 349.
11. DeVellis, R. F. *Scale Development: Theory and Applications*, 2nd ed., vol. 26 of *Applied Social Research Methods Series*. Sage Publications, 2003.
12. Egelman, S., Cranor, L. F., and Hong, J. You've been warned: An empirical study of the effectiveness of web browser phishing warnings. In *Proceeding of The 26th SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, ACM (New York, NY, USA, 2008), 1065–1074.
13. Egelman, S., Jain, S., Portnoff, R. S., Liao, K., Consolvo, S., and Wagner, D. Are you ready to lock? understanding user motivations for smartphone locking behaviors. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer & Communications Security*, CCS '14, ACM (New York, NY, USA, 2014).
14. Everitt, B. S. *The Cambridge Dictionary of Statistics*. Cambridge University Press, Cambridge, UK, 2002.
15. Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., and Strahan, E. J. Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods 4*, 3 (1999), 272.
16. Felt, A. P., Ha, E., Egelman, S., Haney, A., Chin, E., and Wagner, D. Android permissions: user attention, comprehension, and behavior. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, SOUPS '12, ACM (New York, NY, USA, 2012).
17. Fornell, C., and Larcker, D. F. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research 18*, 1 (1981), pp. 39–50.
18. Hair, J. F., Tatham, R. L., Anderson, R. E., and Black, W. *Multivariate Data Analysis*, 6 ed. Prentice Hall, 2006.
19. Horn, J. L. A rationale and test for the number of factors in factor analysis. *Psychometrika 30*, 2 (1965), 179–185.
20. Hu, L.-t., and Bentler, P. M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal 6*, 1 (1999), 1–55.
21. Joireman, J., Shaffer, M. J., Balliet, D., and Strathman, A. Promotion orientation explains why future-oriented people exercise and eat healthy evidence from the two-factor consideration of future consequences-14

scale. *Personality and Social Psychology Bulletin 38*, 10 (2012), 1272–1287.

22. Kumaraguru, P., and Cranor, L. F. Privacy Indexes: A Survey of Westin's Studies. Tech. Rep. CMU-ISRI-5-138, Carnegie Mellon University, December, 2005. `http://reports-archive.adm.cs.cmu.edu/anon/isri2005/abstracts/05-138.html`.
23. Malhotra, N. K., Kim, S. S., and Agarwal, J. Internet users' information privacy concerns (iuipc): The construct, the scale, and a causal model. *Information Systems Research 15*, 4 (December 2004), 336–355.
24. McKinley, R. K., Manku-Scott, T., Hastings, A. M., French, D. P., and Baker, R. Reliability and validity of a new measure of patient satisfaction with out of hours primary medical care in the united kingdom: development of a patient questionnaire. *British Medical Journal 314*, 7075 (1997), 193.
25. Meade, A. W., and Craig, S. B. Identifying careless responses in survey data. *Psychological methods 17*, 3 (2012), 437.
26. Modic, D., and Lea, S. E. G. How neurotic are scam victims, really? the big five and internet scams. `http://ssrn.com/abstract=2448130`, September 2012.
27. Moon, B., McCluskey, J. D., and McCluskey, C. P. A general theory of crime and computer crime: An empirical test. *Journal of Criminal Justice 38*, 4 (2010), 767–772.
28. Mueller, R. O. *Basic principles of structural equation modeling: An introduction to LISREL and EQS*. Springer, 1996.
29. Netemeyer, R. G., Bearden, W. O., and Sharma, S. *Scaling Procedures: Issues and Applications*. SAGE Publications, 2003.
30. Ng, B.-Y., Kankanhalli, A., and Xu, Y. C. Studying users' computer security behavior: A health belief perspective. *Dec. Sup. Systems 46*, 4 (2009), 815–825.
31. Patton, J. H., Stanford, M. S., et al. Factor structure of the barratt impulsiveness scale. *Journal of clinical psychology 51*, 6 (1995), 768–774.
32. Peer, E., Vosgerau, J., and Acquisti, A. Reputation as a sufficient condition for data quality on amazon mechanical turk. *Behavior Research Methods 45*, 4 (December 2013).
33. Raîche, G., Walls, T. A., Magis, D., Riopel, M., and Blais, J.-G. Non-graphical solutions for cattell's scree test. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences 9*, 1 (2013), 23.
34. Ray, J. J. Reviving the problem of acquiescent response bias. *The J. of Social Psychology 121*, 1 (1983), 81–96.
35. Robinson, J. P., Shaver, P. R., and Wrightsman, L. S. Criteria for scale selection and evaluation. In *Measures of personality and social psychological attitudes*. Academic Press, 1991, ch. 1, 1–16.
36. Saucier, G. Mini-markers: A brief version of goldberg's unipolar big-five markers. *Journal of Personality Assessment 63*, 3 (1994), 506–516.
37. Scott, S. G., and Bruce, R. A. Decision-making style: The development and assessment of a new measure. *Educational and psychological measurement 55*, 5 (1995), 818–831.
38. Sherizen, S. Criminological concepts and research findings relevant for improving computer crime control. *Computers & Security 9*, 3 (1990), 215–222.
39. Stanton, J. M., Stam, K. R., Mastrangelo, P., and Jolton, J. Analysis of end user security behaviors. *Computers & Security 24*, 2 (2005), 124–133.
40. Strahan, R., and Gerbasi, K. C. Short, homogeneous versions of the marlowe-crowne social desirability scale. *Journal of clinical psychology* (1972).
41. United States Computer Emergency Readiness Team. Tips. `https://www.us-cert.gov/ncas/tips`. Accessed: September 12, 2014.
42. United States Department of Homeland Security. National Cyber Security Awareness Month 2014. `http://www.dhs.gov/national-cyber-security-awareness-month-2014`, September 8 2014. Accessed: September 12, 2014.
43. Verizon. Security. `http://www.verizon.com/Support/Residential/Internet/FiosInternet/General+Support/Security/Security.htm`, 2014. Accessed: September 12, 2014.
44. Wang, L., Fan, X., and Willson, V. L. Effects of nonnormal data on parameter estimates and fit indices for a model with latent and manifest variables: An empirical study. *Structural Equation Modeling: A Multidisciplinary Journal 3*, 3 (1996), 228–247.
45. Weir, J. P. Quantifying test-retest reliability using the intraclass correlation coefficient and the sem. *The Journal of Strength & Conditioning Research 19*, 1 (2005), 231–240.
46. Wilson, M., and Hash, J. Building an Information Technology Security Awareness and Training Program. Special Publication 800-50, National Institute of Standards and Technology, Gaithersburg, MD, US, October 2003. `http://csrc.nist.gov/publications/nistpubs/800-50/NIST-SP800-50.pdf`.
47. Wogalter, M. S. Communication-Human Information Processing (C-HIP) Model. In *Handbook of Warnings*, M. S. Wogalter, Ed. Lawrence Erlbaum Associates, 2006, 51–61.
48. Woodruff, A., Pihur, V., Consolvo, S., Brandimarte, L., and Acquisti, A. Would a privacy fundamentalist sell their dna for $1000...if nothing bad happened as a result? the westin categories, behavioral intentions, and consequences. In *Proceedings of the 2014 Symposium on Usable Privacy and Security*, USENIX Association (2014), 1–18.