Persuading end users to act cautiously online: A fear appeals study on phishing

Jurjen Jansen¹,² and Paul van Schaik³

¹Faculty of Humanities and Law, Open University of the Netherlands
²Cybersafety Research Group, NHL Stenden University of Applied Sciences & Dutch Police Academy
³School of Social Sciences, Humanities and Law, Teesside University
e-mail: j.jansen@nhl.nl; p.van-schaik@tees.ac.uk

Abstract

Purpose – The purpose of this study is to test the protection motivation theory in the context of fear appeal interventions to reduce the threat of phishing attacks. In addition, it was tested to what extent the model relations are equivalent across fear appeal conditions and across time.

Design/methodology/approach – A pre-test post-test design was used. In the pre-test, 1,201 Internet users filled out an online survey and were presented with one of three fear appeal conditions: strong fear appeal, weak fear appeal and control condition. Arguments regarding vulnerability of phishing attacks and response efficacy of vigilant online information-sharing behaviour were manipulated in the fear appeals. In the post-test, data were collected from 786 Internet users and analysed with partial least squares path modelling.

Findings – The study found that PMT model relations hold in the domain of phishing. Self-efficacy and fear were the most important predictors of protection motivation. In general, the model results were equivalent across conditions and across time.

Practical implications – It is important to consider online information-sharing behaviour because it facilitates the occurrence and success of phishing attacks. The results give practitioners more insight into important factors to address in the design of preventative measures to reduce the success of phishing attacks. Future research is needed to test how fear appeals work in real-world settings and over longer periods of time.

Originality/value – This paper is a substantial adaptation of a previous conference paper (Jansen and Van Schaik, 2017).

Keywords Information security behaviour, Fear appeals, Protection motivation theory, Phishing, Online information-sharing behaviour, Multi-group analysis, Human aspects

Paper type Research paper
1. Introduction

As more services are offered online and personal data are increasingly stored by digital means, people become more technology-dependent, but also more susceptible to security incidents (Furnell et al. 2007). Nonetheless, people play an important role in protecting themselves against such incidents, because they form a crucial link in the information security chain.

This study focusses on the protection against a specific online threat, namely phishing: the process of retrieving personal information using deception through impersonation (Lastdrager, 2014). Phishing is considered predominantly dangerous to Internet users (Arachchilage et al. 2016) and forms a world-wide problem (APWG, 2015) for different sectors, such as retail and banking.

Security education, training and awareness, and the implementation – and proper application – of precautionary online behaviour are critical in the protection against phishing attacks (Purkait, 2012). Although these efforts will not solve the phishing problem on its own (Alsharnouby et al. 2015), aware and vigilant Internet users who practice precautionary online behaviour are believed to better identify phishing attempts (Purkait, 2012). However, transforming the Internet population into an aware and vigilant audience is not easy, as it is not precisely known which interventions work best.

This study contributes to improving online security by investigating to what extent fear appeals can persuade Internet users to perform safe online behaviour and using protection motivation theory (Maddux and Rogers, 1983; Rogers, 1975) as its theoretical basis. Fear appeals are ‘informative communication[s] about a threat to an individual’s well-being’ (Milne et al. 2000, p. 107) that also contain information to promote perceptions of efficacy. Attention to fear and fear appeals is currently lacking in the information security domain (Johnston et al. 2015), but gains in popularity (Wall and Buche, 2017).

We focus on one type of behavioural context: sharing or disclosing personal information online. Personal information includes personally identifying, financial and demographic information (Norberg et al. 2007). When people put their personal information online, it makes it easy for perpetrators to, for example, (spear) phish someone (Shillair et al. 2015). Rocha Flores et al. (2014) who conducted an experiment in an organizational setting demonstrated that when more target information was added to an attack, the likelihood of falling for that attack increased. Furthermore, phishing research has shown that for a scam to be effective, it is crucial that people provide their personal information, for instance, user credentials in response to the scam (e.g., Jansen and Leukfeldt, 2015). Thus, acting cautiously
regarding personal information-sharing online is essential to (a) protect oneself from being attacked by means of phishing and (b) prevent phishing attacks from succeeding.

This paper highlights the results of a pre-test post-test design using fear appeal manipulations. The main goal is to gain insight into the effects of fear appeal manipulations on Internet users’ protection motivation. We do this by developing a research model and then testing this model. In addition, we test the model on different subgroups based on fear appeal manipulations, which is recommended to further differentiate the findings (Hair et al. 2014). In addition, we test whether there are differences in model results between pre-test and post-test data. In sum, our study addresses the following research questions.¹

Research Question 1: to what extent do end users share their personal information online?
Research Question 2: to what extent do the PMT model relations hold in the domain of phishing?
Research Question 3: to what extent are the model relations equivalent across fear appeal conditions?
Research Question 4: to what extent are the model relations equivalent across time?

2. Background literature

The purpose of protection motivation theory (henceforth PMT) is to clarify fear appeals, but it has also been used as a more general model to study decision-making under risk (Maddux and Rogers, 1983). PMT has been recently used in the information security domain and is considered to be a useful theory for predicting different types of precautionary behaviour (e.g., Jansen and Van Schaik, 2017).

PMT posits that intentions to perform precautionary behaviour (protection motivation) are initiated by the threat appraisal process: an evaluation of the perceived vulnerability and severity of a possible threat that is triggered by a fear appeal, in this case in relation to phishing. This is followed by the coping appraisal process, in which a particular response to mitigate or minimize the threat is evaluated, based on the perceived efficacy of this response, the perceived self-efficacy of executing or adopting the response and the costs that are associated with performing the coping response, in this case being vigilant towards sharing or disclosing personal information online.

¹ The effects of fear appeal messages on PMT variables on T1 data are presented in Jansen and Van Schaik’s (2017) work and on T2 data is the subjects of one of the authors’ publication (currently under review).
Based on the notions above, the research model is presented in Figure 1. According to the model, protection motivation is a positive function of perceived vulnerability, perceived severity, fear, response efficacy and self-efficacy, and a negative function of response costs. Similarly, fear is a positive function of perceived vulnerability and perceived severity.

Prior research shows that response efficacy and self-efficacy are the most influential predictors for precautionary online behaviour (e.g., Boehmer et al. 2015; Jansen and Van Schaik, 2017). This is also true in the health domain; in particular, the meta-analyses of PMT research by Floyd et al. (2000) and Milne et al. (2000) and the meta-analysis of fear appeal research by Witte and Allen (2000) demonstrate that the coping variables generally show stronger relations with adaptive behaviours than the threat variables do. However, besides increasing the perceived efficacy of a recommended response, raising perceived threat in a fear appeal is still important because threat appraisal initiates coping appraisal. Finally, Witte and Allen (2000) stress that fear appeals will only work when complemented by an equally strong efficacy message. Therefore, our fear appeals include efficacy information, which is elaborated in the next section.

3. Method
Here we describe the methods used to answer the research questions. First, we discuss the survey questionnaire, procedure and participants. Second, we discuss the design of the fear appeals. Third, we discuss data analysis, validity and reliability of our measures.

3.1 Survey questionnaire and procedure
This study used a pre-test post-test design. Thus, data were collected at two moments in time. The first measurement was conducted between February 28 and March 13 2017 (T1) and the second measurement between April 4 and April 21 2017 (T2).

A survey design was used to experimentally manipulate fear appeals. Sampling was done by an external recruitment service of online survey panels. Participants received panel points that can be used for discounts at Web shops and for donations to charities as compensation for their voluntary participation. The participants were randomly assigned to one of three experimental groups: an intervention using strong arguments (strong fear appeal), an intervention using weak arguments (weak fear appeal) and no intervention (control
condition). Stratified sampling was applied for group composition – controlling for gender and age – resulting in equivalent groups as demonstrated by the results from subsequent analysis of variance (ANOVA). Manipulations of argument strength are common in fear appeal intervention studies (Milne et al. 2000) and are expected to have an effective impact on message processing (Petty & Cacioppo, 1986). In information security studies, however, it is common to test one (or more) treatment(s) against no treatment (Johnston et al. 2015). We combine these two approaches.

The survey at T1 first requested from participants their demographic characteristics, to allow us to balance the quotas of the strata across the conditions. Next, participants answered questions regarding their Internet experience and online personal-information-sharing behaviour. This was followed by the fear appeal manipulation – a written text within the survey environment. One group read a strong fear appeal message, one group read a weak fear appeal message and a control group received no message.

Immediately after the message, participants filled out questions – on a 5-point Likert scale (1 [totally disagree] – 5 [totally agree]) – representing PMT’s core variables. The questionnaire items were based on the works of Anderson and Agarwal (2010), Ifinedo (2012), Johnston et al. (2015), Milne et al. (2002), Ng et al. (2009) and Witte (1996). They were translated in Dutch and were presented in random order. The protection motivation questions and questions on previous online information sharing behaviour included a time frame of four weeks, since time is an important element of behaviour (Fishbein & Ajzen, 2010) and because the post-test (T2) took place four weeks after the pre-test (T1).

At T2, participants filled out a survey with the same PMT-items and their information-sharing behaviour in the past month. This was done to check whether the results are equivalent over time and to study whether participants’ intentions stated at T1 had been acted upon over the last four weeks. The order of the items was changed at T2 to counter possible memory effects.

3.2 Survey participants
Data collection at T1 was conducted with 1,219 Dutch Internet users. After excluding 18 participants because of unsuitable responses, the net response was 1,201. The study at T2 was filled out by 880 Dutch Internet users. However, 94 participants were excluded, resulting in a

---

2 The results on subsequent behaviour are presented and discussed in one of the authors’ other publications (currently under review).
net response of 786 and a net retention rate of 65%. The data analyses are based on the 786 participants who filled out the questionnaire on both occasions.

Of the 786 participants, 48.6% women and 51.4% men participated. The mean age of participants was 49.5 years ($SD = 15.8$) and the age range was 19-76 years. Their levels of education were low (14.0%), medium (33.5%) and high (52.5%). No significant differences were found for the demographic variables between the three measurement groups, at either T1 or T2.

3.3 Fear appeal design
Like most PMT studies, our study involved manipulating a written communication, targeting PMT-variables. Both the strong and weak fear appeal message included factual information on the vulnerability and severity of phishing attacks, appealing to threat appraisal. The combination of manipulated threat appraisal and coping appraisal variables showed the largest effect on outcomes in earlier studies (Sheeran et al. 2014). Therefore, our messages also contained information on how to mitigate phishing attacks by means of being vigilant when sharing personal information online (the suggested coping response). For coping appraisal, specific information was included appealing to response efficacy and self-efficacy.

In the strong fear appeal message, strong arguments were presented regarding perceived vulnerability, whereas the weak fear appeal message used weak arguments nuancing the chances to be victimized by a phishing attack. For coping appraisal, the primary focus was on arguments regarding response efficacy, because this variable showed strongest predictive ability in previous research. The strong fear appeal used strong arguments framing the response as being very effective, that is not sharing personal information online will lead to not being attacked by phishing and any phishing attack that may happen not being successful. In contrast, the weak fear appeal used weak arguments downgrading the level of efficacy. The manipulations are based on results from the work of Bursztein et al. (2014) and Kloosterman (2015).

3.4 Data analysis, validity and reliability
Partial least squares path modelling (henceforth PLS), using SmartPLS 2.0 (Ringle et al. 2005), was used to test the research model and to emphasize differences in model results between the three conditions and the two time points. After the removal of one protection

---

3 The pilot testing of fear appeal messages is the subject of one of the authors’ publication (currently under review). Detailed information about the fear appeals and measures are available from the authors upon request.
motivation item (PM3) with a high cross-loading, the measurement model with T1 data had a simple factor structure, and good composite reliability, convergent validity and discriminant validity. Subsequent analysis using SPSS showed no signs for multi-collinearity issues. The same goes for the measurement model with T2 data, except that in this case another protection motivation item (PM4) and a response cost item (RC2) needed to be removed. Detailed information about the measurement models are available from the authors upon request.

We used a standard bootstrapping procedure (N = 5,000) to test the significance of the structural models’ parameters (Henseler et al. 2009). After testing the overall structural models for each experiment condition, we compared differences between the conditions using t-tests (two-tailed). These were calculated in Excel using Hair et al.’s formula\(^4\), because SmartPLS 2.0 does not execute PLS multi-group analysis (Hair et al. 2014).

4. Results

We first analyse the participants’ Internet and online information sharing behaviour. Next, the general PLS results are discussed. This is followed by the results from the multi-group analysis.

4.1 Internet and online information-sharing behaviour

Participants made use of the following six online services (T1, N = 1,201): e-mail (99.8%); online banking (96.7%); buying products on online marketplaces and/or Web shops (93.7%); instant messaging (e.g., WhatsApp and Facebook Messenger) (87.1%); social media (e.g., Facebook, Instagram and LinkedIn) (84.4%); and selling products on online marketplaces and/or Web shops (57.8%).

According to Table 1, most participants indicated to have shared their e-mail address (82.4%) and home address (58.5%) online in the previous year. This is followed by having shared their bank account number (38.8%) and citizen service number (18.5%). Log-in credentials (4.5%) and PIN codes and/or security codes (1.7%) were shared the least in the previous year.\(^5\) With the exception of citizen service number, over fifty per cent of those who had shared their personal information online in the previous year, also indicated to have done so in the previous month. In total, 180 participants (15.5%) indicated that they had not shared

---

\(^4\) Retrieved from https://www.pls-sem.net/downloads/1st-edition-a-primer-on-pls-sem/

\(^5\) Note that the percentages may differ when adding up those presented in the Table 1 because of rounding mechanisms.
any of the requested information online in the previous year, rising to 411 (34.2%) in the previous month.

According to Table 2, participants shared their personal information online in different ways. E-mail, Web shops, websites and instant messaging were mentioned most frequently. In terms of active information-sharing, participants used both familiar and unfamiliar locations. Familiar locations were more used often in applications for communications, whereas unfamiliar locations were more often used concerning Web shops and websites.

4.2 PMT model
The structural models using all the data of T1 and T2 – using PLS – are presented in Table 3. A substantial amount of variance in protection motivation was explained at both measurements ($R^2 \approx 60\%$). The amount of variance explained in fear was considerable, with 35% at T1 and 47% at T2.

Most of the coping variables are significant and in the expected direction predicting protection motivation, with the exception of response costs at T1. Threat variables were of less direct influence on protection motivation, with the exception of perceived severity at T1. However, fear – together with self-efficacy – was a strong predictor of protection motivation. In addition, perceived vulnerability and perceived severity were significant predictors of fear in both T1 and T2.

4.3 Multi-group analysis
The structural models of the three conditions are presented in Table 4. A substantial amount of variance in protection motivation was explained in all three conditions ($R^2 \approx 60\%$). The

Note that these results are calculated using the data of participants that participated both at T1 and T2 ($N = 786$). The PLS results of all participants who participated at T1 ($N = 1,201$) are presented in the work of Jansen and Van Schaik (2017).
amount of variance explained in fear was considerable, with 42% for the weak fear appeal condition and quite similar for the strong fear appeal and control conditions ($R^2 \approx 35\%$).

Insert Table 4 about here.

First, the results show that self-efficacy and fear were significant predictors of protection motivation in all conditions, with self-efficacy being the strongest. Second, perceived vulnerability and response cost were not a significant predictor of protection motivation in any of the three conditions. Third, response efficacy was a significant predictor of protection motivation in the weak fear appeal condition and perceived severity was significant in the control condition. Fourth, perceived vulnerability and perceived severity were significant predictors of fear in all conditions.

The equality of model parameters between different groups is tested with Hair et al.’s formula, specifically for strong fear appeal, weak fear appeal and control condition (see Table 5). Only one significant difference arose, i.e., the relation between perceived vulnerability and fear was stronger in the weak fear appeal condition compared with the strong fear appeal condition.

Insert Table 5 about here.

We were also interested in whether the model relations are equivalent across time. Therefore, we now present the structural models of the three conditions as measured at T2 using the same procedure as for T1, see Table 6. The amount of variance explained in protection motivation was again substantial, with 52% for the strong fear appeal condition and quite similar for the weak fear appeal and control conditions ($R^2 \approx 60\%$). A considerable amount of variance in fear was explained once more in all three conditions ($R^2 \approx 47\%$).

Insert Table 6 about here.

The results from T2 are quite similar to those from T1. First, the results show that self-efficacy and fear were significant predictors of protection motivation in all conditions, with self-efficacy being the strongest. Second, perceived vulnerability, perceived severity and response cost were not a significant predictor of protection motivation in any of the three
conditions. Third, response efficacy was a significant predictor of protection motivation in the control condition. Fourth, perceived vulnerability and perceived severity were significant predictors of fear in all conditions.

The equality of model parameters between the different measures is again tested with Hair et al.’s formula (T1 versus T2), see Table 7. Two significant differences were observed in the strong fear appeal condition. The effect of perceived vulnerability on fear was stronger at T2 than at T1, which also goes for the effect of fear on protection motivation. One marginally significant difference was found in the control condition. The effect of fear on protection motivation was stronger at T2 than at T1.

_________ Insert Table 7 about here. __________

5. Limitations
Because our study took place within participants’ social context, we created a realistic setting in which Internet users read the fear appeal and answered questions about their cognitions and behavioural intentions. However, this means that we could not control for the effect of other messages related to safe online practices which were not part of intervention, but which participants may have encountered in their day-to-day use of the Internet. To rule out potential threats to internal validity but also external validity, future studies could adopt more variants of fear appeal manipulations.

Another limitation can be found in the number of participants who responded at T2. Although we managed to obtain a net retention rate of 65%, we, consequently, missed the responses of 35% at T2. This might have affected the outcomes to some extent. Furthermore, it is unclear how the fear appeals would have worked in real-world situations (Wall and Buche, 2017). In addition, future studies could also take into account messenger and message factors, that might influence the persuasive effect of fear appeals.

Both a strength and a weakness of the current study is that it focussed on one type of behaviour, as precautionary online behaviour (against phishing) consists of a range of behaviours (Crossler et al. 2017). The strength is related to the fact that predictors of one type of behaviour might not influence another type of behaviour (Blythe et al. 2015). Therefore, we now have a better understanding of what motivates end users to perform specific individual behaviour. A weakness is that it does not represent precautionary online behaviour as a whole. Rather it studies a type of precautionary behaviour in isolation, possibly hindering the theoretical development of the overall structure of preventing phishing (Posey et al.)
2015). We did, for example, not focus on recognizing phishing e-mails or phishing websites. Indeed, phishing is a problem that cannot be solved by a single solution at one level (Purkait, 2012). On the other hand, recent research continues to demonstrate that identifying phishing attempts is an extremely difficult task for Internet users (Alsharnouby et al. 2015). An important point of discussion for behavioural-information security researchers is whether research on precautionary online behaviour should focus on a single behaviour or on multiple behaviours. Crossler et al. (2017) strongly advise to focus on multiple behaviours concurrently; however, this may make the research designs exceedingly demanding for research participants.

6. Conclusions and discussion

In response to Research Question 1 (To what extent do end users share their personal information online?), our results from T1 show that the participants often shared their personal information online. This primarily goes for address details, which is unsurprising since people need to find each other this way, especially on the Internet. More sensitive data are shared to a lesser extent, that is bank account numbers and citizen service numbers. Again, these are often necessary for instance to buy products or to make use of governmental services. However, it also became clear that respondents share log-in credentials (4.4%) and PIN codes or security codes (1.8%). These results indicate that a number of participants engage in potentially harmful online behaviour. This is also true for participants who share their personal information with unfamiliar (and potentially untrustworthy) sources, which commonly occurred.

In response to Research Question 2 (To what extent do the PMT model relations hold in the domain of phishing?), from the overall PLS analyses we observed that self-efficacy was the most important predictor for protection motivation, which is a common finding in PMT-studies. Thus, the more a person believes he or she is able of carrying out the measure, the more likely protection motivation is (e.g., Ifinedo, 2012; Jansen and Van Schaik, 2016). Response efficacy, also often found to be a significant important predictor, impacted protection motivation to a lesser extent. Hence, a significant effect of response efficacy on protection motivation was only visible in the weak fear appeal condition; although the multi-group analysis did not indicate that there were significant differences between the three conditions for this predictor. Response cost was not a significant predictor of protection motivation at T1. Thus, it seems that the participants’ protection motivation is not influenced
perceived costs over and above the influence of other PMT variables. This can also be explained by the fact that response costs were not addressed in the fear appeals.

Next to self-efficacy, fear had the second greatest effect on protection motivation. However, the threat variables had almost no influence on protection motivation, except for perceived severity at T1. This could possibly be explained by threat variables having both a positive and negative relation with protection motivation, depending on the responses already taken against a particular threat (Milne et al. 2000). Nevertheless, besides increasing the perceived efficacy of a recommended response, raising perceived threat in a fear appeal is still important because threat appraisal initiates coping appraisal (Floyd et al. 2000) and because it appeals to personal relevance, which is important for communications on information security (Johnston et al. 2015).

In response to Research Question 3 (To what extent are the model relations equivalent across fear appeal conditions?), we found one significant difference between the conditions. In particular, the significant relation between perceived vulnerability and fear was stronger in the weak fear appeal condition compared with the strong fear appeal condition. Therefore, generally the model relations were equivalent across conditions.

In response to Research Question 4 (To what extent are the model relations equivalent across time?), we observed three significant differences. Two of these occurred in the strong fear appeal condition. The significant effects of perceived vulnerability on fear and of fear on protection motivation were stronger at T2 than at T1. The other difference was observed in the control condition, where the significant effect of perceived vulnerability on fear was stronger at T2 than at T1. Therefore, the model relations were mostly equivalent and otherwise increased in strength.

In sum, we established that potentially unsafe information-sharing behaviours occur to varying degrees. It is important to consider addressing these behaviours because they facilitate the occurrence and success of phishing attacks. We also established, now in the domain of phishing, the power of specific PMT-variables to predict protection motivation (in particular, self-efficacy and fear), that were previously found to be influential in information security studies as well is in other domains. It is important to consider these predictors in designing security education, training and awareness campaigns to reduce the number of successful of phishing attacks.

In conclusion, future studies should take into consideration the longitudinal aspect regarding the effect of interventions. We found that the model relations were mostly equivalent over time and otherwise increased in strength in a month’s time. However,
research indicates that such effects may wane over time (e.g., Bullée, 2017). It is important for behavioural information security researchers to examine how such effects evolve over time, for instance, over a six or twelve month period and, most likely, how and how frequent interventions should be applied in such a way that Internet users are maintaining high levels of alertness, while not becoming too annoyed by it.

**Acknowledgements**

This study is part of the Dutch Research Program on Safety and Security of Online Banking. This program is funded by the Dutch banking sector (represented by the Dutch Banking Association), the Police Academy and the Dutch National Police. The funders primarily took on a facilitating role in the entire research process and occasionally provided feedback on written materials, such as the questionnaire and the manuscript.

**References**


