Toward a Security Behavior Intentions Scale Robust to Linguistic Differences

Yukiko Sawaya
KDDI R&D Laboratories, Inc.
2-1-15 Ohara Fujimino
Saitama, 356-8502, Japan
yu-sawaya@kddilabs.jp

Mahmood Sharif
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA, 15213, USA
mahmoods@cmu.edu

Nicolas Christin
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA, 15213, USA
nicolasc@cmu.edu

Ayumu Kubota
KDDI R&D Laboratories, Inc.
2-1-15 Ohara Fujimino
Saitama, 356-8502, Japan
kubota@kddilabs.jp

Akihiro Nakarai
KDDI R&D Laboratories, Inc.
2-1-15 Ohara Fujimino
Saitama, 356-8502, Japan
ak-nakarai@kddilabs.jp

Akira Yamada
KDDI R&D Laboratories, Inc.
2-1-15 Ohara Fujimino
Saitama, 356-8502, Japan
ai-yamada@kddilabs.jp

Abstract
We are interested in understanding how cultural differences impact people's attitudes toward security. As a first, preliminary, step toward this goal, we translated Egelman and Peer's Security Behavior Intention Scale (SeBIS) into Japanese, to determine whether general attitudes and behaviors could be replicated in Japanese. Through a pilot study of 1,654 Japanese-speaking users we discovered that the model construct previously reported was no longer valid. We found these inconsistencies were caused by (a) negative interrogative sentences and (b) questions which led to ambiguous translations. Using a couple of simple modifications, we propose a revised SeBIS questionnaire to attempt to correct both factors. We tested the revised scale on a new cohort of Japanese and English speaking participants. Our results suggest that the revised scale is robust to linguistic differences, albeit with different outcomes across cultures.

Author Keywords
Security behavior; Confirmatory factor analysis.

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

Understanding people's attitudes toward security and privacy is essential to devise effective human-centered defenses. We posit that cultural differences may considerably impact those attitudes, yet, the impact of these cultural differences has so far been under-studied. As a first step, we attempt to generalize the Security Behavior Intention Scale (SeBIS, [1]) by investigating to which extent it can apply to Japanese users. We find that merely translating the SeBIS scale into Japanese does not work: that is, statistical tests of model fit lead to rejecting the model previously validated for English speakers. We then focus on potential reasons for this discrepancy and discover that the original English scale uses double-negations and presents some ambiguities, which could have led to this negative result. We then refine the SeBIS scale to correct these potential problems, test the revised scale with both Japanese and US-based users, and show that the resulting scale is now robust to translating between English and Japanese, but that, interestingly, the results remain different between both population groups.

Preliminary Experiment

We started by translating the original, English, SeBIS scale into Japanese. The translation was conducted by the native Japanese speakers in our team; this ensured that the translation was performed by subject matter experts.

Relying on a survey company, we first randomly selected a pool of 30,000 potential participants in Japan. We then recruited, on February 13, 2016, 1,654 participants from this pool, so that our recruited sample's demographics (gender and age) statistically matches those of Japanese Internet users as reported in the 2010 national census [6]. The survey company administered the survey online on our behalf. We then performed confirmatory factor analysis (CFA) to evaluate the goodness of fit of the SeBIS scale for Japanese users.

Table 1 summarizes our results. Let us focus for now on the two leftmost columns. The first column presents Cronbach's α reliability measure and several data fit indices (RMSEA, SRMR, CFI and TLI) for the original, English, SeBIS scale [1]. The second column presents the same indices for the literal Japanese translation ("preliminary experiment") of the SeBIS scale. Contrary to the English scale, the RMSEA and SRMR are both above the cutoff points recommended by Hu and Bentler [3]; and the CFI and TLI are both below the 0.90 cutoff recommended by Netemeyer et al. [5]. Thus, these statistical tests indicate poor model fit, and lead us to reject the literal translation of the SeBIS scale.

Possible Causes

We next investigate the causes for the poor model fit we observed. To do so, we formulate the following hypotheses:

H1: Demographic differences between the two populations cause the poorness of fit. The original SeBIS paper relied on participants aged 19 to 71 (average: 34.3, SD: 10.78) and 46.8% were female and 52.8% were male. On the other hand, our participant pool – aiming to be representative of the Internet demographics in Japan – ranges from 15 to 69 (average: 44.3, SD: 14.80, male: 51.5%, female: 48.5%). That is, our sample is slightly older on average than the population sample used by Egelman and Peer; if age negatively impacts security behavior, we would expect to get worse goodness of fit values.

H2: Inadequate filtering of participants resulted in poor goodness of fit. Contrary to Egelman and Peer [1], our initial translation did not feature attention questions aimed at discarding participants who answer without paying enough

Experiment #	Original	Preliminary	H1	H2	H3	H4	
scale	SeBIS(en) [1]	SeBIS(ja)	SeBIS(ja)	SeBIS(ja)	RSeBIS(ja)	RSeBIS(en)	recommended
N	500	1,654	505	1,654	1,654	408	
Cronbach's α	0.801*	0.794*	0.814*	0.685*	0.893*	0.86*	>0.6 [1]
RMSEA	0.058*	0.087	0.089	0.095	0.062	0.061	<0.06 [3]
SRMR	0.050*	0.090	0.097	0.107	0.042*	0.054*	<0.08 [3]
CFI	0.920*	0.818	0.828	0.753	0.949*	0.934*	>0.9 [5]
TLI	0.902*	0.777	0.789	0.698	0.938*	0.919*	>0.9 [5]

Table 1: Model fit indicators. The first column represents the values indicated by Egelman and Peer. The second to fourth column show values for tests across various population samples using a literal Japanese translation of the scale. The fifth column shows the values for our proposed revised scale in Japanese, and the sixth column shows the fit for our revised scale in English. Stars indicate good model fit.

attention to their answer. We hypothesize that such "bogus" answers might have caused the poorness of fit.

H3: Linguistic particularities impact the SeBIS scale. Literal translation does not account for certain linguistic particularities. For instance, answers to negative interrogative sentences may differ between languages. Likewise, certain concepts (e.g., "security") may be more ambiguous in a language than in another. We hypothesize that the poorness of model fit is (at least partly) due to such particularities.

Hypothesis Testing

To test H1, we sampled a set of 505 participants from our larger participant pool, designed to match the demographics of the population sample Egelman and Peer used [1], that is, ages ranging from 18 to 69 (average: 34.6, SD: 10.87) and 44.8% were female and 55.2% were male. The results, shown in Table 1 (column H1), indicate a poor fit, similar to our preliminary experiment. We reject H1.

To test H2, we recruited another cohort of 1,654 participants without overlap with those who participated in our preliminary experiment or in our test of H1. On March 9–10, 2016 we rerun the test of the literal Japanese translation of the SeBIS scale, adding attention questions identical to those used by Egelman and Peer [1] to weed out participants who did not read questions carefully. The results, presented in Table 1, indicate that the model fit remains poor, despite the inclusion of these attention questions. We thus rule out H2.

We are left with H3. We first performed correlation analysis to determine if certain questions were more problematic than others. We discovered that reverse-scored questions, in which lower scores indicate more insecure behavior, had a negative correlation with the other questions (in which higher scores indicate more insecure behavior). In short, reverse-scored questions appear to have been a possible cause of problems in our original translation.

One of the root causes may be the use of negative interrogation sentences, whose answers are opposite in English and Japanese. For instance, to the question "aren't you tired?" an English speaker could answer "no, I am not tired." A Japanese speaker, to denote the exact same state, would answer (literally translating) "yes, I am not tired." Another possible issue we identified is that some of the reverse-scored questions appear more ambiguous when translated

literally in Japanese, than they are in the original English phrasing. For example, the question "I know what website I'm visiting based on its look and feel, rather than by looking at the URL bar" combines somewhat abstract concepts ("based on its look and feel") with a more concrete proposition ("looking at the URL bar"). We conjecture that this shift may have confused Japanese speakers.

To test H3, we rephrased all of the questions that were originally reverse-scored so that they matched the logical order of the other questions, and attempted to make them as unambiguous as possible; we call the revised scale Refined SeBIS (RSeBIS).

We recruited another 1,654 participants, following the same demographics as in our preliminary experiment, and tested RSeBIS on March 9 and 10, 2016. The results, shown in Table 1 (column H3) indicate a good fit, pretty much equivalent to the goodness of fit of the original scale in English.

Running RSeBIS in the US

After retaining H3, we wanted to validate that the refined scale remains effective in English. To this end, we rephrased reverse-worded questions in the original SeBIS scale, and ran the refined survey on Amazon's Mechanical Turk to test the following hypothesis.

H4: RSeBIS presents good model fit in English and for US populations as well.

We recruited 408 participants from the US, with demographics comparable to those used in the original work [1]. We used Cronbach's α and the Composite Reliability measures to validate RSeBIS' reliability. We found that the reliability measures exceed the desired minimum expected from a reliable scale ($\alpha > 0.7$ for the entire scale, $\alpha > 0.6$ for each sub-scale, and Composite Reliability > 0.6).

Further, as shown in Table 1, the values of all the goodness-of-fit measures are very close to, or well within the ranges of recommended values. Therefore, we retain H4.

Differences between US and Japan While RSeBIS provides a good model fit with both US and Japanese users, the results to the survey are nevertheless quite different. Using Mann-Whitney-Wilcoxon tests we discovered that Japanese users tend to espouse more insecure behavior for a number of subscales (device securement, password generation, proactive awareness, updating behavior), and over the entire SeBIS scale ($W = 523,710$, $p < 1.1 \times 10^{-15}$ after applying Bonferroni correction). Interpreting this result remains future work, but we do note with interest that it seems to mirror, to some extent, independent observations by Harbach et al. in the context of mobile phone security [2].

Related Work

Besides the aforementioned recent work by Harbach et al. [2], a few studies on cross cultural differences of users with respected to security exist. Of note, Murayama et al. [4], empirically compare the "sense of security" of non-computer science students in the US and Japan. Using exploratory factor analysis, Murayama et al. identify four factors contributing to the sense of security both in Japan and in the US, two of which being common to both cultures.

Conclusion and Future Work

We discovered that accounting for linguistic particularities plays an essential role in making behavioral scales like the SeBIS scale robust across languages. Our work is a first foray toward designing culturally independent behavioral scales, and, in turn, analyzing the impact of cultural differences on security attitudes and effectiveness of mitigations.

REFERENCES

1. Serge Egelman and Eyal Peer. 2015. Scaling the Security Wall: Developing a Security Behavior Intentions Scale (SeBIS). In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2873–2882. DOI: http://dx.doi.org/10.1145/2702123.2702249

2. Marian Harbach, Alexander De Luca, Nathan Malkin, and Serge Egelman. 2016. Keep on Lockin' in the Free World: A Multi-National Comparison of Smartphone Locking. In *Proceedings of the 34th Annual ACM Conference on Human Factors in Computing Systems (CHI'16)*. ACM, New York, NY, USA.

3. Li-Tze Hu and Peter M. Bentler. 1999. 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.

4. Yuko Murayama, Natsuko Hikage, Yasuhiro Fujihara, and Carl Hauser. 2007. The structure of the sense of security, anshin. In *CRITIS'07 Proceedings of the Second international conference on Critical Information Infrastructures Security*. Springer, 83–93.

5. Richard G. Netemeyer, William O. Bearden, and Subhash Sharma. 2003. *Scaling procedures: Issues and applications*. Sage Publications.

6. Statistics Bureau. 2011. 2010 National Census in Japan (in Japanese). http://www.stat.go.jp/data/kokusei/2010/index.htm. (2011). [Online; accessed 5/10/2016].