Click me if you can!

How do users decide whether to follow a call to action in an online message?

Thomas Pfeiffer\textsuperscript{1}, Heike Theuerling\textsuperscript{2}, and Michaela Kauer\textsuperscript{2}

\textsuperscript{1} Center for Advanced Security Research Darmstadt, Germany \texttt{thomas.pfeiffer@cased.de}
\textsuperscript{2} Technische Universität Darmstadt, Institute of Ergonomics, Germany \texttt{\{h.theuerling,kauer\}@iad.tu-darmstadt.de}

Abstract. Being able to predict how internet users react when confronted with a potentially dangerous call for action in an online message (such as an e-mail) is important for several reasons. On the one hand, users have to be protected from fraudulent e-mails such as phishing. On the other hand, over-cautious users would be difficult to communicate with on the internet, so senders of legitimate messages have to know how to convince recipients of the authenticity of their messages. Extensive research already exists from both of these perspectives, but each study only explores certain aspects of the complex system of factors influencing users' reactions. In this paper the results of our efforts to integrate the various existing findings into one comprehensive model are presented, along with the results of a preliminary qualitative evaluation of some of the model's predictions using quantitative as well as qualitative measures and eye-tracking.

Keywords: decision model, e-mail, phishing, social engineering, e-commerce, trust, risk

1 Introduction

Electronic messages such as emails or messages on social networks like Twitter or Facebook are an important communication medium for private as well as business-to-business (B2B) and business-to-consumer (B2C) communication. However, the existence of diverse kinds of malicious messages (e.g. phishing, financial fraud, attachments containing malware) exposes users who blindly trust every message they receive and do everything the message prompts them to (e.g. clicking a link and entering data into a web form, opening an attachment or transferring money) to serious risks. Therefore, recipients have to decide carefully which message they should trust and which calls to action they should follow, whereas senders of legitimate messages have to find ways to gain their recipients' trust and ideally make them follow their calls to action. To inform efforts both to prevent misplaced trust and to gain legitimate trust, it is important to understand which attributes of a message, its sender, its recipient or the context in which the message is received increase or decrease the likelihood that the recipient trusts the message and follows the call to action contained in it.

In an approach to gain a comprehensive understanding of these attributes and processes, we have integrated current research mainly from the usable security and e-commerce disciplines into a general path-analytic model to predict a users' decision whether to follow a call to action in an online message or not.

In section 2 we summarize the existing literature on the topic and in particular the outcomes of existing empirical research. Section 3 describes the method used to create the initial version of
the predictive model. The model itself is described in section 4. Section 5 illustrates an early study conducted to qualitatively explore some of the model’s paths using eye-tracking.

2 Background

As described in a previous publication [1], our work draws mainly from two fields of scientific research: E-commerce and usable security. Both fields have approached user’s decisions in potentially risky online environments. While e-commerce research concentrates on users’ trust in legitimate websites and messages, usable security researchers focus on users’ reactions to social engineering attacks such as phishing e-mails. While e-commerce research is mainly concerned with websites instead of messages, we hypothesize that findings from websites can be transferred to e-mails, since studies that used both e-mails and websites as stimuli (e.g. [2, 3]) found similar effects for both.

In research from both fields, several factors affecting usage of a website or following a call to action in an e-mail have been found empirically. In accordance with the Theory of Planned Behavior [4], Kim et al. [5] found purchase on an e-commerce site to be predicted by intention to purchase. Intention to purchase in turn was found to be positively influenced by trust [5–7], perceived/expected benefit [8, 5] and negatively influenced by perceived risk [5, 8]. Aiken et al. [9] conceptualized willingness to provide information to a website as the behavioral component of trust. Similarly, intention to adopt e-services was found to be positively influenced by expected benefit/usefulness [10], trust (directly [11] or mediated by usefulness [10]), and risk (directly [11] and via usefulness [11, 10]). Hardee et al. [12] found both perceived risk and expected benefit to influence intention to engage in secure behavior.

In the studies by Blais et al. and Figner et al. [13, 14] risk taking is seen as an evaluation of the benefits expected and the perceived risk. Chang et al. [7] found risk perception to influence attitude towards reading commercial e-mail and thereby intention to read it. Weber et al. [15] postulated differences in risk preference to be caused by attitudes toward perceived risk or by different perceptions of the risk. Blais et al. [13] found risk perception to be a good predictor of risk taking. The same applies to a study [16] where risk taking was operationalized as visiting a website with an SSL error. Depending on content and context riskiness of a situation is perceived differently between and within individuals [15, 18]. A different perception of the respective risk was found to be the reason for gender differences in risk-taking [18]. According to Weber [15] there is a wide range of factors influencing these perceptions, like outcome feedback from previous risky decisions, aspiration levels, trust, expectations, and loss functions for outcomes that deviate from expectations. More than the perceived risk Hanoch et al. [19] found the respective perceived benefit to mediate the propensity to take risk.

Furthermore, increased trust was found to decrease perceived risk [5, 7] and vice versa [20, 10, 7]. Trust was found to increase expected benefit [10], but Fogg et al. [21] also found "usefulness of information" to be a predictor of a website’s perceived credibility.

According to the model of organizational trust by Mayer et al. [22], trust is preceded by the trustee’s perceived ability, benevolence and integrity. Gill et al. [23] confirmed the model empirically and found that the trustor’s propensity to trust is only effective in influencing the trustor’s intention to trust in a specific trustee in the absence of clear signals of the trustee’s high or low ability, integrity and benevolence. However, other studies [5, 24, 7, 25] found a general positive influence of propensity to trust on specific trust. Schlosser et al. [26] confirmed the influence of perceived ability on the intention to buy from a website (which they call “trusting intention”) for users searching for specific information on a website, and the influence of perceived benevolence for users generally
browsing a site. Fogg et al. [21] found a website owner’s “company motive” to influence the site’s perceived credibility. Implied investment into a website was found to predict trust (via perceived integrity) [26, 9, 27].

Many studies investigated the effects of different attributes of the website or message itself on perceived risk, expected benefit and especially on trust. The information or content quality of a website was found to have influence on trust [5, 28, 21, 6, 7, 29](via perceived ability), as were spelling and grammar errors [30, 31]. Language which felt persuading to readers in websites [21] or e-mails [31, 2] specifically was found to decrease trust, as was the request to enter a lot of sensitive information on a website [29]. Design and structure of websites [28, 21, 29, 32] or emails [31] was found to be another factor influencing trust. The presence of a footer was found to increase trust in both websites [29, 21] and e-mails [31, 2] (via perceived ability), as was the presence of a third-party trust seal on websites [9, 21, 27–29] and e-mails [2]. Personalization of content was also found to increase trust in the authenticity of e-mails [31, 30]. Perceived privacy and security protection is another factor that increases perceived ability, benevolence and integrity and thus trust in websites [5, 29, 26] and decreases perceived risk [5]. Further investigating some of the aforementioned factors, Tsow et al. [2] found that design, trust seals, URLs and HTTPS only affected trust in emails and websites when their narrative strength was low, leading users to focus on these secondary indicators for their trust decision, whereas high narrative strength increased trust in general.

The sender’s address of e-mails was found to be an important factor influencing trust [31, 30, 33, 3], but Vishwanath et al. [34] found that only recipients with sufficient knowledge of phishing and/or general computer self-efficacy paid attention to the e-mail’s source. Classification of an e-mail by an e-mail service/program as spam was found to reduce trust in that e-mail as well [3].

Attributes of the – purported – sender of an email or owner of a website have been found to play an important role in eliciting trust from recipients/users. The most prominent of these attributes is reputation or brand which was found to increase trust [5, 27, 6, 7, 3, 30, 29] and reduce perceived risk [5, 7]. Another important factor is the recipient’s/user’s familiarity with the purported sender/website owner [5, 21, 27, 33], as well as perceived similarity to the sender/owner [24, 27]. Jagatic et al. [33] also found that men were more likely to follow calls to action contained in e-mails that purportedly came from women.

Three kinds of attributes of the recipient of an e-mail or user of a website were also found to affect their decisions regarding e-mails or websites: Knowledge/experience, personality traits and demographic attributes. Users with more knowledge of and/or experience with social engineering threats [35, 36], or better general computer knowledge [33, 36] were found to be less likely to fall for phishing e-mails. General internet experience was found to decrease vulnerability to phishing attacks [36] and have an inverted-U-shaped relation with trust in online firms [9]. A higher level of general education was found to either reduce [36] or increase [25] vulnerability to phishing attacks.

User’s perceived protection from opportunism while using e-commerce was found to increase trust in e-commerce sites [27, 6], whereas fear of financial risks was found to decrease their susceptibility to phishing [35, 36]. Workman [25] found the personality traits commitment, obedience, and propensity to trust to positively influence susceptibility to phishing attacks. Women were found to be more susceptible to phishing attacks than men [33, 36], but Sheng et al. [33] found that this effect is mostly mediated by lower technology knowledge and training. Age was found to correlate either with higher [25] or lower susceptibility to phishing [33, 36], though Sheng et al. [36] found the effect to be completely mediated by higher education, exposure to anti-phishing training, internet experience and fear of financial risks.
High e-mail load was found to increase a person’s susceptibility to phishing attacks, and people who perceived their computer as less vulnerable to attacks were found to perceive less risk from ignoring their browser’s SSL certificate warnings [16].

3 Method

The first step in the creation of our model was the extraction of empirically found and/or verified antecedents of users’ behavior in risky online scenarios from the existing literature, as well as the direction and form of their relationships among themselves and to the actual behavior. In the next step, similar constructs where merged in order to reduce the number of constructs in the model. Some of the studies we used put their findings in the context of theoretical models of intra-personal processes but others did not, making the integration of the findings non-trivial. Therefore we tried to integrate the empirical findings reported without theoretical models into the theoretical models we found, based on theoretical considerations. We then inserted the construct of ”Threat-Awareness” (see section 4 for details) into the model to further explain connections found in the literature. Since the model at this point was deemed too complex for easy presentation and discussion, antecedents were aggregated to groups for simplified representation of the model. This simplified representation was then discussed with two groups of scientists from the fields of user-centered trust in interactive systems and ergonomics and refined according to their comments.

4 Results

Figure 1 shows the simplified representation of the model. The predicted behavior is labeled ”Follow call to action”, it is preceded by ”Intention to follow”. The intention to follow the call to action is in turn positively influenced by the recipient’s trust in the message’s sender and the message itself, as well as the recipient’s expected benefit from following the call. It is negatively influenced by the risk a user perceives when following the call to action, moderated by the amount of possible loss. Trust and perceived risk influence each other negatively, trust affects expected benefit positively, and perceived risk influences it negatively.

The most important group of variables influencing the aforementioned variables are the attributes of the message itself. The actual content (the call to action itself and the context in which it is embedded, as well as personalization and the narrative strength of the content) influence trust as well as perceived risk and expected benefit. Formal aspects such as design, grammar and spelling correctness, language, the presence of logos, fine-print footers or third-party trust seals all affect trust. Third-party seals also affect perceived risk.

Another important group of variables are attributes of the message’s purported sender. In accordance with Gill et al. [23], trust in the sender is influenced by the sender’s perceived ability, integrity and benevolence. Those are in turn influenced by the purported sender’s reputation or brand, the recipient’s familiarity with and perceived similarity to him or her or the sender’s gender (especially male recipients are more likely to trust female senders).

We introduced the variable ”threat awareness” into the model to reflect the individual differences in the approach to trust decisions. Threat-aware recipients are both motivated and able to evaluate authenticity of a message by technical means and only factor attributes of the message’s purported sender into their trust decision if they perceive the message as being authentic. Threat-unaware recipients, on the other hand, are not aware that a message’s sender information and content can be
Fig. 1. Simplified representation of the model to predict whether a user follows a call to a potentially risky action contained in an online message.
forged and thus trust a message if they trust its purported sender. Threat-awareness is a combination of the knowledge of or experience with the threat of forged online messages on the one hand the technical knowledge and/or experience necessary to effectively evaluate a message’s authenticity. This is reflected in the model as multiple moderating effects: Perceived authenticity moderates the the effect of attributes of the purported sender on trust, and that moderating effect is in turn moderated by threat awareness. Additionally, threat awareness moderates the effect of attributes of the message on perceived authenticity. In fact, perceived authenticity does not play any role for the decision of trust-unaware recipients. Besides knowledge and experience, current threat-awareness is also influenced by contextual factors such as time pressure or distractions affecting the recipient, or the narrative strength of the message.

Another group of variables are the recipient’s personality traits. For example, recipients’ general propensity to trust affects their trust in the sender/message, risk-taking propensity affects risk perception. Recipients with high obedience and commitment generally show higher intention to follow calls to action.

Attributes of the context in which the message is received influence trust in the sender/message, perceived risk, benefit, and current threat awareness. Variables in this group are for example the contextual plausibility of the email, the perceived vulnerability of the currently used computer to attacks.

Demographic attributes of the recipient such as age, gender, education or marital status correlate with perceived risk as well as threat awareness.

5 Study

We conducted a preliminary study with a small sample in order to test whether a combination of eye-tracking and open-ended as well as closed questions can be used to evaluate the model qualitatively.

In this study, participants had to evaluate the authenticity of three e-mails. We expected participants which scored higher on a threat-awareness test to focus their attention more on aspects of the e-mail which can be used to identify authenticity more reliably (such as sender or recipient address) and less on aspects which are very easy to fake, such as design elements, than participants which scored low on the test.

Furthermore we expected the reasons given for their decisions to reflect factors found in the model.

5.1 Study design

The study was a laboratory test taking place at the Technische Universität Darmstadt in Germany. Participants were presented with screen shots of three different e-mails in PDF format. Each e-mail was purportedly sent by PayPal Inc. One of them was authentic, the other two were phishing emails. All emails were in German, we use translated text here.

The authentic featured the usual PayPal email design and informed the user about changes in their terms of service, including correct links to PayPal websites and the complete new terms of service. It was well-written, without grammatical or spelling errors, addressed the recipient by full name and was signed with “Your PayPal team”.

The first phishing email appeared to be from “www.paypal.de” <service@verifiedbyvisa.com> and addressed the recipient with ”Dear PayPal member”. It included the PayPal logo and some
yellow and light-blue elements, but not the original PayPal design. It told the recipient that PayPal’s
danger prevention system had detected suspicious credit card charges in the recipient’s account and
therefore the recipient should log into their account to regain access, followed by link. A sidebar
contained security advice, asking recipients not to give their password to fraudulent websites. The
language was of very bad quality, to the point that many sentences were barely understandable.

The second phishing mail appeared to be from “Pay.Pal-Sicherheit EURO” <kunde@pay-pal-
sicherer-euro.be> addressed the recipient with “dear user”. It told the recipient about an impending
suspension of their account due to a missing data synchronization and asked them to synchronize
their data by 11/08 to prevent the suspension by clicking a link (which did not contain a URL in
the link text). This email contained no layout elements. There were no grammatical errors and no
spelling errors other than some missing spaces.

After looking at each email (while their eye-movements were tracked using a remote eye-tracking
device), participants were asked whether they perceived the email as authentic and would follow
the call to action or not. They were also asked to state reasons for their decision. After the task
was completed, the participants took a threat awareness quiz which consisted of questions covering
both knowledge of threats encountered online and knowledge of technical aspects of internet security.
The answers to both sets of questions are added to a general threat awareness score. Furthermore
information about age, gender, education, income, internet usage and experience as well as which
electronic devices they owned was collected.

5.2 Results

We were able to analyze the data of ten participants, all of them male, between the age of 22 and
27, all of them were college students. All participants use the internet daily. Three participants
had three to seven years of internet experience, four participants had seven to ten years and three
participants had more than ten years of internet experience.

Eight participants identified all e-mails correctly. One participant perceived the first phishing
e-mail as authentic and indicated that he would follow the call to action, for the reason that the
e-mail advises recipients not to give away their passwords. One participant perceived the second
e-mail as authentic, but indicated he would not follow the call to action because he did not perceive
the reason it gave for the data synchronization as reliable. This means that in fact, no participant
would have actually given his data to phishers. Therefore no influences on susceptibility to phishing
could be tested.

We compared gaze dwell times for certain areas of interest from the eye-tracker between par-
ticipants which scored low on the threat awareness test and those which scored high (split at the
50th percentile). Contrary to what we expected, we found that participants who scored higher on
the threat-awareness test tended to focus more on the actual content of the e-mail (49% vs. 41% of
the dwell time on the legitimate mail, 40% vs. 30% for the first phishing mail and 43% vs. 33% for
the second phishing mail) and less on header information (4% vs. 9%, 1% vs. 6% and 11% vs. 13%)
then participants with lower scores. Both groups paid equally little attention to design elements
such as logos or colored areas.

The reasons given for the perception of mails as authentic or forged were in line with our model.
The phishing mails were identified as such mainly because of spelling/grammar mistakes (eight
participants), attempts to persuade recipient to submit data or click a link (five participants), their
design/layout (three), the poor quality of arguments (three) and lack of personalization (three).
Accordingly, the authentic mail was identified as such mainly because it did not ask recipients to
enter any information or click any link (five participants), because it mentioned legal matter such as terms of service and relevant laws (four participants), because of its elaborateness (three), its design/layout (three), its language quality (two) and because it contained a personalized salutation (two participants).

5.3 Discussion

The reasons participants stated for their authenticity judgments were all in line with the previous research our model is based on. All of the reasons were related to the content of the emails, to their language or their design. In the case of our study, these criteria actually lead to correct decisions, since the phishing mails we used were only poor imitations of authentic PayPal e-mails. However, all the criteria mentioned except for not asking to click any link or submit any information can be spoofed relatively easily by having a native speaker write the text and by copying the design and layout of authentic e-mails. The fact that none of the participants mentioned a suspicious sender or recipient address as the reason for their decision to judge an e-mail as fraudulent, and that participants who scored higher on the threat awareness test actually spent less time looking at the header area of the e-mails seems counter-intuitive at first. On the other hand, these participants may have known that even a sender address can be spoofed (which wasn’t the case in our examples, though), and since they did not have the chance to check out for example the more reliable return-path header (it was not included in the screen shot), they may have dismissed the header information as unreliable. It also has to be noted that the high/low threat-awareness groups were based only on relative scores, since standardized scores do not yet exist for the quiz we used.

Limitations. Since our stimuli turned out to be too easy to identify correctly, we cannot draw any conclusions regarding effectiveness of different decision strategies from our study. To this end, stimuli with more subtle differences between authentic and fake ones would have to be used. We cannot explore the effects of demographic attributes based on our sample either, since it was very small and very homogeneous, so in further studies we would try to recruit larger and more heterogeneous samples.

The insights we gained from our very limited preliminary study holds the promise that eye-tracking, combined with qualitative and quantitative measures of decision strategies and influencing factors can give valuable information for understanding recipients’ behavior when facing potentially fraudulent and dangerous online messages.

6 Conclusion and Outlook

The model presented here integrates as well as extends previous decision models for user reactions to potentially dangerous websites or online messages in order to provide increased predictive power as well as generalizability compared to previous models. This model constitutes basic/fundamental research with various ways of application. Plans already exist to use it to inform algorithms which will warn users about potentially dangerous actions, or prevent cues used to evaluate trustworthiness, risk and usefulness of an electronic message from misleading users. Knowledge about factors influencing users’ decisions is also useful when creating material to teach users to improve their decision-making process: They should be taught to apply decision criteria which often lead to good
decisions and avoid those which often lead to bad decisions. Since the model encompasses both malicious and legitimate messages, it can also guide the design of online messages which legitimately attempt to call users to appropriate action (e.g. in commercial or official communication).

The next step is a larger-scale quantitative empirical evaluation of the model using structural equation modeling techniques to verify the predicted influence paths and determine their relative strengths. Further qualitative studies using eye-tracking could additionally provide further insights into the detailed processes underlying the influence of the different factors.

Acknowledgements. We thank Kathrin Tello and Julian Eurich for conducting the study.

References