Transactions on Replication Research

Conceptual Replication DOI: 10.17705/1atrr.00037 ISSN 2473-3458

Fear Appeals and Information Security Behaviors: An Empirical Study on Mechanical Turk

Sagar Samtani
Information Systems and Decision Sciences
University of South Florida
ssamtani@usf.edu

Hongyi Zhu
Department of Management Information Systems
University of Arizona
zhuhy@email.arizona.edu

Shuo Yu
Department of Management Information Systems
University of Arizona
shuoyu@email.arizona.edu

Abstract:

This study aims to conduct a methodological replication of the information security study conducted by Johnston and Warkentin (2010). This study leveraged the use of the fear appeals model (FAM) in the context of information security as they pertain to the individual use of anti-spyware software. We adopt all measures, instruments, statistical tests, theory, and models from the original study, but apply them to the Amazon Mechanical Turk population. The results from this replication study are not consistent with the original study, in that two of the five posited hypotheses have opposite effects than those originally found; threat severity is shown to have a positive effect on both response efficacy and self-efficacy, where in the original study, this is shown to have a negative effect on both. The results imply that there may be differences in which populations the study was conducted, thus requiring additional samples and statistical tests.

Keywords: Fear Appeal Model, Information Security, Amazon Mechanical Turk, Anti-spyware Software, Methodological Replication

The manuscript was received 01/29/2017 and was with the authors 9 months for two revisions.

1 Introduction

As information technology use increases, cybersecurity is becoming a concern for companies and individuals alike. Some key aspects of cybersecurity include password security, vulnerability assessment, and understanding hacker motivations. One aspect of cybersecurity receiving growing attention focuses on ensuring that employees within organizations utilize the proper security tool such as a password manager, anti-virus software, firewall, or anti-spyware software. The lack of proper use of such tools has the potential to cause significant harm to an individual or the organization as a whole. As a result, Information Systems (IS) researchers have attempted to understand the behaviors and motivations of individuals as they pertain to the compliance of using these different software tools.

One work aiming to understand an individual's anti-spyware software usage compliance is a 2010 paper published in *Management Information Systems Quarterly (MISQ)* by Allen C. Johnston and Merrill Warkentin entitled *"Fear Appeals and Information Security Behaviors: An Empirical Study."* This research is interested in the compliance of end users with regard to anti-spyware technology, aiming to answer the research question, "How do fear appeals modify end user behavioral intentions associated with recommended individual computer security actions?" (Johnston and Warkentin, 2010, p. 550). Fear appeals are "persuasive messages designed to scare people by describing terrible things that will happen to them if they do not do what the message recommends" (Witte, 1992, p. 329). To answer such a question, Johnston and Warkentin conducted a survey-embedded experiment with three groups of randomly assigned members of academic departments. The experiment was driven by the theories of fear appeals model (FAM) and protection motivation theory (PMT).

While the FAM and PMT have been applied in various information security contexts, little work has attempted to replicate the models and methods used by these studies. As such, this paper aims to be a methodological replication, wherein all theories, methods, and hypotheses are adopted from the original study. Replication is one of the tenets of the scientific method, which allows for scientific consensus to emerge and provides validation of previous studies (Dennis and Valacich, 2014). This paper is selected for replication for two reasons. First, it focuses on cybersecurity, a high-impact and emerging IS research area. Second, this paper follows an experimental approach, which is conducive to replication, as experiments allow for a greater deal of control as compared to other behavioral approaches (Dennis and Valacich, 2014).

As this paper aims to be a methodological replication, we first present the model and hypotheses adopted from the original study (Figure 1 and Table 1, respectively). However, rather than adopt the sample of academic faculty, staff, and students, we select a sample of Amazon Mechanical Turk users. Amazon Mechanical Turk has been identified to contain participants beyond the United States (US) academic subjects. This characteristic provides the opportunity to identify if the results presented in the original study are generalizable to populations beyond those in the US academic realm.

Table 1. Research Hypotheses

N	Hypotheses
H_1	Response efficacy will have a positive effect on end user intentions to adopt recommended individual computer security actions with respect to spyware.
H_2	Self-efficacy will have a positive effect on end user intentions to adopt recommended individual computer security actions with respect to spyware.
H_3	Social influence will have a positive effect on end user intentions to adopt recommended individual computer security actions with respect to spyware.
H_{4a}	Perceptions of threat severity will negatively influence perceptions of response efficacy.
H_{4b}	Perceptions of threat severity will negatively influence perceptions of self-efficacy.
H_{5a}	Perceptions of threat susceptibility will negatively influence perceptions of response efficacy.
H_{5b}	Perceptions of threat susceptibility will negatively influence perceptions of self-efficacy.

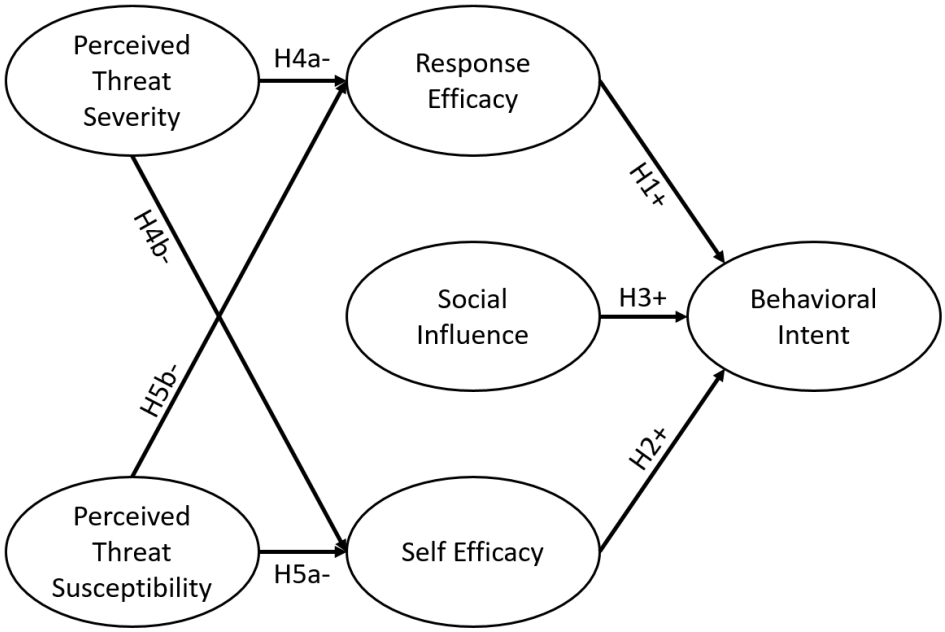


Figure 1. Research Model

The remainder of this paper is organized as follows. We first present the model and hypotheses posited in the original paper. We then present our methodology, focusing on the adjustments made for the Mechanical Turk context. Next, we summarize our results and discuss the implications. Subsequently, we highlight limitations for our work and point to several promising next steps.

2 Method

Researchers in the original study conducted an experiment by sending out surveys and fear appeals via email to members of various academic departments. Three groups were randomly assigned. Group 1, which had 215 participants, received a pre-test, a manipulation (i.e., the fear appeal message), and a post-test. Group 2 (N=30) received the pre-test and post-test, and Group 3 (N=30) received the manipulation and the post-test. The pre and post-tests have the same survey questions, while the manipulation was a fear appeal in the form of an email (can be seen in Appendix A).

As this is a methodological replication, instead of faculty and students, we opted for Amazon Mechanical Turk participants because it is easier to get a large sample from various backgrounds with fast response. Amazon Mechanical Turk is an online crowdsourcing market in which employees (workers) complete tasks for employers (requesters) (Steelman et al., 2014). The tasks that workers can do can be cleaning data, creating keywords, audio transcriptions, etc. Requesters hold the right to deny or increase payment based on work quality. Amazon Mechanical Turk provides an ideal replication testbed, as it tests to see if the results are truly generalizable beyond that of an academic institution.

The use of Mechanical Turk (MTurk) requires some design changes to the original study. We adjusted the manipulation for the participants by removing university-specific information from the fear appeal manipulation. In addition, the MTurk participants are asked to read the manipulation for at least one minute before they proceed to the post-test survey. Attention checks are incorporated to ensure that participants are truly paying attention during the study (e.g., “Select ‘Strongly Agree’ if you think 1 + 1 = 3 is true, otherwise please select ‘Strongly Disagree’”). We employed payment mechanisms to ensure that the participants are adequately compensated for their time in the study. All participants are paid $0.50 for adequately completing the task (i.e., without dropping out or failing the attention checks), which is a fair rate for a 5-8 minutes’ survey. Additionally, we set up an MTurk Task Description wherein potential participants of the study can see the task’s details and how much they will be compensated. Participants are also alerted to their rights, and all major sections of a consent form are presented to the participants. All components of the original and adjusted study are presented in Appendix B.

After making all necessary changes, the Task is launched and kept open for three weeks. Different batches of the Task are randomly released at different points during the day to minimize the chance of tainted sampling pools (Steelman et al., 2014). MTurk workers who completed more than 1,000 approved tasks with an approval rate greater than 95% are qualified to participate. Like the original study, we maintain the same experimental design of randomly assigning participants into one of three groups. Johnston and Warkentin (2010) suggested that the three groups have a minimum of 200, 30, and 30 participants, respectively. Group 1 has a pre-test, manipulation (fear appeal), and a post-test, Group 2 has receives the pre-test and post-test, and Group 3 receives the manipulation and the post-test. We estimate that the participants can complete the pre-test or the post-test survey in 2-3 minutes and finish reading the manipulation message in 1-2 minutes. Participants in Groups 1 and 2 may finish the post-test survey faster since the survey is identical to the pre-test survey that they have completed.

The overall collection is summarized in Table 2. There was a high rate of usable responses across all three groups. The final usable responses for groups 1, 2, and 3 came out to 206, 33, and 37, respectively. Such numbers meet the minimum participant requirement by the original paper (200, 30, 30), and are in line with the original study (215, 30, 30). They are sufficient for the required statistical tests. The survey durations of usable responses are consistent with our estimated completion time for groups 1, 2, and 3.

Table 2. Collection Details

	Group 1	**Group 2**	**Group 3**
Total Responses	284	51	43
Did Not Finish	22	2	5
Failed Attention Questions	56	16	1
Final Usable Responses	206	33	37
Percent Usable	72.54%	64.71%	86.05%
Average Duration (Usable)	06:43	04:36	04:28
Maximum Duration (Usable)	20:11	08:07	15:32
Minimum Duration (Usable)	03:30	01:57	02:05

Delving deeper into the demographics of usable responses reveals that the age range is primarily between 18-39, the majority of the participants are male (154/276), most have a bachelor's degree or higher, and the majority have greater than three years of work experience (213/276). The age and gender distributions are consistent with the original study, while the education and experience numbers are not stated in the original study. The majority of the participants are from the United States (216/276). We summarized our OCM survey procedures in Appendix D as recommended by Steelman et al. (2014).

Table 3. Demographic Information

Age	**Gender**	**Education**	**Experience**	**Country of Origin**
18-29: 101	**Male: 154**	High School: 20	< 6 months: 18	**USA: 216**
30-39: 103	Female: 122	Some College: 80	6-12 months: 9	India: 44
40-49: 43		**Bachelors: 130**	1-2 years: 18	Philippine: 5
50-59: 21		Masters: 36	2-3 years: 18	Others: 10
60+: 8		Doctorate: 9	**> 3 years: 213**	
		Other: 1		

3 Results

All statistical tests and analyses conducted in the original paper are also conducted in this replication paper. They are also presented in the same manner. All statistics are conducted in WarpPLS and/or SPSS.

3.1 Instrument Validity

Johnston and Warkentin (2010) defined social influence (SINF) as a formative construct, with comparisons of performance expectancy (PERF) and attitude (ATTI). We replicated their model selection of PLS. Group 1's (N = 206) post-test response data were used for validity tests. The construct measures PERF, SINF, and ATTI were constituted by multiplying related PLS item weights by item values and summed, following the instruction provided by Johnston and Warkentin (2010). The inter-item and item-to-construct correlation matrix are shown in Table 4, with item-to-construct correlations as grayed out cells. Loch et al. (2003) suggested that if items of the same construct have a significant correlation with their respective composite construct value, then the convergent validity of those formative constructs is adequately shown. As all items are correlated significantly ($p < 0.01$) with their respective construct composite value, this condition was met. This is consistent with the original work.

Table 4. Inter-Item and Item-to-Construct Correlation Matrix

	PERF1	PERF2	PERF3	PERF	SINF1	SINF2	SINF	ATTI1	ATTI2	ATTI3	ATTI4
PERF1	-										
PERF2	0.413	-									
PERF3	0.447	0.767	-								
PERF	0.701	0.897	0.898	-							
SINF1	0.387	0.229	0.222	0.322	-						
SINF2	0.581	0.243	0.254	0.406	0.424	-					
SINF	0.558	0.278	0.278	0.424	0.882	0.8	-				
ATTI1	0.311	0.663	0.601	0.645	0.219	0.207	0.253	-			
ATTI2	0.312	0.555	0.511	0.56	0.093	0.169	0.149	0.709	-		
ATTI3	0.512	0.58	0.491	0.63	0.244	0.266	0.3	0.624	0.659	-	
ATTI4	0.337	0.538	0.458	0.539	0.187	0.21	0.233	0.733	0.815	0.696	-
ATTI	0.412	0.661	0.583	0.67	0.21	0.24	0.264	0.872	0.901	0.835	0.922

PERF = Performance Expectancy; SINF = Social Influence; ATTI = Attitude

Loch et al. (2003) noted that if the item-to-construct correlations are higher with each other than other construct measure's composite values, then discriminant validity can be established. As with the original study, this condition is also met.

Following the procedure set by Johnston and Warkentin (2010), we also conducted construct validity tests. Specifically, we examined whether items loaded onto their intended constructs (Straub et al., 2004). If the item loadings are greater than 0.70 on their factors and the average variance extracted (AVE) for each construct is more than 0.50, then convergent validity is demonstrated (Gefen and Straub, 2005). Additionally, if the root of each construct's AVE is higher than the inter-construct correlations and the item loadings are higher on their appropriate constructs than others, then discriminant validity is adequately proven (Gefen and Straub 2005). Table 5 indicates that all of the aforementioned conditions were met, showing that our study has strong convergent and discriminatory validity. These findings are consistent with the original work.

Table 5. Loadings, Cross-Loadings, and AVEs for Multi-Item Constructs							
	TSEV	**TSUS**	**SEFF**	**RESP**	**BINT**	**Replication AVE**	**Original AVE**
TSEV1	0.938	0.044	0.019	0.002	0.018	0.8464	0.846
TSEV2	0.925	-0.015	0.021	0.001	-0.041		
TSEV3	0.895	-0.030	-0.042	-0.003	0.024		
TSUS1	0.038	0.894	-0.077	0.038	-0.001	0.7832	0.780
TSUS2	-0.048	0.870	0.025	0.019	0.029		
TSUS3	0.009	0.890	0.053	-0.056	-0.027		
SEFF1	0.046	-0.004	0.873	0.086	-0.085	0.7586	0.846
SEFF2	0.016	-0.019	0.911	-0.027	-0.026		
SEFF3	-0.067	0.025	0.827	-0.061	0.119		
RESP1	-0.072	-0.019	0.025	0.929	-0.084	0.7957	0.792
RESP2	0.043	0.071	0.102	0.874	0.078		
RESP3	0.034	-0.051	-0.129	0.871	0.012		
BINT1	-0.005	0.005	0.025	0.011	0.960	0.9216	0.873
BINT2	-0.007	0.002	0.048	-0.029	0.963		
BINT3	0.012	-0.007	-0.073	0.018	0.956		
TSEV = Threat Severity; TSUS = Threat Susceptibility; SEFF = Self-Efficacy; RESP = Response Efficacy; BINT = Behavioral Intent; AVE = Average Variance Extracted							

Scale reliability was assessed by examining the reliability scores in the PLS output. Prior literature indicates that acceptable composite reliability scores are greater than or equal to 0.70 (Fornell and Larker 1981; Gefen and Straub 2005). Table 6 indicates that the composite reliability scores of these reflective variables are acceptable and consistent with the original study.

Table 6. Reliability and Inter-Construct Correlations							
			Inter-Construct Correlations				
Construct	**Original CRel**	**Replication CRel**	**TSEV**	**TSUS**	**SEFF**	**RESP**	**BINT**
TSEV	0.943	0.943	0.920				
TSUS	0.914	0.915	0.269	0.885			
SEFF	0.942	0.904	0.357	0.061	0.871		
RESP	0.897	0.921	0.251	-0.012	0.463	0.892	
BINT	0.954	0.972	0.325	0.290	0.405	0.402	0.960
TSEV = Threat Severity; TSUS = Threat Susceptibility; SEFF = Self-Efficacy; RESP = Response Efficacy; BINT = Behavioral Intent							

3.2 Manipulation Check

Consistent with the original study, we performed a manipulation check to identify whether the changes seen were due to the manipulation administered to the participants or due to the instrument. In our study, the participants were asked about the contents of the manipulation message after the post-test. Table 7 summarizes our replication's manipulation check's against the original. Overall, TSEV, TSUS, and RESP are all significant in both the replication and original. However, there is a difference in SEFF; the replication results indicate that it is not significant. We speculate that this is due to the usage of MTurk participants rather than the ones in the original study.

Table 7. Manipulation Check Results				
	Replication Results		**Original Results**	
IV	**F-test**	**Significance**	**F-test**	**Significance**
TSEV	8.028	$p < 0.05$	6.850	$p < 0.01$
TSUS	3.881	$p < 0.05$	6.174	$p < 0.05$
SEFF	0.072	$p > 0.1$	8.988	$p < 0.01$
RESP	3.661	$p < 0.10$	10.344	$p < 0.01$
TSEV = Threat Severity; TSUS = Threat Susceptibility; SEFF = Self-Efficacy; RESP = Response Efficacy				

3.3 Fear Appeal Manipulation and Test of Internal Validity

A within-subjects MANCOVA of group 1 (N=206, pre-test—manipulation—post-test) assessed the effectiveness in manipulating perceptions of TSEV, TSUS, SEFF, and RESP. Two individual characteristics, experience with anti-spyware software and age, were used as covariates in the analysis. Results can indicate the effectiveness of the fear appeal (Appendix B) in creating changes in end user perceptions of RESP, SEFF, TSEV, and TSUS. These findings are reported in Appendix C, however, indicate that only RESP, TSEV, and TSUS increased significantly ($p < 0.10$), whereas SEFF was not significant ($p > 0.10$), which is inconsistent with the original work. In Johnston and Warkentin (2010), end user's perceptions of RESP, SEFF, TSEV, and TSUS all increased significantly. Our replication results show inconsistencies of user's perception on SEFF with Johnston and Warkentin (2010).

Finally, we tested the differentials in the independent variables based on a MANOVA using Groups 1 and 2 (pre-test—manipulation—post-test and manipulation—post-test, respectively) in Appendix C. The result was also inconsistent with the original work. From our results, the inter-group MANOVA tests on SEFF and RESP were significant ($p < 0.05$), in contrast to the original result that tests on TSEV, TSUS, SEFF, and RESP were all not significant ($p > 0.10$). These results suggest that the pre-test condition may at least partially be a significant factor for the experiment design, and the internal validity of the experiment was not ensured. This may show that the finding that SEFF is insignificant may invalidate the fact that it was opposite to the original.

3.4 PLS Analysis: Test of FAM Nomological Network

A PLS analysis involving post-test Group 1 data (N = 206) was used to test the structural model and its hypotheses. Via bootstrapping resampling, the analysis produced estimates of both the path coefficients as well as the explained variance in RESP, SEFF, and BINT. Of the seven hypotheses, two were found to be significant in the opposite directions to that predicted by hypotheses (H4a and H4b), thus not supported, as shown in the overall findings in Table 8. The remaining hypotheses were supported. This is inconsistent with the original work, where H5a and H5b were both not supported, while the remainder (including H4a and H4b) were supported.

Table 8 and Figure 2 shows that the model explains approximately about 34 percent, 10 percent, and seven percent of the variance in BI, SEFF, and REFF, respectively. The highest explanatory power of 34 percent is the path for social influence, response efficacy, and self-efficacy leading to behavioral intent. Consistent with H1, response efficacy has a significant positive effect on behavioral intent ($\beta = .11$, $p < .10$). Similarly, H2 and H3 are supported as both self-efficacy ($\beta = .33$, $p < .01$) and social influence ($\beta = .31$, $p < .01$) have significant positive effects on behavioral intent, which is consistent with the original work.

Table 8. Overview of Findings						
	Replication Values			**Original Values**		
Hypothesis (with Direction)	**Path Coefficient (β)**	**P-Value**	**Supported?**	**Path Coefficient (β)**	**P-Value**	**Supported?**
H_1: RESP → BINT (+)	0.11	$p < 0.10$	Supported	0.213	$p < 0.01$	Supported
H_2: SEFF → BINT(+)	0.33	$p < 0.01$	Supported	0.187	$p < 0.01$	Supported
H_3: SINF → BINT (+)	0.31	$p < 0.01$	Supported	0.298	$p < 0.001$	Supported
H_{4a}: TSEV → RESP (-)	0.32	$p < 0.01$	Not Supported	-0.286	$p < 0.01$	Supported
H_{4b}: TSEV → SEFF (-)	0.24	$p < 0.01$	Not Supported	-0.437	$p < 0.001$	Supported
H_{5a}: TSUS → RESP (-)	-0.17	$p < 0.01$	Supported	-0.079	$p > 0.10$	Not Supported
H_{5b}: TSUS → SEFF (-)	-0.19	$p < 0.01$	Supported	-0.112	$p > 0.10$	Not Supported
TSEV = Threat Severity; TSUS = Threat Susceptibility; SEFF = Self-Efficacy; RESP = Response Efficacy						

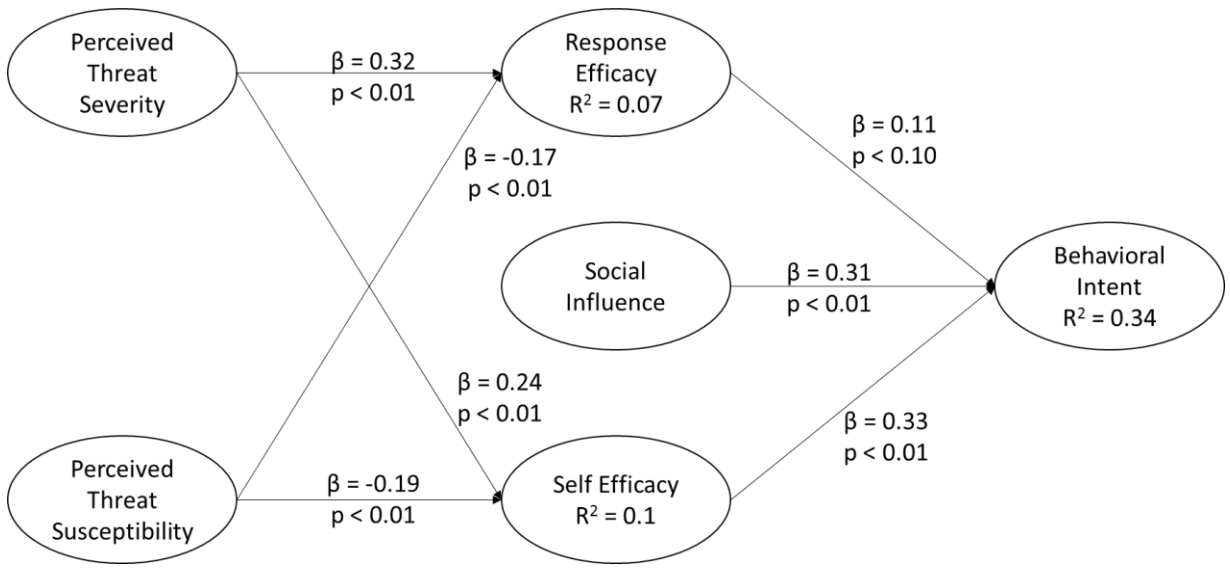


Figure 2. Results of FAM PLS Structural Model Analysis

Structural model analysis results indicate that positive relationships exist between the two threat constructs and the two efficacy constructs. Thus, H4a and H4b are not supported. Threat severity explains approximately 10 percent of the variance in self-efficacy and seven percent of the variance in response efficacy.

4 Discussion

This study methodologically replicated the fear appeals model proposed by Johnston and Warkentin (2010). This model contextualizes the PMT danger control process in the technology adoption literature. In contrast with the original study, this replication used Amazon Mechanical Turk participants instead of student, staff, and faculty participants. Replication results indicate several discrepancies from the original study, each

needing further interpretations and elaborations. We summarize all consistencies and discrepancies in Table 9.

Table 9. Comparison of Original and Replication Studies

Dimension	Original Study	Replication Study	Consistent?
Theoretical Foundations	Protection Motivation Theory (PMT)	Same	Yes
Sample size	G1:215, G2:30, G3:30	G1:206, G2:33, G3:37	Yes
Survey Platform	University faculty, staff, and students	Amazon Mechanical Turk	No
Analysis Tool	Not listed	SPSS, WarpPLS	Unknown
Hypothesis 1	Supported	Supported	Yes
Hypothesis 2	Supported	Supported	Yes
Hypothesis 3	Supported	Supported	Yes
Hypothesis 4a	Supported	Not Supported	No
Hypothesis 4b	Supported	Not Supported	No
Hypothesis 5a	Not Supported	Supported	No
Hypothesis 5b	Not Supported	Supported	No

Several key differences can be identified from Table 9. First, we chose workers on Amazon Mechanical Turk as our participants instead of university faculty and students. This modification provides a better heterogeneity on the sample set, considering wider ranges of ages, educational experiences, anti-spyware software usage experiences, and potentially provides extra external validity on the original work. The fear appeal message is adapted from a college context to a MTurk context (e.g., removal of all academic-related information, email to message) to meet the requirement of general anti-spyware software users.

Second, unlike the results presented by Johnston and Warkentin (2010), our results were not consistent with PMT. We identified that the effects between perceived threat severity and the two efficacy constructs (response efficacy and self-efficacy), however supportive from the initially posited hypotheses, are different from what is discovered in the original PMT model and in Johnston and Warkentin (2010). Opposite to negative relationships, our experiment data suggest that threat constructs positively correlate with efficacy constructs significantly, which indicates a stronger threat severity does not reduce users' response/ self-efficacy, but enhances it. This may be due to the idiosyncrasy of the participant sample set. One possible explanation is that, compared to faculty/student participants, MTurk workers may be more technology-savvy. With this characteristic, the threats may strengthen their confidence of handling situations properly, instead of weakening it. In this sense, a potential boundary condition is identified for the original PMT model, as well as for Johnston and Warkentin (2010). However, additional statistical tests for various participant sets are necessary to truly identify if this is a boundary condition.

Third, statistical tests show that the fear appeal manipulation in our replication work is not consistent with Johnston and Warkentin (2010) on affecting users' self-efficacy. Similarly, the tests between Group 1 and Group 3 suggest that the pre-test condition may at least partially be a significant factor for the experiment design, and the internal validity of the experiment was less ensured. Both of the results are partially inconsistent with the original work. Although a potentially larger sample size may produce a better test result, these results may still lead to further discussions on research design effectiveness and replicability.

4.1 Limitations

As with any study, this paper has limitations. At the time of this writing (2017), spyware threats are not as significant as other forms of malware such as ransomware. However, we argue that the initial motivation of the original study is still valid, as negligent human behavior (e.g., improper anti-spyware software usage) is often the weakest link in any cybersecurity defense.

The usage of MTurk to create our participant pool has its limitations. Research has shown that a non-trivial portion of MTurk workers are from India and/or spoofing to look like they are from the US. Additionally, the tech-savviness of MTurk workers may explain their unique response to a security threat. If this were true,

then the findings of this study would be localized to this population; they may not generalize to other studies, but rather, create a new boundary condition. This may also explain the inconsistency of results in this study compared to the original. Testing on additional populations may provide further explanations.

4.2 Next Steps

There are several promising next steps for this study. First, as the groups in the original study were run in the experiment back to back, we are unsure if these results will be different if they are run with some distracting time in between the tasks. We are particularly interested in the control group, as they received the pre-test and the post-test back to back. As such, future work can run an additional group through Amazon Mechanical Turk. This group would receive the pre-test, spends some time watching unrelated videos (e.g., cat videos), and takes the post-test. All appropriate statistical tests can evaluate the key aspects of this setting. To ensure successful execution of the experiment in the MTurk context and to account properly for those MTurk workers who complete tasks quickly, scholars are advised to add additional attention checks and validations of desired manipulations.

Throughout this replication, we noticed several statistical tests that can further bolster the replication and provide for good scientific grounding. First, we noticed that there was a survey was used in the experiment. However, the original paper did not do any statistical tests for common method bias. Employing Harman's Single Factor method can help deal with any potential issues of common method bias. Secondly, checking for variance inflation factors can help identify constructs are highly correlated with each other. Constructs with high correlation can cause differences in the beta weights. Ideally, the value should be <3.1. Finally, if there is no common method bias and the variance inflation factors prove to be less than <3.1, then we have a better grounding to state that our findings are truly different from the original study.

5 Conclusion

This replication paper methodologically replicated the information security paper of Johnston and Warkentin (2010). The original study conducted a three group, randomly assigned experiment in which participants of the study were members of various academic departments. We applied the same theory, experimental methodology, and statistical tests to Amazon Mechanical Turk users. However, we found results inconsistent with the original study, namely in the consistency of results as they pertained to the relationships between threat and efficacy constructs. As such, additional data collection on multiple populations and statistical tests are required to determine if there were any underlying issues with our analysis.

Acknowledgments

We would like to thank Dr. Sue Brown for her feedback and guidance.

References

Dennis, A. R., & Valacich, J. S. (2014). A replication manifesto. *AIS Transactions on Replication Research*, *1*(1), 1–5.

Gefen, D., & Straub, D. (2005). A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. *Communications of the Association for Information Systems*, *16*, 91–109.

Johnston, A. C., & Warkentin, M. (2010). Fear appeals and information security behaviors: An empirical study. *MIS Quarterly*, *34*(3), 549–566.

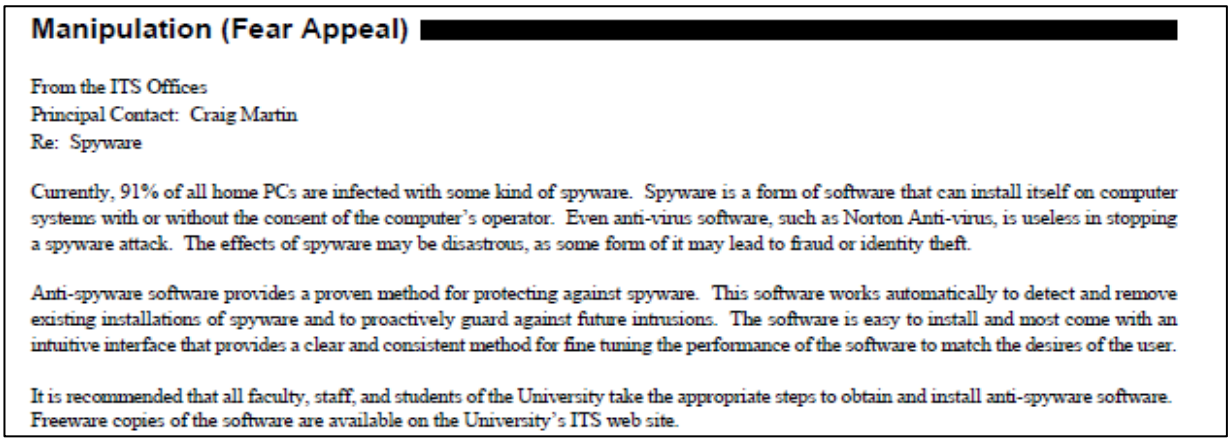
Loch, K. D., Straub, D. W., & Kamel, S. (2008). Diffusing the internet in the Arab world: The role of social norms and technological culturation. In *Global Information Systems: The Implications of Culture for IS Management* (pp. 143–177).

Steelman, Z. R., Hammer, B., & Limayem, M. (2014). Data collection in the digital age: Innovative alternatives to student samples. *MIS Quarterly*, *38*(2), 355–378.

Straub, D., Boudreau, M.-C., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the Association for Information Systems*, *13*(24), 380–427.

Witte, K. (1992). Putting the fear back into fear appeals: The extended parallel process model. *Communication Monographs*, *59*(4), 329–349.

Appendix A: Fear Appeals



Manipulation (Fear Appeal) [redacted]

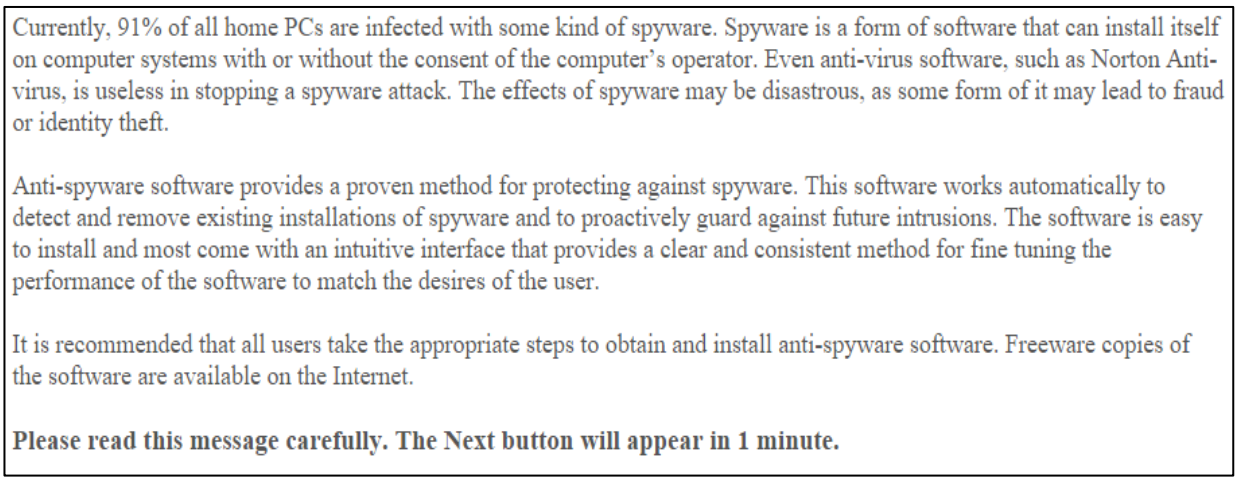
From the ITS Offices
Principal Contact: Craig Martin
Re: Spyware

Currently, 91% of all home PCs are infected with some kind of spyware. Spyware is a form of software that can install itself on computer systems with or without the consent of the computer's operator. Even anti-virus software, such as Norton Anti-virus, is useless in stopping a spyware attack. The effects of spyware may be disastrous, as some form of it may lead to fraud or identity theft.

Anti-spyware software provides a proven method for protecting against spyware. This software works automatically to detect and remove existing installations of spyware and to proactively guard against future intrusions. The software is easy to install and most come with an intuitive interface that provides a clear and consistent method for fine tuning the performance of the software to match the desires of the user.

It is recommended that all faculty, staff, and students of the University take the appropriate steps to obtain and install anti-spyware software. Freeware copies of the software are available on the University's ITS web site.

Figure A1. Original Fear Appeal



Currently, 91% of all home PCs are infected with some kind of spyware. Spyware is a form of software that can install itself on computer systems with or without the consent of the computer's operator. Even anti-virus software, such as Norton Anti-virus, is useless in stopping a spyware attack. The effects of spyware may be disastrous, as some form of it may lead to fraud or identity theft.

Anti-spyware software provides a proven method for protecting against spyware. This software works automatically to detect and remove existing installations of spyware and to proactively guard against future intrusions. The software is easy to install and most come with an intuitive interface that provides a clear and consistent method for fine tuning the performance of the software to match the desires of the user.

It is recommended that all users take the appropriate steps to obtain and install anti-spyware software. Freeware copies of the software are available on the Internet.

Please read this message carefully. The Next button will appear in 1 minute.

Figure A1. Replicated Fear Appeal

Appendix B: Surveys

Instrument (with Notes Added Regarding Formative Versus Reflective Scales)

Section 1: General Purpose

Think about your usage and maintenance responsibilities for a specific computer system. Please select a single score from 1 to 5 where 1 means you **Strongly Disagree** with the statement and 5 means you **Strongly Agree** with the statement.

	Strongly Disagree (1)		Neutral (3)		Strongly Agree (5)
1. I maintain important data on a specific computer.	[ ]	[ ]	[ ]	[ ]	[ ]
2. I am responsible for the detection, prevention, and/or removal of spyware from that computer.	[ ]	[ ]	[ ]	[ ]	[ ]
3. I am concerned for the security of the data on that computer.	[ ]	[ ]	[ ]	[ ]	[ ]

Figure B1. Original Survey

Section 2: Spyware Threat Concerns

The following statements concern spyware and spyware protection. Anti-spyware use refers to installing, running, updating, and/or configuring the software. Please select a single score from 1 to 5 where 1 means you **Strongly Disagree** with the statement and 5 means you **Strongly Agree** with the statement.

	Strongly Disagree (1)		Neutral (3)		Strongly Agree (5)
Threat Severity (reflective)					
1. If my computer were infected by spyware, it would be severe (TSEV1).	[ ]	[ ]	[ ]	[ ]	[ ]
2. If my computer were infected by spyware, it would be serious (TSEV2).	[ ]	[ ]	[ ]	[ ]	[ ]
3. If my computer were infected by spyware, it would be significant (TSEV3).	[ ]	[ ]	[ ]	[ ]	[ ]
Threat Susceptibility (reflective)					
4. My computer is at risk for becoming infected with spyware (TSUS1).	[ ]	[ ]	[ ]	[ ]	[ ]
5. It is likely that my computer will become infected with spyware (TSUS2).	[ ]	[ ]	[ ]	[ ]	[ ]
6. It is possible that my computer will become infected with spyware (TSUS3).	[ ]	[ ]	[ ]	[ ]	[ ]
Self-Efficacy (reflective)					
7. Anti-spyware software is easy to use (SEFF1).	[ ]	[ ]	[ ]	[ ]	[ ]
8. Anti-spyware software is convenient to use (SEFF2).	[ ]	[ ]	[ ]	[ ]	[ ]
9. I am able to use anti-spyware software without much effort (SEFF3).	[ ]	[ ]	[ ]	[ ]	[ ]
Response Efficacy (reflective)					
10. Anti-spyware software works for protection (RESP1).	[ ]	[ ]	[ ]	[ ]	[ ]
11. Anti-spyware software is effective for protection (RESP2).	[ ]	[ ]	[ ]	[ ]	[ ]
12. When using anti-spyware software, a computer is more likely to be protected (RESP3).	[ ]	[ ]	[ ]	[ ]	[ ]
Performance Expectancy (formative)					
13. I would find the use of anti-spyware software useful in my job (PERF1).	[ ]	[ ]	[ ]	[ ]	[ ]
14. Using anti-spyware software enables me to accomplish tasks more quickly (PERF2).	[ ]	[ ]	[ ]	[ ]	[ ]
15. Using anti-spyware software increases my productivity (PERF3).	[ ]	[ ]	[ ]	[ ]	[ ]
Social Influence (formative)					
16. People who influence my behavior think that I should use anti-spyware software (SINF1).	[ ]	[ ]	[ ]	[ ]	[ ]
17. In general, the University has supported using anti-spyware software (SINF2).	[ ]	[ ]	[ ]	[ ]	[ ]
Behavioral Intent (reflective)					
18. I intend to use anti-spyware software in the next 3 months (BINT1).	[ ]	[ ]	[ ]	[ ]	[ ]
19. I predict I will use anti-spyware software in the next 3 months (BINT2).	[ ]	[ ]	[ ]	[ ]	[ ]
20. I plan to use anti-spyware software in the next 3 months (BINT3).	[ ]	[ ]	[ ]	[ ]	[ ]
Attitude (formative)					
21. Anti-spyware software makes work more interesting (ATTI1).	[ ]	[ ]	[ ]	[ ]	[ ]
22. Working with anti-spyware software is fun (ATTI2).	[ ]	[ ]	[ ]	[ ]	[ ]
23. I like working with anti-spyware software (ATTI3).	[ ]	[ ]	[ ]	[ ]	[ ]
24. Working with anti-spyware software is enjoyable (ATTI4).	[ ]	[ ]	[ ]	[ ]	[ ]

Figure B2. Original Survey

Section 3: Demographic Information

The demographic information in this section will only be used in aggregate form and will not be used to identify individual respondents. Please select only one item in each category. Experience refers to your experience using anti-spyware software. Department refers to the department in which you are employed or enrolled.

Gender [ ] male
[ ] female

Experience [ ] < 6 months
[ ] 6–12 months
[ ] > 1 year to 2 years
[ ] > 2 years to 3 years
[ ] > 3 years

Age [ ] 18 to 29
[ ] 30 to 39
[ ] 40 to 49
[ ] 50 to 59
[ ] 60 and over

Education [ ] high school
[ ] some college
[ ] bachelor's degree
[ ] master's degree
[ ] doctorate
[ ] other

Department [ ] COBI
[ ] CVM
[ ] ITS
[ ] AOCE
[ ] other

Thank you for participating in this study.

Figure B3. Original Survey

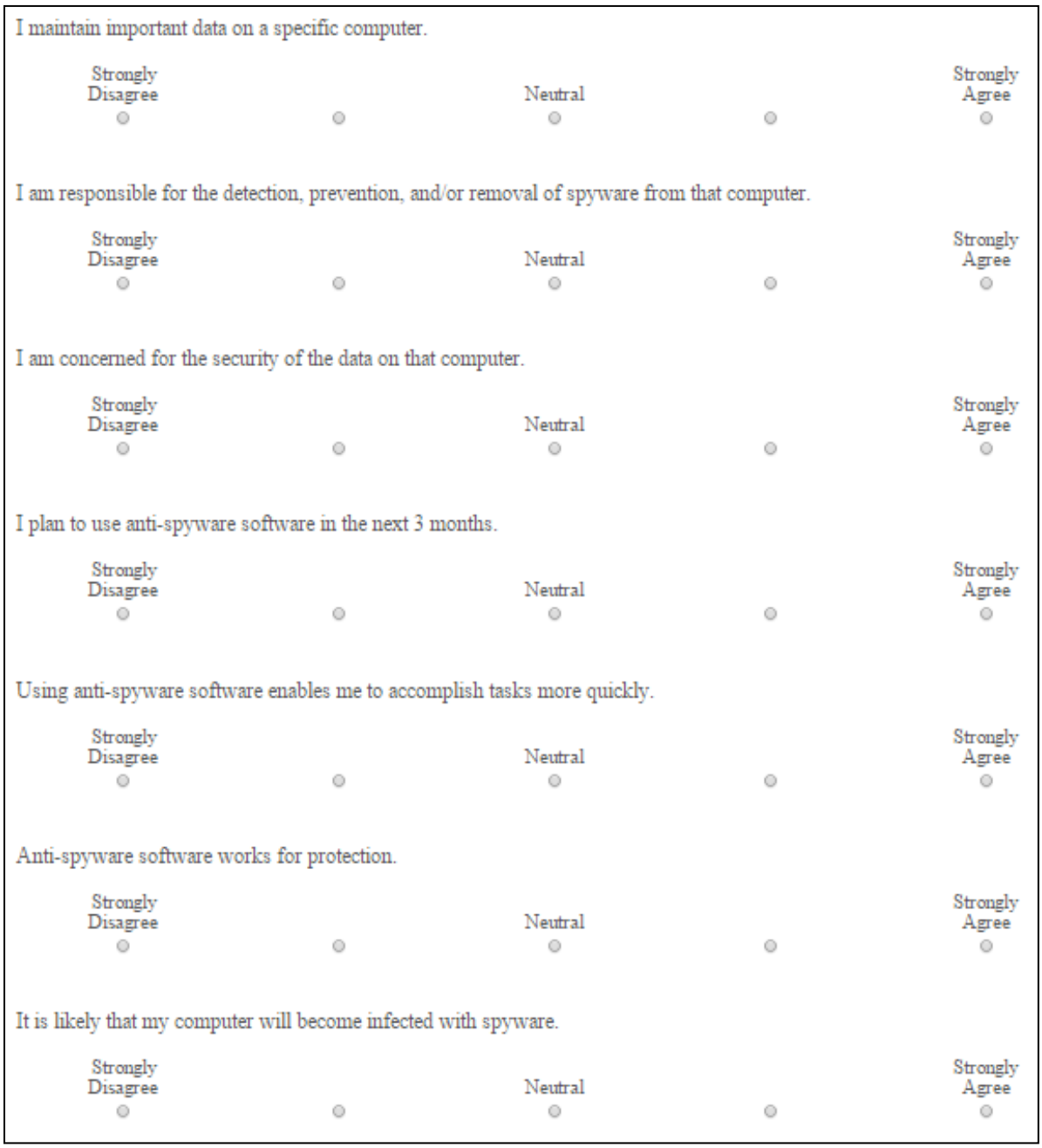


Figure B4. Replicated Survey

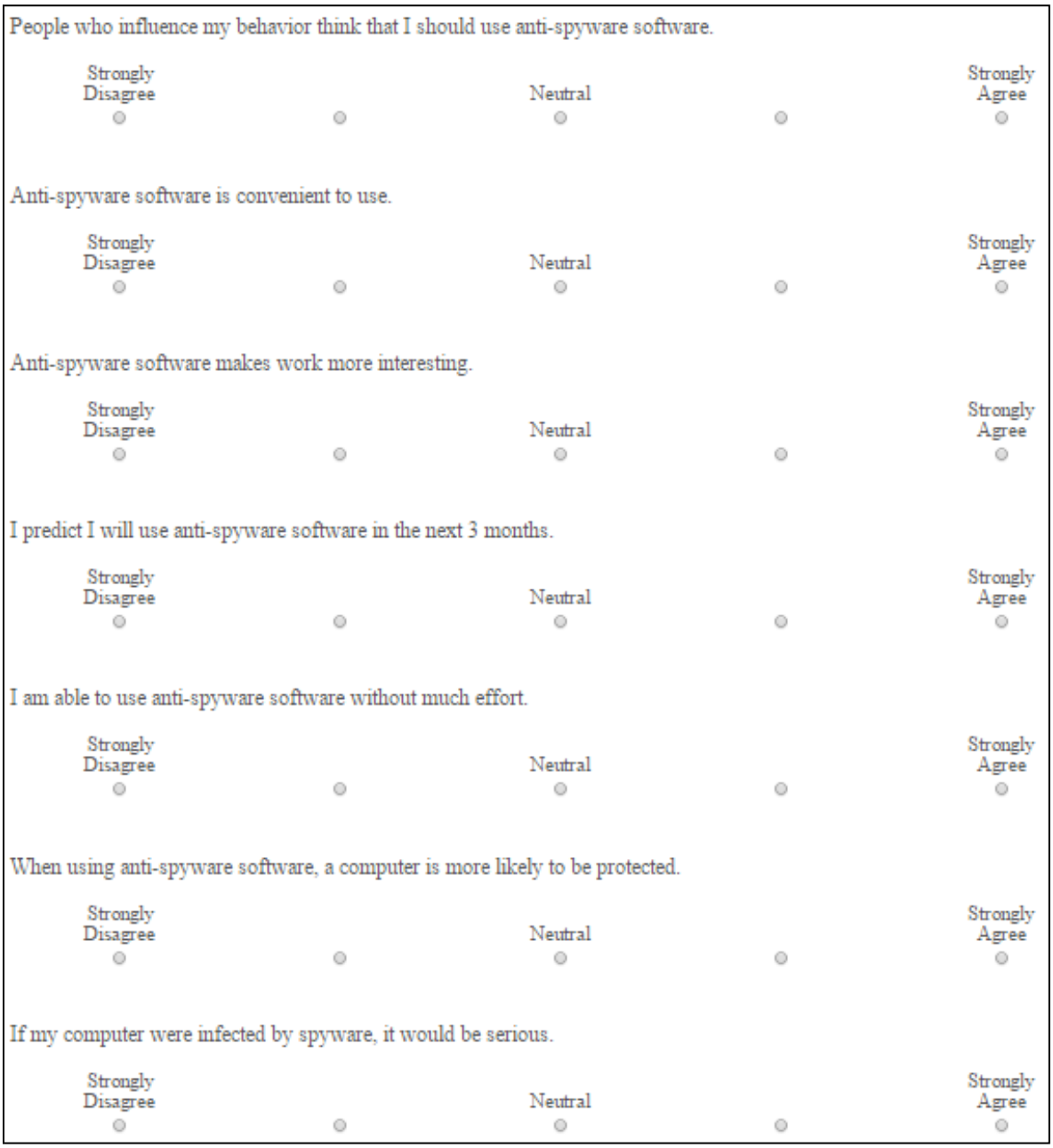


Figure B5. Replicated Survey

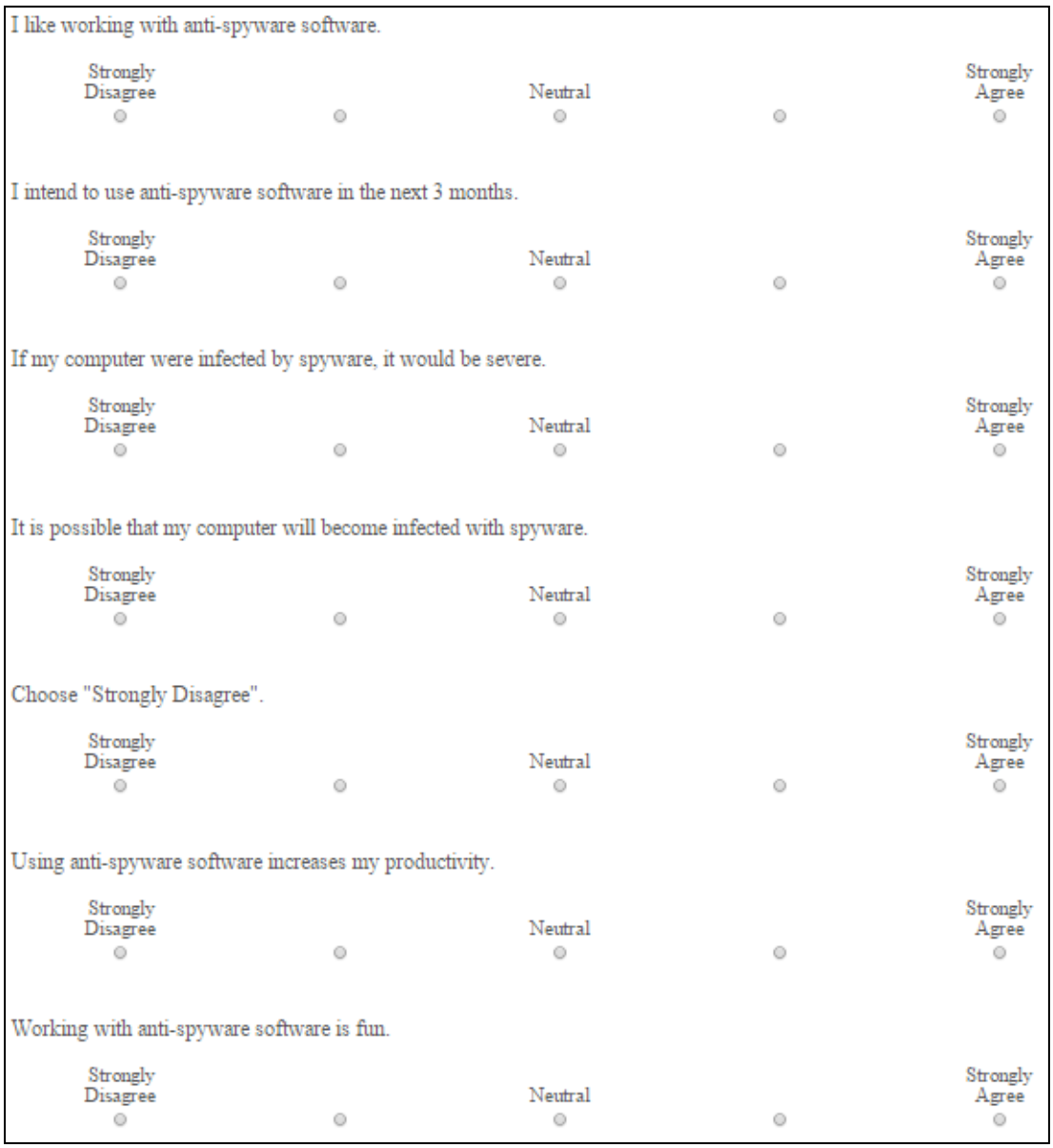
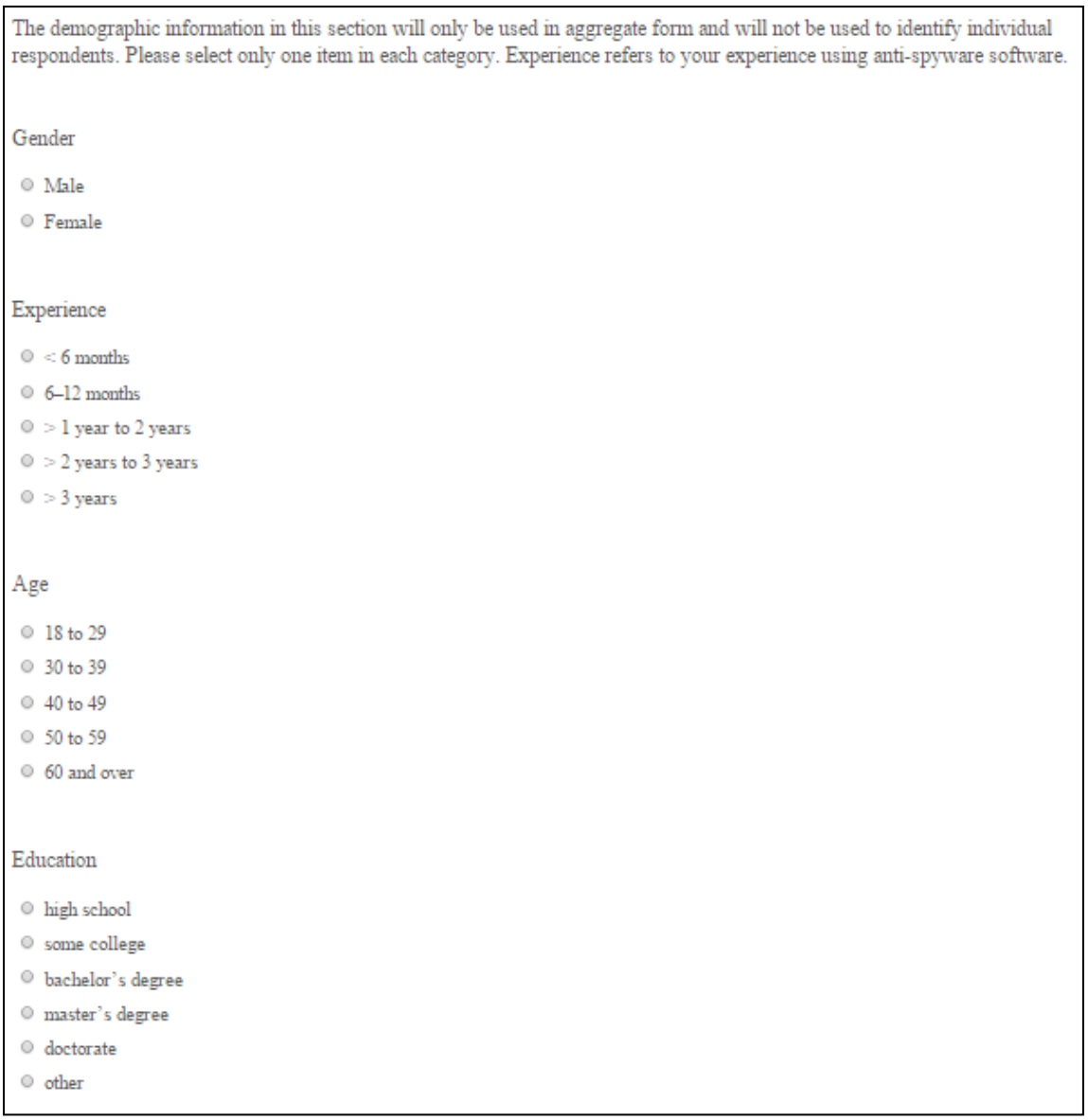


Figure B6. Replicated Survey

Figure B7. Replicated Survey


The demographic information in this section will only be used in aggregate form and will not be used to identify individual respondents. Please select only one item in each category. Experience refers to your experience using anti-spyware software.

Gender

- Male
- Female

Experience

- < 6 months
- 6–12 months
- > 1 year to 2 years
- > 2 years to 3 years
- > 3 years

Age

- 18 to 29
- 30 to 39
- 40 to 49
- 50 to 59
- 60 and over

Education

- high school
- some college
- bachelor's degree
- master's degree
- doctorate
- other

Figure B8. Replicated Survey

Appendix C: Additional Results

Table C1. Results of Fear Appeal Manipulation Effectiveness				
Construct	**Pre-test Mean**	**Post-test Mean**	**F-Test**	**Significance**
TSEV	3.922	4.042	8.741	0.03
TSUS	3.212	3.574	52.098	<0.001
SEFF	3.888	3.896	0.045	0.833
RESP	4.218	4.252	0.791	0.375
TSEV = Threat Severity; TSUS = Threat Susceptibility; SEFF = Self-Efficacy; RESP = Response Efficacy				

Table C2. Test of Internal Validity		
Construct	**Inter-Group F-Test**	**Significance**
TSEV	1.010	0.316
TSUS	0.683	0.409
SEFF	1.736	0.189
RESP	0.025	0.876
TSEV = Threat Severity; TSUS = Threat Susceptibility; SEFF = Self-Efficacy; RESP = Response Efficacy		

Appendix D: OCM Reporting

Table D. OCM Reporting Summary	
Reporting Items	**Summary**
1. Participant demographics	
a. Country of origin distribution	Reported in Table 3
b. Income	Not in the original study
c. Age	Reported in Table 3
d. Gender	Reported in Table 3
e. Marital status	Not in the original study
f. Employment status	We did not have this information.
2. Participation restrictions	
a. Location or country of origin	We did not restrict on the location.
b. Computer system requirements	We did not restrict on the computer system.
c. Survey experience	Reported. Mechanical Turk Workers who completed more than 1000 approved tasks with an approve rate higher than 95% are qualified to participate.
3. Payment incentives	
a. Monetary specifications	Reported. All participants are compensated $0.50 for adequately completing the task (i.e., without dropping out or failing the attention checks).
b. Bonus incentive conditions	No other bonuses.
4. Task timeline	
a. Average time to completion	Reported in Table 2
b. Minimum and maximum time to completion	Reported in Table 2
5. Data quality questions and checks	
a. Human verification tasks	Based on 2.c, our participants are not likely to be robots.
b. Attention verification tasks	Reported. Attention check examples: • Choose "Strongly Disagree". (Figure A6) • Select the choice between "Strongly Agree" and "Neutral. (Figure A7)
c. Embedded software techniques (e.g., CAPTCHA)	Not applicable to this research
d. Description of correct vs. failed responses	Reported in Table 2. Incomplete responses and responses failed on attention checks are failed responses.
6. Detailed data cleaning procedures	
a. Number of responses received before cleaning	Reported in Table 2
b. Checks for compliance with participant restrictions (e.g., country of origin validation)	We did not restrict on 2.a and 2.b. Amazon Mechanical Turk system automatically filtered participants based on 2.c.
c. Description of protection from previous survey responses	Reported. Different batches of the Task are released randomly at different times of day.

About the Authors

Sagar Samtani. Sagar Samtani is an Assistant Professor at the University of South Florida. His research interests are in developing proactive Cyber Threat Intelligence (CTI) by using and developing deep learning, text mining, and network science analytic procedures. His work has been published in *Journal of Management Information Systems* and *IEEE Intelligent Systems*.

Hongyi Zhu. Hongyi Zhu is currently a PhD student majoring in MIS at the University of Arizona. His major research interest lies in deep learning-based mobile health analytics, which aim to model, recognize, and analyze senior citizen's home-based activities with unobtrusive sensing techniques such as accelerometers and passive environment sensors. He is also interested in machine learning, data analytics, and visualization. His work has been published in *Journal of Nanoparticles Research*.

Shuo Yu. Shuo Yu is currently a PhD student majoring in MIS at the University of Arizona. His main research interest lies in mobile health analytics using deep learning techniques, which aim to detect adverse physical events and predict condition risks for senior citizens based on acceleration data collected from miniature motion sensors. He also has an interest in general data analytics, text mining, and natural language processing. His work has been published in *Institute of Electrical and Electronics Engineers Journal of Biomedical and Health Informatics*.