Computers in Human Behavior 66 (2017) 75–87

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Priming and warnings are not effective to prevent social engineering attacks

M. Junger [a, *], L. Montoya [b], F.-J. Overink [a, 1]

[a] University of Twente, Faculty of Behavioral, Management and Social Sciences (BMS), The Netherlands
[b] University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science, The Netherlands

ARTICLE INFO

Article history:
Received 6 June 2016
Received in revised form
26 August 2016
Accepted 7 September 2016
Available online 24 September 2016

Keywords:
Priming
Warning
prevention
Social engineering
Phishing
Disclosure of personal information

ABSTRACT

Humans tend to trust each other and to easily disclose personal information. This makes them vulnerable to social engineering attacks. The present study investigated the effectiveness of two interventions that aim to protect users against social engineering attacks, namely priming through cues to raise awareness about the dangers of social engineering cyber-attacks and warnings against the disclosure of personal information. A sample of visitors of the shopping district of a medium-sized town in the Netherlands was studied. Disclosure was measured by asking subjects for their email address, 9 digits from their 18 digit bank account number, and for those who previously shopped online, what they had purchased and in which web shop. Relatively high disclosure rates were found: 79.1% of the subjects filled in their email address, and 43.5% provided bank account information. Among the online shoppers, 89.8% of the subjects filled in the type of product(s) they purchased and 91.4% filled in the name of the online shop where they did these purchases. Multivariate analysis showed that neither priming questions, nor a warning influenced the degree of disclosure. Indications of an adverse effect of the warning were found. The implications of these findings are discussed.

1. Introduction

The present study investigates whether users can be helped to protect their personal information against direct requests. It tests the effectiveness of two interventions that aim to protect users against social engineering attacks, namely priming through cues to raise awareness about the dangers of online activities and warnings against the disclose of personal information.

Social engineering has been defined as '*The science of using social interaction as a means to persuade an individual or an organization to comply with a specific request from an attacker where either the social interaction, the persuasion or the request involves a computer-related entity*' (Mouton, Leenen, Malan, & Venter, 2014). The success of a social engineering attack often depends on a target either being willing or tricked into disclosing personal information. Many cyber-attacks begin with users who unknowingly or mistakenly disclose personal information to attackers. For instance, attackers send users phishing emails containing a link to a webpage that requests the user to fill in personal information (Hong, 2012; Purkait, 2012). Today, it is estimated that there are thousands of unique phishing emails send to users on a daily basis. For instance, in March 2016 229,265 unique phishing e-mail reports (campaigns) were received by Anti-Phishing Working Group (APWG) from consumers and 123,555 unique phishing websites were detected (APWG., 2016). Most targeted industries were the retail industry (42.71% of all mails) and the financial industry (18.67% of all mails), meaning that attackers impersonated for instance a retail store or a bank.

Research has shown that the online and offline worlds are connected (Mesch, 2012). Offline trust, such as trust in social institutions and trust in individuals, is associated with trust online (Mesch, 2012). This is true for cybersecurity as well: from a security perspective physical and digital security are interconnected (Dimkov, 2012). Montoya, Junger, & Hartel, (2013) argued that crime can be situated on a continuum from only traditional or physical crime to completely digital crime. Many types of crime today have aspects of both. For instance, in a random sample of

[*] Corresponding author. University of Twente, Faculty of Behavioral, Management and Social sciences (BMS), Po Box 217, 7500 Ae Enschede, The Netherlands.
E-mail address: M.Junger@UTwente.nl (M. Junger).

[1] F.- J. Overink was a student of the University of Twente at the time that this study.

http://dx.doi.org/10.1016/j.chb.2016.09.012
0747-5632/© 2016 Published by Elsevier Ltd.

crimes reported to the police, 41% of all frauds and 16% of the threats have in part a digital modus operandi (MO). To commit burglaries, offenders hardly ever use ICT. But in 2.9% of the residential burglaries, however, bank cards were stolen which were later used to steal money from a bank account (Montoya et al., 2013). It is therefore important to understand how users react offline to understand the online threats. This is also illustrated with evidence on identity theft. Information necessary to execute a crime such as phishing or other forms of identity theft often come from the victims themselves. For instance, in a random survey of the Australian population on identity theft, victims reported that the information originated from email (18.3%), from information placed on a website, such as an online shopping website (15.7%), from information placed on social media (e.g. Facebook, Linked-in) (6.9%) and/or from text messages (SMS) (6.4%) (R. G. Smith & Hutchings, 2014). Information also originated from direct contact with the victim, namely from a face-to-face meeting (e.g. a job interview or a door knock appeal) (7.5%) or a telephone conversation (10.5) (R. G. Smith & Hutchings, 2014; pp., table 18). Although the percentages cannot be summed up (because attackers may have used several methods) these figures show that in about half of the incidents of identity theft, information used was provided voluntarily by victims, and in a sizable proportion of these cases, information was provided in a direct, not online contact. Consequently, user's perceptions about what constitutes sensitive personal identifiable information (PII), and their reactions to requests for sensitive PII matters for computer security. A better understanding and quantification of privacy and security perceptions is needed (Cranor, 2016).

Phishing attacks become increasingly sophisticated. For instance, spear-phishing mails or ‘targeted attacks’, are an increasingly popular (Hong, 2012; Wueest, 2014). In targeted phishing attacks, attackers attempt to better mimic genuine emails by using personal details from customers. Genuine emails from online shops usually mention the name of the customer and what was purchased. They often refer to the bank account number, and mention only the last three digits of the account (in the Netherlands), for safety reasons. The more an attacker knows from his potential victim, the better he can mimic genuine emails. Examples of phishing emails can be found on http://www.social-engineer.org/wiki/archives/Phishing/Phishing-eBay.html.
Although getting this additional information takes time, an advantage of targeted attacks is that they are relatively successful (Jagatic, Johnson, Jakobsson, & Menczer, 2007; Rocha Flores, Holm, Nohlberg, & Ekstedt, 2015). An experiment using social network information, showed that people are 4.5 times more likely to fall for a phishing message sent from an existing contact compared to standard phishing attacks. Out of 512 students at the Corps of Cadets at West Point receiving a spear phishing email mentioning a problem with their Grade Report, 80% clicked on the link in the email (Ferguson, 2005).

Many security-related organizations have lamented on the vulnerability of users to social engineering attacks and their tendency to disclose information (Adams & Sasse, 1999; Kirlappos & Sasse, 2012). To overcome this situation, customers are informed about the occurrence of cybercrime and phishing emails and they receive tips and instructions on how to protect themselves against disclosing PII. For example, most banks and online shops have webpages devoted to security. Often, they use leaflets and warnings to inform users or customers about what they should and should not do to protect themselves against possible attacks (see for instance: http://pages.ebay.com/help/account/recognizing-spoof.html).
However, it is not well known how effective these warnings messages are. The present study tests warning and priming for cybercrime to study their effectiveness in a sample of shoppers selected in a shopping area.

1.1. Disclosure of personal information

Trust can be defined as “a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or the behavior of another” (Rousseau, Sitkin, Burt, & Camerer, 1998). Trust determines the way in which an individual approaches other people (Glaeser et al., 2000; Fehr et al., 2003; Kosfeld et al., 2005).

Humans tend to conform and are relatively trustworthy by nature: trust has evolutionary survival value e.g. children need to trust others in order to be able to learn (Dawkins, 1993; Morgan & Laland, 2012; Ostrom, 1998; Penner, Dovidio, Piliavin, & Schroeder, 2005). (Fetchenhauer & Dunning, 2009; Glanville & Paxton, 2007; P. L.; Harris et al., 2012; P. L.; Harris & Corriveau, 2011). Most researchers believe that adults start with the presumption of truth (Burgoon & Levine, 2010; Mills, 2013). In general, having trust in others has positive outcomes for individuals (Dohmen, Falk, Huffman, & Sunde, 2012; Fetchenhauer & Dunning, 2009; Frattaroli, 2006; Glanville & Paxton, 2007; Ostrom, 1998).

Disclosure or self-disclosure can be defined as the process of communicating information about oneself verbally to another person (Cozby, 1973). Besides having relatively high trust, humans seem to have low thresholds for disclosing personal information and do it relatively often. Most of the studies on disclosure have been done in the field of psychology and mental health. These studies showed that self-disclosure has positive outcomes (Dindia, Allen, Preiss, Gayle, & Burrell, 2002; Cozby, 1973; Omarzu, 2000; Sprecher, Treger, & Wondra, 2013; Worthy, Gary, & Kahn, 1969).

Although in general trusting others has positive outcomes (Dindia et al., 2002), personal information can be abused relatively easily (Acquisti, Brandimarte, & Loewenstein, 2015; Gross & Acquisti, 2005; Hann, Hui, Lee, & Png, 2002). Research has been done to investigate the degree to which users are prepared to disclose personal information online and the situations in which disclosure increases or decreases.

John, Acquisti, and Loewenstein (2011) presented four experiments in which disclosure was measured by investigating whether subjects answered questions on deviant behavior such as ‘Having sex with the current husband, wife, or partner of a friend’ or ‘Making a false insurance claim’. Three experiments showed that users disclosed more personal information on unprofessional looking websites, which are arguably more likely to misuse it than on professional looking websites which were less likely to misuse it (John et al., 2011, p. 868). In other words, ‘individuals are prone to disclose in contexts that downplay privacy concerns—ironically, even when such contexts are likely higher in both objective and perceived disclosure danger’ (John et al., 2011, p. 868). Interestingly, in an experiment in which users were cued to think about privacy, these contextual differences i.e. type of website-disappeared and all users had similar rates of disclosure. John et al. (2011, p. 868) concluded that their results ‘stand in contrast to the considerable body of privacy research that is premised on the assumption of rational choice’, which states that people make trade-offs between privacy and other concerns, implying that disclosure is the result of this rational weighing of costs and benefits, in which objective costs – such as an unprofessional looking website - should prevent or at least decrease disclosure.

Joinson, Reips, Buchanan, and Schofield (2010) combined a survey with an experiment. Their findings differ from John et al. (2011), and are more in line with rationality. They report that self-disclosure was reduced when the context involved a weak privacy policy and low trust. In all other combinations of trust (high versus low) and privacy (high versus low), self-disclose was higher.

This study used similar questions on deviance as John et al. (2011). The authors conclude that their study supports the notion – from Olson, Grudin, and Horvitz (2005) - that it is not so much the nature of the information disclosed that matters, but the person to whom information is disclosed. When that person is trusted, users easily disclose (Joinson et al., 2010).

In a review of the literature, Acquisti et al. (2015) conclude that disclosure levels are highly malleable. For instance, users are sensitive to default settings: they tend to stick to the pre-programmed default settings. They are easily manipulated by website design, some websites frustrate or confuse users into disclosing personal information. In addition, the perception of having control over one's personal information reduces concerns over privacy and increases disclosure (Acquisti et al., 2015).

1.2. *User's problems and interventions*

Because humans start with the assumption of truth - see Section 1.1 - it probably will always be difficult for them to detect fraud (Burgoon & Levine, 2010). In addition, because of their high degree of malleability, users need some form of assistance to protect their privacy (Acquisti et al., 2015). Accordingly, security officers and researchers have been looking for preventive measures (Abraham & Chengalur-Smith, 2010) to help them overcome specific problems. It has been argued that users have 1) insufficient knowledge of what they can do, 2) are not aware of the consequences of mistakes, 3) are not concerned - sufficiently - to behave in a safe way and 4) lack a social context. Below we discuss these four reasons and the possible interventions to mitigate these risks.

1.2.1. *Lack of knowledge*

Users have insufficient knowledge and lack strategies to identify vulnerabilities and scams (Hong, 2012; Purkait, 2012). For instance they do not know the methods attackers use to execute their crimes (Kritzinger & von Solms, 2010; 2013). Previous research supports this thesis. For instance, in an experimental study of warnings before downloading a pdf-file, many subjects explained that they trusted the antivirus program or thought that pdf files were always safe (Krol, Moroz, & Sasse, 2012).

An experiment by Grazioli and Wang (2001) studied the determinants of success or failure in determining if a website was fraudulent or not. They concluded that subjects are all equally good at inspecting the website or at perceiving cues of deception. However, failure is caused because the generation of hypotheses, based on perceived cues, cannot be assessed. The authors concluded that ‘priming subjects to facilitate hypothesis generation is useless, unless we provide them with effective knowledge to assess the primed hypothesis (Grazioli & Wang, 2001, p. 201).

Acquisti et al. (2015)'s conclusion, based on their literature review, is that many users were unaware of the fact that the information requested from them by researchers could be abused: ‘People are often unaware of the information they are sharing, unaware of how it can be used, and even in the rare situations, when they have full knowledge of the consequences, of sharing, uncertain about their own preferences.’ (Acquisti et al., 2015, p. 513). Due to their lack of knowledge, users follow their ‘default’ option, which is trust, obedience and disclosure of information.

1.2.1.1. Training. To improve knowledge educational programs have been evaluated in a number of experimental studies. Some experiments evaluated training users and showed that this helped them to fall less often for phishing mails. For instance, the “School of phish” (Kumaraguru et al., 2009) compared to what extent three groups of participants each fall for phishing scams (172 participants in control, 172 in one training - and 171 in two-training condition). The control group received no training, one group was trained once and the last group received training twice. The results suggest that training reduces the likelihood of participants falling for phishing scams. However, even after training 20% of participants fell for phishing scams. For the present overview we found 18 experimental studies on anti-phishing training (see Appendix). Among these 18 studies only three had no positive effect on phishing recognition (Caputo, Pfleeger, Freeman, & Johnson, 2014; Davinson & Sillence, 2010; Kearney & Kruger, 2014). It is noteworthy that Kearney and Kruger (2014) found an adverse effect: after training more users proved information in response to a mock phishing attack. Only four of the eighteen studies were executed in a ‘real world’ corporate environment (Aburrous, Hossain, Dahal, & Thabtah, 2010; Caputo et al., 2014; Kearney & Kruger, 2014; Kumaraguru, Sheng, Acquisti, Cranor, & Hong, 2008). Among these four, two (Aburrous et al., 2010; Kumaraguru et al., 2008) reported a successful effect of training, the other two had - as mentioned above - no or an adverse effect.

1.2.1.2. Information. A less comprehensive way to provide knowledge is providing more limited information in the form of leaflets. Bullee, Montoya, Pieters, Junger, and Hartel (2015) looked at social engineering in person and tested the effectiveness of an intervention that consisted of a poster and a leaflet sent by email plus a reminder in the form of a key fob. ‘Attackers’ visited university staff in their office, in an effort to trick employees into handing over their electronic office door-key. The intervention decreased the number of employees who handed over their key to the attacker from 62.5% (control group) to 37.0% (intervention group). To summarize, this experiment suggests that a small intervention may have a relatively large effect.

Bullee, Montoya, Junger, and Hartel (2016) tested the effectiveness of an information campaign to counter a telephone-based social engineering attack directed at university staff. The aim of the mock attack was to request employees to download a file to ‘neutralize malware’. 46.15% of the employees in a control condition followed the instructions of the attacker and downloaded the file. Only 9.1% of those exposed to an intervention 1 week prior to the attack complied, whilst 54.6% of those who were exposed to the intervention 2 weeks prior to the attack complied, a very similar compliance rate to that of the control group. This research suggests that scam awareness-raising campaigns can reduce vulnerability, but only on the short term.

1.2.2. *Awareness of consequences*

Not all researchers found that lack of knowledge is a major cause of victimization. A study in which victims of scams reported about their experiences showed that scam victims often have better than average background knowledge in the area of the scam content’ (Lea, Fischer, & Evans, 2009, p. 137). Accordingly, some studies focused of the more emotional side of prevention: making users aware, or perhaps afraid, of the possible consequences of disclosure of information. Users may lack insight in the severity of the consequences if they fill in their personal information on a spoofed website (Purkait, 2012; West, 2008).

1.2.2.1. Priming. People's behavior can be altered when they are exposed to certain sights, words or sensations (Dolan, Hallsworth, Halpern, King, & Vlaev, 2010; Kenrick, Neuberg, & Cialdini, 2005). These sights, words or sensations prime people: they activate knowledge or certain goals and makes them ready for use (Kenrick et al., 2005). Priming often works outside conscious awareness (Dolan et al., 2010; Kenrick et al., 2005). In the physical world, a large amount of research supports the existence of priming effects (Cameron, Brown-Iannuzzi, & Payne, 2012; Wentura & Degner, 2010). A recent meta-analysis of 167 independent studies revealed that priming tasks were associated with relevant behaviors (r = 0.28) and explicit measures (r = 0.20) (Cameron et al., 2012). However, recently concerns have been raised about the replicability and robustness of findings on priming research (Kahneman, 2012).

Although until recently, research on priming has been relatively uncontroversial, lately, researchers have started to cast doubts on the robustness of priming results. Several issues were raised by Kahneman (2012): 1) the recent exposure of fraudulent researchers (see also Stroebe, Postmes, & Spears, 2012), 2), multiple reported failures to replicate salient results in the priming literature (e.g., Doyen, Klein, Pichon, & Cleeremans, 2012; C. R.; Harris, Coburn, Rohrer, & Pashler, 2013; Pashler, Coburn, & Harris, 2012; Pashler & Wagenmakers, 2012; Shanks et al., 2013), and 3) the growing belief in the existence of a pervasive file drawer problem, which undermines the usefulness of meta-analyses (Bower, 2012; Pashler & Wagenmakers, 2012). In answer to criticism, Cesario and Jonas (2014) argued that the variation in outcomes of priming effects is due to individual differences and to context. Accordingly the differences in outcomes are to be expected and are consistent with current research on the mind (Cesario & Jonas, 2014). In conclusion, in the physical world, priming effects may be less well established than previously stated.

A few studies applied priming to the online world and different types of results were reported. Acquisti, John, and Loewenstein (2012) looked at the effectiveness of raising privacy concerns on the disclosure of information online. Cues to think about phishing consisted of a number of pages with pictures of phishing emails and the request to categorize them as ‘phishing’ or ‘non-phishing’. In this study, cues about phishing led to a decrease in disclosure (Acquisti et al., 2012). Similarly, Parsons, McCormac, Pattinson, Butavicius, and Jerram (2015) found that priming subjects in an experiment by telling them that they participate in a study on recognizing phishing emails leads to improved performance in comparison with a control group that was not primed in this way.

Zhang and Xu (2016) studied the impact of a ‘frequency nudge’ and a ‘social nudge’ on the evaluation of a mobile application. The ‘frequency nudge’ indicated how often a mobile app uses several types of data of the phone (namely, location, read contacts, read call log, vibrate, post notification, and camera). The ‘social nudge’ is the percentage of other users of this app that approve the use of each type of data permissions (Zhang & Xu, 2016). It is interesting that the two nudges had opposing effects. The social nudge led to positive feelings about the app, while the frequency nudge led to negative feelings about the app and lower comfort level in sharing data with the app (Zhang & Xu, 2016). Grazioli and Wang (2001) executed a somewhat similar priming experiment. One third of the subjects in each condition received an abridged version of a Federal Trade Commission pamphlet on Internet fraud, similar to the priming in the present study which focused on cybercrime. Another third received a pamphlet and a watch list of possible manipulations. The last third was the control condition, which did not receive any information. Grazioli and Wang (2001) did not find any effect of priming on several online-trust related variables as well as the ability to detect deception.

Sundar, Kang, Wu, Go, and Zhang (2013) investigated the impact of priming on the disclosure of information on websites. They found that priming subjects for the advantages of personalized websites - either explicitly of by demonstrating it - lowered the level of disclosure on generic websites. Disclosure was highest in the control situation on the generic website. The authors conclude that “the default tendency among online users is to be relatively careless in protecting private information” (Sundar et al., 2013, p. 815), in line with what researchers reported for communication in the physical world (Morgan & Laland, 2012; Ostrom, 1998; Penner et al., 2005). In other words: high disclosure is the norm. Priming subjects on the advantages of personalization of websites made subjects relatively more suspicious than no priming (Grazioli, 2004).

In total we found 7 studies that investigated priming. Three two showed a decrease in disclosure after being primed (Acquisti et al., 2012; Grazioli, 2004; Parsons et al., 2015). Two presented mixed findings (Sundar et al., 2013; Zhang & Xu, 2016) and two showed no effect (Grazioli & Wang, 2001; Leon et al., 2012).

In conclusion, the sparse literature on priming in relation to information security shows that disclosure seems to be responsive to priming in sometimes counter-intuitive ways.

1.2.2.2. Warnings. Warnings are a much more direct way to convey a message than priming. Traditional warnings have been successful in influencing behavior (Argo & Main, 2004; Wogalter, Laughery, & Mayhorn, 2012, pp. 868–894). There are guidelines for effective warnings, which were summarized by (Wogalter et al., 2012, pp. 868–894). Relevant principles include: 1) brevity, warnings should be as brief as possible; 2) design for the low-end receiver, meaning that warnings should not be directed at an ‘average person’ but for people who have lower competence, education, knowledge, and/or the elderly or the disabled, for instance.

Warnings have been also generally used to warn against website unsafety. Kirlappos and Sasse (2012) used a warning to inform about website safety. Warnings helped improve user behavior, but a relatively large part of the users did not adjust his/her behavior when monetary rewards were at stake. A similar conclusion was drawn by (Christin, Egelman, Vidas, & Grossklags, 2012).

A study by Zhang, Wu, Kang, Go, and Sundar (2014) found an adverse effect. This study investigated users' behavior in the presence of warnings for mobile sites. The security cue was operationalized by the presence or absence of a security warning banner showing that a trusted security certificate could not be detected. In contrast to expectations, on the stimulus website when the security cue was present, participants disclosed more social media information, i.e., number of Facebook friends, Twitter ID, number of Twitter followers, number of people followed on Twitter, (Zhang et al., 2014).

Krol et al. (2012) found that 81.7% of their subjects ignored a warning when downloading a pdf file. Similarly, Wu, Miller, and Garfinkel (2006) found that users ignore toolbar warnings. Other research also concluded that browser warnings overall did not have positive effects (Egelman & Schechter, 2013; Egelman, Cranor, & Hong, 2008; Xiao & Benbasat, 2015).

1.2.3. Users are not motivated

Security is not always the main concern for users. They may prefer convenience or are tricked for financial motives or other basic human motives (West, 2008). Förster, Liberman, and Friedman (2007) state - in line with Goals System Theory - that goals are organized in a hierarchical network. Many authors have argued that - during the course of their work - warnings, for instance, interrupt users and are therefore considered a nuisance, keeping them away from their primary goal, which is to get their job done (Bada, Sasse, & Nurse, 2015; Krol et al., 2012). Similar findings were reported by (Krol et al., 2012), who found that pop-up warnings are usually disrupting users from their primary task. Due to their disrupting nature, users tend to skip them to continue with their primary task. This may explain why priming or warnings showed these varying effects on security behavior.

1.2.4. Users need the social context: social proof

Social proof is our tendency to look to others - the social context - for cues on what to use and how to behave (Cialdini & Goldstein, 2004; Nolan, Schultz, Cialdini, Goldstein, & Griskevicius, 2008). In several studies, Cialdini and colleagues (Nolan et al., 2008) reported that observing others, or receiving information about what others do, leads to increased levels of that specific behavior. Previous research also reported that observing other disclosing information leads to increased levels of disclosure (Acquisti et al., 2012). As far as we know, one study experimented with the social proof principle online, to guide them to disclose less personal information. Das, Kramer, Dabbish, and Hong (2014) experimented on Facebook and reported that showing people the number of their friends that used security features lead to 37% more viewers to explore the promoted security features compared to raising awareness about security issues.

1.3. *Psychological mechanisms and effectiveness of interventions*

From the above, it is concluded that interventions to decrease users' online risk have had limited success. Several other possible explanations can be offered to explain the difficulties facing policy makers to improve user's behavior.

1.3.1. *Personal relevance*

Users may have only a vague idea about cybercrime and its possible causes. In general, it has been established that most people pay attention to warnings or other messages when these are perceived to be personally relevant and they do not pay attention to messages that lack personal relevance (Sagarin, Cialdini, Rice, & Serna, 2002). For instance non-alcohol users do not pay attention to warnings labels required on alcoholic beverages (Stewart & Martin, 1994). The explanation is that these warnings are not relevant to them. In line with this, Krol et al. (2012) reported that subjects who had previously experienced fraud, scams or viruses less often download a pdf file than those who had no such experience. To take personal relevance into account some training programs mentioned above ('School of phish' (Kumaraguru et al., 2009)) train users only when they were victims of a mock phishing attack, which, according to the authors, contributed to training success. Replication by Caputo et al. (2014) and Davinson and Sillence (2010) which did not restrict training to 'victims', could not replicate previous successes. Accordingly, personal relevance may indeed be key to training success.

1.3.2. *Optimism bias*

People believe that negative events are less likely to happen to them than to others, and they believe that positive events are more likely to happen to them than to others (Weinstein, 1980). This occurs for many types of risk, including crime victimization (Chapin & Coleman, 2009; Weinstein, 1980) and infringement of one's privacy (Baek, Kim, & Bae, 2014). Weinstein (1987) listed four major causes for an optimism bias: (1) the belief that if the problem has not yet appeared, one is exempt from future risk; (2) the perception that the problem is preventable by individual action; (3) the perception that the hazard is infrequent; and (4) lack of experience with the hazard (Weinstein, 1987, p. 496). In support of the relevance of the fourth cause (no experience), Baek et al. (2014) showed that personal experience with privacy infringement online decreases the optimism bias. Accordingly, many people believe they are relatively invulnerable, which in turn, decreases the personal relevance of priming for or warning against cybercrime. Herley (2009) argued that given that experience with cybercrime victimization is relatively rare, this optimism bias is in fact a rational reaction, given that there are also costs of staying on the safe side online, such as extra time spend on additional security tasks.

1.3.3. *Distraction*

People cannot attend to everything at once. This phenomenon has also been called 'selective attention' (Stewart & Martin, 1994). It is also known that people are not very good at executing more than one task at the time (Pashler, 1998). To perform a task correctly, distracting cues must be inhibited in order to be able to perform the task (May, Kane, & Hasher, 1995). Distraction can lead to 'attentional blindness' or 'looking without seeing' (Lavie, 2010). A related approach is provided by the Goals System Theory, as explained in Section1.2.3.

1.3.4. *The person to disclose to*

Joinson et al. (2010) concluded that it is not the nature of the information that matters but that person to whom information is disclosed what is important. When that other person is trusted, users easily disclose (see also Section 1.1).

In conclusion, there is a limited amount of research showing the practical effectiveness of information security preventive interventions on end-users' behavior (H. J. Smith, Dinev, & Xu, 2011). The experimental literature on training or warning users against disclosing personal information has produced inconclusive results. Furthermore, the existent research has methodological limitations. About half of the studies found (seven out of fifteen) were executed with university students in a university laboratory. Using students restricts the representativeness of the findings to a specific age and educational group. A sizable minority (about six studies) was executed online, generally using Amazon's mechanical Turk. Although the mechanical turk has clear advantages (Buhrmester, Kwang, & Gosling, 2011), recent research showed limitations in terms of reliability (Rouse, 2015) and some doubts as to the quality of the data from the Indian respondents — who constitute about 46% of the subject pool. In addition, most studies collected data with the help of a PC. Only one study collected information in a telephone interview (Bullee et al., 2016) or face to face (Bullee et al., 2015). Only one study collected information specifically from consumers, namely (Xiao & Benbasat, 2015). There is a pressing need for more research in this area.

The goal of present study is to test two interventions that aim to prevent disclosure of personal information that could be used in social engineering attacks, such as spear-phishing. The main question is whether priming or a warning leaflet with information cautioning against disclosing personal information and social engineering, prevent users from being socially-engineered. With respect to priming, we assume that priming about cybercrime will raise negative emotions, increase awareness about possible victimization and, in turn, lead to less disclosure.

The present study contributes to the literature in a number of ways. First, many studies have relied on behavioral intentions, which in the case of disclosure often do not match with actual behavior (H. J. Smith et al., 2011). The present study recorded actual disclosure by asking subjects to disclose personal information in a questionnaire. Second, much previous research on disclosure used questions about deviant behavior (i.e., sexual behavior, illegally download music) (John et al., 2011; Joinson et al., 2010) which is not directly relevant for security related information. Other studies investigated disclosure in response to a phishing email. The present study chose a different method and requested information about one's email address, online shopping and one's bank account number by asking subjects to fill in a questionnaire. Third, previous studies often collected data online, while the present study used personal contact. Previous studies often studied a student population, but the present study investigated a sample more representative of the general population. We asked passersby in a city center to fill in our questionnaire.

We received permission for the study from the Ethical

Committee of the Faculty of Electrical Engineering, Mathematics and Computer Science of the University of Twente. Subjects were debriefed after having filled in the questionnaire.

2. Method

2.1. Sample

The data were collected by approaching visitors at the central shopping area a medium-sized town (between 100,000 and 250,000 inhabitants) in the East of the Netherlands and asking whether they would like to participate in the research. The research was carried out on five different days between May 18 and May 26, between 10AM and 5 PM, to increase the representativeness of the study. Special attention was paid to the distribution of the different conditions over different days and times of the day. We aimed to collect data on about 100 subjects for each of the three experimental conditions.

2.2. Experimental procedure

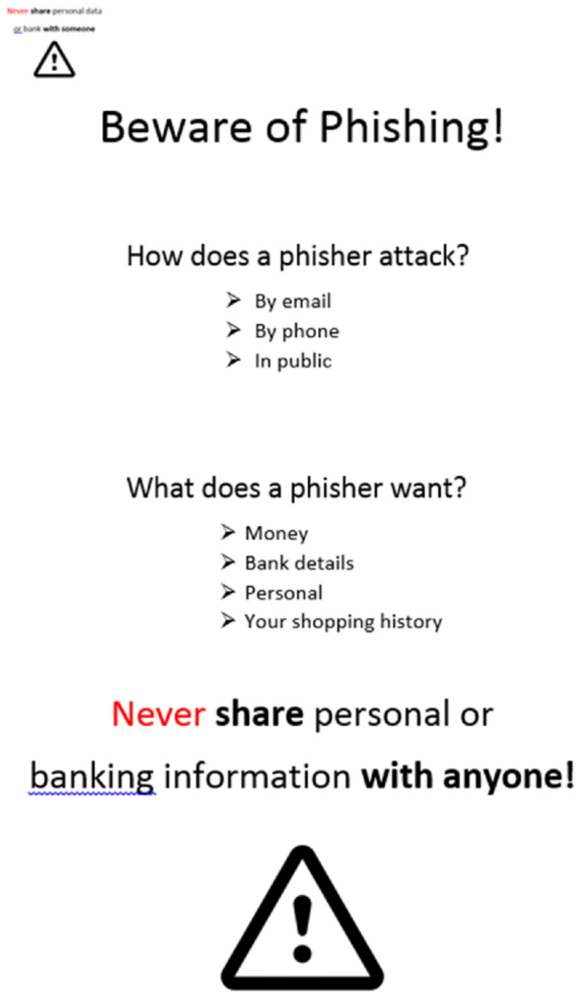
2.2.1. Experimental conditions

The questionnaire was based on Beunder, Kerkers, and Orij (2014). It consists of two parts: 1) a common part that all subjects answered, and 2) four questions on cybercrime and privacy on the internet, that were meant to prime subjects on the topic of cybercrime. There were three conditions: two experimental conditions and a control condition.

1) The priming condition. In the first experimental condition subjects had to answer 4 questions about cybercrime, namely: 1) Are you familiar with the term phishing? 2) Are you aware of the amount of personal information you share on the Internet and that is publicly accessible? 3) Do you use Facebook? If so, what are generally your privacy settings? 4) Have you ever been scammed on the Internet (for example through phishing)? These questions were placed in the middle of the questionnaire. As mentioned above, their aim was to prime subjects on the topic of cybercrime victimization.
2) The warning condition. In the second experimental condition, prior to getting the questionnaire, subjects were handed over a leaflet of A4 format (Fig. 1) before they received the questionnaire. In addition, a small part of this leaflet was placed at the top of each page, as a reminder (Fig. 2). This leaflet was inspired by (Bullee et al., 2015) who created a poster for an intervention in a study aiming to reduce the success of social engineering attacks. The leaflet attempted to be brief, focused on the right issues and attempted to be as simple and direct as possible, following suggestions on successful warnings (Wogalter et al., 2012, pp. 868–894). For instance, it was mentioned explicitly ‘do not provide personal information or bank account information to someone you do not know’, which should increase compliance (Stewart & Martin, 1994).
3) A the control condition, the subjects answered only the questions of the common part of the questionnaire i.e. they were not either primed or given the leaflet.

2.2.2. Measures

Disclosure was measured with four questions as well as the sum score of the positive scores (‘total risk’). These questions were chosen as they can be easily abused in spear phishing attacks. The four questions were the following.



Never share personal data
or bank with someone

Beware of Phishing!

How does a phisher attack?

- By email
- By phone
- In public

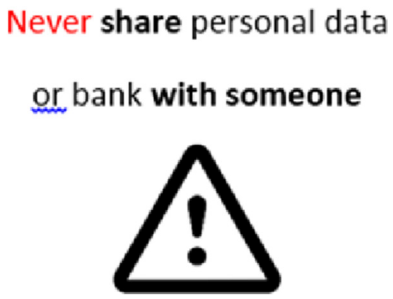
What does a phisher want?

- Money
- Bank details
- Personal
- Your shopping history

Never **share** personal or

banking information **with anyone!**

Fig. 1. Warning leaflet.



Never **share** personal data

or bank **with someone**

Fig. 2. Small warning message.

1) Writing down an email address. A request to write down one's E-mail address was made at the beginning of the questionnaire. Subjects were first asked if they wanted to receive a copy of the results of the study, and then if they could fill in their email address, which was also meant to prevent double subjects. At that moment, subjects in the priming condition had not yet received the four questions to prime them on cybercrime but the subjects in the ‘warning condition’ had already received the warning, as the leaflet was distributed before handing over the questionnaire.
2) Bank account information. The bank account information was asked towards the end of the questionnaire. To protect subject's

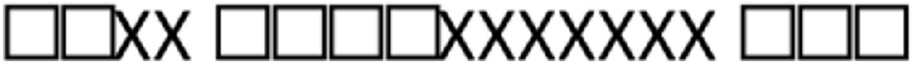


Fig. 3. Bank account number, the subject was asked to fill in the squares.

privacy, only half of the digits of the bank account number (in total 18 digits) were asked, (see Fig. 3). For instance, in the present case, an attacker could easily fake an email from a specific shop, referring to products that were bought including the digits of the bank account. 8% of the phishing attacks are "purchase confirmation" attacks (Atkins & Huang, 2013). We assume that practically everybody in the Netherlands has a bank account from the age of 15. A recent survey showed that of the 15–24 year olds in the Netherlands, only 1% does not have a bank account (Schors & Werf, 2014).

3) When subjects previously had shopped online, they were asked what they had bought. A broad range of options were offered and there was space on the questionnaire for filling in something else.
4) Subjects were asked about the name of the web shop where they bought their products. The answer categories included major Dutch online shops and a space was provided to fill in other shop names.

For these four dependent variables, answer categories were: 'filled in' (1) or 'not filled in' (0). To preserve subject's privacy, no specific information was coded or used.

5) 'Total risk'. A 'total risk' variable is the sum of the positive answers on the previous questions, namely the number or times subjects did disclose personal information on the four dependent variables. Total risk is a continuous variable with scores from 0 to 4. These questions, with the exception of the email address (see above), were all placed at the end of the questionnaire.

Information on age, sex and educational level was recorded. Besides socio-demographic variable, two questions were asked about computer knowledge (see Table 1) and the amount of time subjects spend online (in term of days a week and hours a day).

Table 1
Characteristics of subjects, means and % (N = 278).

	Control	Priming	Warning	%. Mean
All				
Sex. % females	56.3%	57.4%	61.4%	58.3%
Mean age (years) **	30.2	26.3	34.2	30.1
Mean number of hours online	4.6	5.3	5.5	5.1
Mean number of days a week online	6.5	6.6	6.4	6.5
Education. High, in %	18.8%	22.3%	28.7%	23.1%
Manage a computer well to very well	54.2%	53.2%	55.7%	54.3%
Total	96	94	88[a]	278
Online shoppers only				
Sex. % females	56.3%	57.6%	61.0%	58.2%
Mean age (years) *	27.1	26.5	31.3	28.2
Mean number of hours online	4.8	5.2	5.4	5.1
Mean number of days a week online	6.6	6.6	6.5	6.6
Education. High, in %	19.5%	22.8%	32.9%	24.7%
Manage a computer well to very well	57.5%	54.3%	63.6%	58.2%
Total	87	92	77*[b]	256

$^{*}p < 0.05$, **, $p < 0.01$ $^{***}p < 0.001$.
[a] N = 87 for education.
[b] N = 76 for education.

2.2.3. *Analysis*

To describe the data frequencies and cross tables are presented. To analyze the effectiveness of the intervention, Logistic regression and Poisson regression analysis were performed. Control variables were age and, to control for non-linear effects, age-square (see also below). Analysis on online shopping, reporting the name of the web shop and the total risk sum were performed on the online shoppers only.

3. Results

3.1. *The sample*

A total of 290 persons filled in the questionnaire. Two subjects were not online at least once a week, and accordingly, they were deleted from the analysis. Due to 10 missing values (6 on age and 4 on education), 278 subjects remained in the analysis. A subgroup of 256 (92%) of the subjects shopped online.

Regarding the representativeness of the sample, 58% of the subjects were females, which is higher than that of the general population. They were also slightly younger than the general city population, where 22.2% of the population is younger than 20, in the present study this is 26.3%; 16.2% of the population is older than 65, in the present study this is 5.4%.

The control group consisted of 96 subjects, the priming condition of 94 subjects and the warning condition of 88 subjects. Table 1 shows the characteristics of all subjects and of those who shop online. The three groups contain equal number of males and females and they do not differ with respect to educational level, computer knowledge, and time spent online. However, they do differ in terms of age. The mean age is 30 years old, however, the warning group is somewhat older (mean age = 34.2) and the priming group is somewhat younger (mean age = 26.3). Accordingly, to assess the effectiveness of our interventions, it is necessary to control for age. Because the graphs suggested a non-linear relationship, we also controlled for age square.

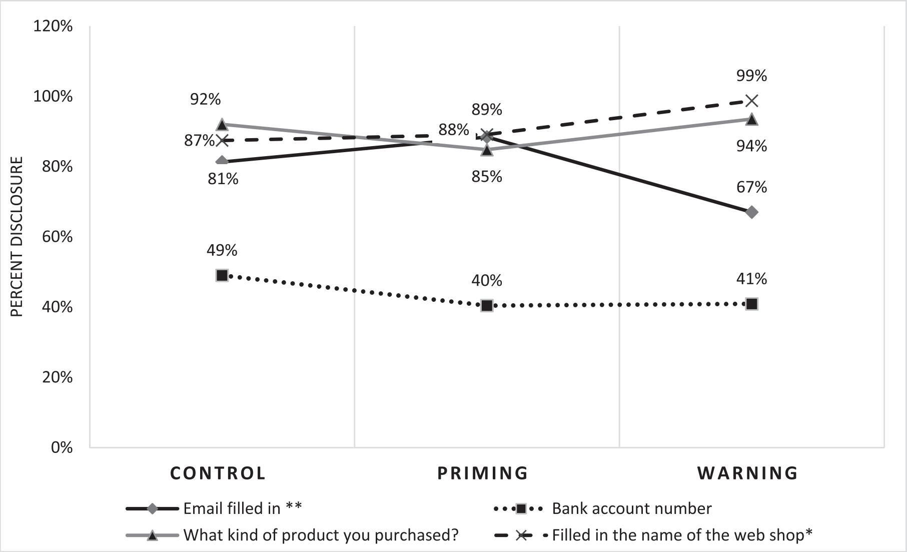
Among the subjects in the priming condition, 67% mentioned that they had heard about phishing and 10% mentioned having been victim of online fraud (N = 94).

There were small to moderate (Cohen, 1977) correlations between the four dependent variables. Reporting an email address was positively related to reporting a bank account ($r = 0.20$, $p < 01$) and reporting the product that one bought was also positively associated with reporting the name of the web shop ($r = 0.36$, $p < 0.01$). Reporting one's bank account was also positively associated with mentioning the product bought online ($r = 0.15$, $p < 0.05$). Of the total 6 correlations, three were not statistically significant.

3.2. *Effectiveness of the interventions*

Fig. 4 shows the bivariate relationship between the experimental condition and the dependent variables. Two of the four outcomes are significantly related to the experimental condition. In the warning condition, a smaller number of subjects report their email address, namely 67%, in contrast to 81.3% (control group) and 88.3% (priming group) ($p < 0.01$).

43.5% of all subjects filled in their bank-account information, but there were no differences between the conditions. Among the



Chi Square test: * p<.05, **, p<.01 *** p<.001

Fig. 4. Disclosure by experimental condition, in %. All respondents (N = 278), selection online shoppers (N = 256). Chi Square test: $^{*}p < 0.05$, **, $p < 0.01$ $^{***}p < 0.001$.

online shoppers, 89.8% of the subjects filled in the kind of product(s) they purchased. In the warning condition, significantly more subjects disclosed the name of the web shop, namely 98.7% in contrast to 87.4% in the control group and 89.1% in the priming group. The total risk score is also unrelated to experimental condition (results not shown).

Multivariate analyses were executed to study the effectiveness of the interventions, while controlling for differences in age and age square (Tables 2 and 3). The interventions did not affect the number of subjects reporting their email address, their bank account, the type of product bought online and the total risk score. Reporting the name of the online shop is related to the experimental condition but in an unexpected way. The logistic regression shows that the presence of a warning increases the likelihood of disclosing the name of the online web shop, in line with the cross tabular analysis presented in Fig. 4.

Table 3
Poisson regression predicting the effect of a warning or priming on the total risk score (number of items disclosed) on online shoppers (N = 256).

	IRR	p	Std. Error
Age	1.01		0.15
Age square	1.00		0.00
Priming	0.95		0.08
Warning	0.94		0.09
Constant	2.73	***	0.28

IRR = incidence rate ratio.
$X^2 = 3.56$, $p < 0.001$.

Table 2
Effect of a warning or priming on disclosure of reporting an email address and a bank account number, logistic regression analysis. All respondents (N = 278), selection online shoppers (N = 256).

All respondents	Email				Bank account number			
	OR	p	95% C.I. OR		OR	p	95% C.I. OR	
			Lower	Upper			Lower	Upper
Age	0.92		0.82	1.05	1.12	*	1.02	1.23
Age square	1.001		0.999	1.002	0.999	*	0.998	1.000
Priming	1.52		0.66	3.50	0.63		0.35	1.15
Warning	0.55		0.27	1.14	0.67		0.36	1.23
Constant	35.39	**			0.19			
−2 Log likelihood = 256.69; Cox & Snell R Square = 0.10; Nagelkerke R Square = 0.15								
−2 Log likelihood = 366.15; Cox & Snell R Square = 0.05; Nagelkerke R Square = 0.07								
Online shoppers only	**Product mentioned**				**Online shop**			
Age	1.15		0.97	1.36	1.15		0.95	1.38
Age square	0.998		0.996	1.000	0.998		0.996	1.000
Priming	0.44		0.17	1.19	1.13		0.44	2.88
Warning	1.21		0.35	4.19	11.43	*	1.39	94.09
Constant	1.67				1.00			
−2 Log likelihood = 155.03; Cox & Snell R Square = 0.05; Nagelkerke R Square = 0.10								
−2 Log likelihood = 131.80; Cox & Snell R Square = 0.07; Nagelkerke R Square = 0.16								

$^{*}p < 0.05$; $^{**}p < 0.01$.

4. Discussion & conclusion

Humans tend to trust each other and tend to disclose personal information relatively easily (Acquisti et al., 2015; Morgan & Laland, 2012; Ostrom, 1998; Penner et al., 2005). In general this is believed to have beneficial effects for themselves (Frisina, Borod, & Lepore, 2004; Pennebaker & Seagal, 1999) and society at large (Acquisti et al., 2015). However, it makes them vulnerable to social engineering attacks when they provide personal information that can be abused, for instance, in spear-phishing attacks that imitate emails from an online shop. The present studied investigated whether priming or a warning leaflet prevent users from disclosing personal information in a social engineering attempt.

To this end, 278 subjects filled in a questionnaire in a shopping area in a medium-sized town in the East of the Netherlands, of which 256 subjects also had shopped online in the past. Three experimental conditions were tested: a) a priming condition, in which subjects had to answer 4 questions on cybercrime and privacy issues, b) a warning condition, in which prior to the filling in the questionnaire, subjects received a leaflet with a warning against giving away personal information, and c) a control condition that contained none of the two previous interventions but only the remaining questions. Disclosure was measured by asking the subjects' email address, 9 digits from their 18 digit bank account number, and for those who shopped online, what they had purchased and in which web shop. The total risk score was the sum of the number of times personal information was disclosed. These questions were chosen because a potential phisher could use this information to write a convincing spear phishing email impersonating a web shop and refer to the product that was bought and reporting the last three digits of the bank account.

4.1. High disclosure

Relatively high disclosure rates were found: about 80% or more of all subjects disclosed information upon request: 79.1% of subjects filled in their email address, and among the online shoppers, 89.8% of filled in the kind of product(s) they purchased and 91.4% filled in the name of the online shop where they made these purchases. However, 'only' 43.5% disclosed the 9 digits of their bank account, which is still considerable given the attractiveness of this information for criminals.

The level of disclosure found is relatively high. In relation to the email address disclosure, our results are comparable to previous research. John et al. (2011) reported in their experiment that between 32% (computerized questionnaires) and 88% (online questionnaire) of subjects reported their email address. To the best of our knowledge, only one study tested whether bank employees were willing to disclose their bank account information. In a sample of 50 respondents, 32% gave away their full-e-banking credentials (user name and password) in a telephone phishing experiment. Another 16% gave away their user name but refrained from giving their password (Aburrous et al., 2010). These findings match the findings on the present study although the requested information differs slightly. To the best of our knowledge, only one study tested disclosure with questions on online shopping or one's bank account number.

The present findings are in line with the general thesis that humans tend to be trustworthy (Morgan & Laland, 2012; Ostrom, 1998; Penner et al., 2005). Similarly, research on lying and deception concluded that humans have a "truth bias": people have a strong tendency to judge a message as truthful (Burgoon & Levine, 2010).

4.2. Lack of effectiveness

The most remarkable finding of the present study is the global lack of effectiveness of priming or of a warning to prevent disclosure of personal information. Multivariate analysis showed that neither the priming questions, nor the warning influenced the degree of disclosure, with one exception. In one situation, we found that a warning worked in the opposite way: a warning increased the tendency to disclose the name of the online shop where subjects bought online products. Bellow we discuss these results.

4.2.1. Priming

The present study showed that priming was not effective in reducing disclosure of personal information. Although past research suggested that priming was supported almost unanimously, recent research has casted doubt on the robustness of priming effects (see Section 1.2.2). Although some studies found support for priming on cybercrime to decrease disclosure (Acquisti et al., 2012) other studies did not find support for priming effectiveness (Grazioli & Wang, 2001; Zhang & Xu, 2016). Our results are in line with these later studies.

4.2.2. Warnings

We expected that handing over a leaflet with a warning before subjects filled out a questionnaire would prevent them from disclosing personal information. However, the present study showed that a warning did not decrease disclosure, and, upon being asked about the web shop's name, it unexpectedly *increased* disclosure. Many previous studies did not find that browser warnings had positive effects (Hong, 2012). For instance, Krol et al. (2012) found that 81.7% of their respondents ignored a warning when downloading a pdf file. Similarly, Wu et al. (2006) found that users ignore toolbar warnings. In sum our findings are in line with most research in this field.

Although our studies provide comparable research to previous research it remains important to investigate the effectiveness of possible interventions but also to try to unravel which possible processes seem to disrupt the effectiveness of interventions. Why did priming on cybercrime or a warning not raise alarm bells in our subjects? Several options seem plausible.

4.3. Explanations for lack of effectiveness

4.3.1. No intervention

The first possible explanation is that our interventions were not noticed. It often happens that users fail to notice or hardly pay any attention to warning messages (Stewart & Martin, 1994). We believe this is not the case in the present experiment. In the priming condition, subjects answered questions on phishing and privacy. Accordingly, they had to read and to fill in, and we noted that there were no missing values on these questions. Our warning was read before handing over the questionnaire. Similarly, we do not think anyone skipped this (in this condition) and that explained the lack of effectiveness. The priming and the warning were noticed but ignored.

4.3.2. Lack of knowledge

Our interventions did not train users but our warning provided - although briefly - some information. Our observations during data collection suggest that many subjects did not make the connection between the information that was provided and the more general issue of cybercrime or phishing. This suggests that indeed - as was explained in 1.2.1. - users have no clue about what type of information may be useful to attackers. In support of this line of reasoning is the fact that our subjects were much more restrictive in proving information on their bank account - a topic which links more directly to online fraud - than their email address of online shopping information. So, a sizable part of our respondents seemed to understand that disclosing one' bank account information was unwise.

4.3.3. Goal hierarchy

Users may not give priority to security. Usually security is a secondary goal, they may prefer focusing on their primary task or chose for convenience (see Section1.2.3). The present findings suggest that this may explain the present findings. Our subjects were shopping: this was their primary goal. They were interrupted by the researchers to participate in a research and specifically, to fill out a questionnaire. Given that their mind was focusing on their primary task, they filled it out without paying too much attention to privacy considerations. This would be in line with Krol et al. (2012) results who stated that pop-up warnings disrupt users from their primary task (section 1.2.3).

4.3.4. Social proof

Because our subjects were sometimes walking and shopping with others when we approached them, it happened several times that they filled in their questionnaire in pairs or small groups. From our own observations we saw that this seemed to increase disclosure. This happened, for instance, when some subjects thought that we were testing the quality of their memory when requesting information on their bank account. This is in line what other have observed - that observing others disclosing information leads to increased levels of disclosure (Acquisti et al., 2012) (see also Section 1.2.4).

4.3.5. Personal relevance

It is possible that the priming questions and the information on our leaflet were not considered as personally relevant. Research shows that people pay more attention to warnings or other messages when these are perceived to be personally relevant (Sagarin et al., 2002) (see Section 1.3.1). However, this explanation probably does not hold for the findings of the present study. An additional analysis among the subjects in the priming group (N = 94) - the only group to whom questions about previous victimization were asked - did not reveal differences between subjects who were or who were not victimized by previous cybercrime experience. It cannot be concluded to which extent this explanation holds for the present results.

4.3.6. Optimism bias

People tend to believe that negative events are less likely to happen to them than to others, and they believe that positive events are more likely to happen to them than to others (Weinstein, 1980) (see Section 1.3.2). Because cybercrime victimization is relatively rare, this optimism bias is a rational reaction given the cost of staying on the safe side online (Herley, 2009). Although this is a plausible line of reasoning, we did not see any indications of optimism bias in the present study.

4.3.7. Distraction

People cannot attend to everything at once. This has also been called ‘selective attention’ (see Section 1.3.3). In a situation in which subjects are interrupting their shopping to fill in a questionnaire, they might have focused on the questionnaire and ignored the warning. This, in turn, may lead subjects to follow their ‘default’ option, which is obedience and disclosure of information. We believe it is possible that this phenomenon explains the lack of effectiveness of our interventions. It is interesting to note that to distract people is a common method used by attackers to defraud them (Stajano & Wilson, 2011).

4.3.8. The person to disclose to

Joinson et al. (2010) concluded that it is not the nature of the information that is important but that person to whom information is disclosed what is important. In the present study, the researcher who approached the shoppers was a friendly looking young man. In addition, many city inhabitants are aware of the university and may have a favorable view of students. This phenomenon may explain in part the high disclosure rates. However, subjects reported information on their bank account about half as often as they did for other personal information. Despite their confidence, the type of information still seems to matter considerably and the ‘who to disclose to’ can be only part of the explanation of the high disclosure rates.

4.3.9. Adverse effects and reactance or boomerang effects

In one occasion, we found that the warning had an adverse effect: the warning significantly increased the disclosure of the online web shop. Although it should be noted that this adverse effect was found for only one dependent variable, namely the name of the online shop, it is worth paying more attention to this finding. Noteworthy, in the physical world, adverse effect have been reported. For instance Weinstein and Klein (1995) found that 4 interventions to reduce optimism bias had no effect on their subjects. Instead, they reported that focusing attention on risk factors can increase the optimism bias. There is empirical evidence that repeated exposure to the same message at high levels of intensity can produce irritation, counter-argumentation, and behavior that is inconsistent with the message (Fransen, Smit, & Verlegh, 2015; Stewart & Martin, 1994). For instance, in survey participation research, it was showed that for sensitive issues, such as sexual behavior or drug use, stronger assurances of confidentiality elicit higher response rates or better response quality (Singer, 2004). However, for innocuous research, stronger assurances of confidentiality appear to backfire, leading to less willingness participation, and greater expressions of suspicion and concern about what will happen to the information requested (Singer, 2004).

Adverse effects have been noted in online security as well. Wolff (2016) reviewed some of the perverse effects present in the defense of computer systems. For instance, Zhang et al. (2014) investigated the effect of security warnings on a fictitious mobile website. They found that a security warning led to a *higher* perceived threat and *more* disclosure. The presence of a security cue had a significant main effect on users' attitudes toward the website. Specifically, when there was a security cue individuals expressed less positive attitudes, perceived more privacy threats, and had lower intention to use the mobile website (Zhang et al., 2014, p. 113). Participants also disclosed more social media information (i.e., number of Facebook friends, Twitter ID, number of Twitter followers, number of people followed on Twitter) on the experimental website when the security cue was present (Zhang et al., 2014, p. 113). The possibility of adverse effects emphasizes the need to study the effectiveness of interventions before their launch.

4.4. Policy implications

It is concerning that users displayed such high disclosure rates. An important advice is that users need more knowledge about how attackers operate, hence user education is necessary. In developing user education it is important to determine priorities, teaching everything may amount to learn nothing much (Krol et al., 2012). In addition, it may be important to focus on vulnerable groups. Studies by Ronald Dodge, Coronges, and Rovira (2012), Kumaraguru et al. (2009, 2008) showed that anti-phishing training helped when it was delivered to victims of mock attacked. Finally, users are sensitive to what others do Cialdini, Martin, and Goldstein (2015); Goldstein, Martin, and Cialdini (2008); Nolan et al. (2008). In the present study, users seem to have imitated unwise behavior. The same technique may be used to convey the message that others display safe behavior. A large scale experiment on Facebook showed that showing people the number of their friends that used security features was most effective, and drove 37% more viewers to explore the promoted security features compared to the non-social announcement (thus, raising awareness) (Das et al., 2014).

We mentioned above that security interventions have adverse effects. Our findings are important because the types of interventions we tested in the present study are similar to what the media and policy makers are doing: warning against the dangers of cybercrime and ‘improving awareness’. For instance, in early 2010, US President, Barack Obama, created the National Initiative for Cybersecurity Education (NICE). NICE aims to create an operational, sustainable, and continually improving program for cybersecurity awareness, education, training, and workforce development to improve the nation's ability to deal with cybersecurity threats (National initiative for cybersecurity education (NICE), 2009; Paulsen, McDuffie, Newhouse, & Toth, 2012). Similarly, in Europe, the European Union organizes the European Cyber Security Month (ECSM), which is a European Union advocacy campaign that took place in October 2014. It is part of the EU-U.S. collaboration in the field of cybersecurity. One of the main aims of the European Cyber Security Month campaign is to 'generate general awareness about cyber security, which is one of the priorities identified in the EU Cyber Security Strategy' and to 'generate specific awareness on Network and Information Security (NIS), which is addressed in the proposed NIS Directive'. Despite the wide use of awareness campaigns, the effectiveness of these campaigns has not been studied, to the best of our knowledge. The present study and other studies showing adverse effects emphasize the need to provide further empirical evidence for the practitioner community to be able to identify awareness campaign designs that are effective based on specific subject characteristics.

The present study is subject to limitations. We do not know to what extent our sample is representative of the broader population. Shoppers were approached to participate in our study, but not all agreed to participate. It is difficult to evaluate the differences between the participants and those who refused. Comparing the age-range with that of the general population showed no large deviations. Also, in support of generalizability, the findings on disclosure are similar to previous findings. Nevertheless, replication of the findings in different populations is necessary. Also, controlling for additional factors, such as personality factors would be valuable from the point of view of tailoring awareness campaigns as much as possible. Another limitation of our research is that we were not able to ask our subjects the reasons for their disclosure, due to the fact that many subjects wanted to continue with their shopping.

The present study contributes to the literature in several ways. First, it investigates actual disclosure of information, which is usable in a spear-phishing attack, by asking for an email address, the product and the name of the online shop and half of the digits that constitute the bank account number. Several studies, such as Acquisti et al. (2012) and Joinson et al. (2010) requested participants to report deviant behavior such as shoplifting. Many studies disclosure of information that is not directly relevant to online security or investigated general personal identifiable information without focusing on relevant information directly usable to launch online spear-phishing attacks. Second, the present study collected information in a real life setting by interviewing and requesting personal identifiable information from subjects in the real world, of all ages, in a shopping area. This in contrast to many studies which collected data in a laboratory setting, with university students or university staff.

In conclusion, this study found relatively high disclosure rates. When asked for information, a large proportion of subjects provided information on their email address (79.1%), bank account information (43.5%), the type of product(s) they purchase online (89.8%) and the name of the online shop (91.4%). Neither priming nor a warning influenced the degree of disclosure. Users should be educated in what constitutes sensitive information and how it can be abused in online attacks.

Acknowledgments

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 318003 (TREsPASS). This publication reflects only the author's views and the Union is not liable for any use that may be made of the information contained herein.

Appendix 1. List of 18 experimental studies on anti-phishing training.

Aburrous et al. (2010); Alnajim and Munro (2009); Caputo et al. (2014); Davinson and Sillence (2010); R Dodge, Rovira, Radwick, and Shevchik (2011); Greis, Nogueira, and Kellogg (2012); Jansson and von Solms (2011); Kearney and Kruger (2014); Kumaraguru et al. (2009); Kumaraguru, Rhee, Acquisti, et al. (2007); Kumaraguru, Rhee, Sheng, et al. (2007); Kumaraguru et al. (2008); Kumaraguru, Sheng, Acquisti, Cranor, and Hong (2010); Mayhorn, Murphy-Hill, Zielinska, and Welk (2015);S. Sheng, Holbrook, Kumaraguru, Cranor, and Downs (2010); Steve Sheng et al. (2007); Yang, Tseng, Lee, Weng, and Chen (2012)

References

Abraham, S., & Chengalur-Smith, I. (2010). An overview of social engineering malware: Trends, tactics, and implications. *Technology in Society, 32*(3), 183–196. http://dx.doi.org/10.1016/j.techsoc.2010.07.001.

Aburrous, M., Hossain, M. A., Dahal, K., & Thabtah, F. (2010). Experimental case studies for investigating e-banking phishing techniques and attack strategies. *Cognitive Computation, 2*(3), 242–253. http://dx.doi.org/10.1007/s12559-010-9042-7.

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science, 347*(6221), 509–514. http://dx.doi.org/10.1126/science.aaa1465.

Acquisti, A., John, L. K., & Loewenstein, G. (2012). The impact of relative standards on the propensity to disclose. *Journal of Marketing Research, 49*(2), 160–174. http://dx.doi.org/10.1509/jmr.09.0215.

Adams, A., & Sasse, M. A. (1999). Users are not the enemy. *Communications of the ACM, 42*(12), 40–46.

Alnajim, A., & Munro, M.. (2009). An anti-phishing approach that uses training intervention for phishing websites detection. Paper presented at the Information Technology: New Generations, 2009. ITNG'09. Sixth International Conference on.

APWG. (2016). *Phishing activity trends report, 1nd quarter 2016: Anti-Phishing working group (APWG).* https://docs.apwg.org/reports/apwg_trends_report_q1_2016.pdf.

Argo, J. J., & Main, K. J. (2004). Meta-analyses of the effectiveness of warning labels. *Journal of public policy and marketing, 23*(2), 193–208.

Atkins, B., & Huang, W. (2013). A study of social engineering in online frauds. *Open Journal of Social Sciences, 1*(03), 23.

Bada, M., Sasse, M. A., & Nurse, J. R. C. (2015). *Cyber security awareness campaigns: Why do they fail to change behaviour?* UK: Oxford. Retrieved from http://www.cs.ox.ac.uk/files/7194/csss2015_bada_et_al.pdf.

Baek, Y. M., Kim, E.-m., & Bae, Y. (2014). My privacy is okay, but theirs is endangered: Why comparative optimism matters in online privacy concerns. *Computers in Human Behavior, 31*, 48–56.

Beunder, K., Kerkers, M., & Orij, J. (2014). *The effects of awareness on the disclosure of Personally Identifiable Information.* Enschede, Nl: University of Twente.

Bower, B. (2012). The hot and cold of priming: Psychologists are divided on whether unnoticed cues can influence behavior. *Science News, 181*(10), 26–29. http://dx.doi.org/10.1002/scin.5591811025.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3–5.

Bullee, J.-W., Montoya, L., Junger, M., & Hartel, P. (2016, 14–15 Jan 2016). Telephone-based social engineering attacks: An experiment testing the success and time decay of an intervention. Paper presented at the Cyber Security R&D Conference (SG-CRC) 2016, Singapore.

Bullee, J.-W., Montoya, L., Pieters, W., Junger, M., & Hartel, P. H. (2015). The persuasion and security awareness experiment: Reducing the success of social engineering attacks. *Journal of Experimental Criminology, 11*(1), 97–115. http://dx.doi.org/10.1007/s11292-014-9222-7.

Burgoon, J. K., & Levine, T. R. (2010). Advances in deception detection. *New directions in interpersonal communication research*, 201–220.

Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review, 16*(4), 330–350. http://dx.doi.org/10.1177/1088868312440047.

Caputo, D. D., Pfleeger, S. L., Freeman, J. D., & Johnson, M. E. (2014). Going spear phishing: Exploring embedded training and awareness. *IEEE Security and Privacy, 12*(1), 28–38. http://dx.doi.org/10.1109/MSP.2013.106.

Cesario, J., & Jonas, K. J. (2014). Replicability and models of priming: What a resource computation framework can tell us about expectations of replicability. *Understanding Priming Effects in Social Psychology, 129*.

Chapin, J., & Coleman, G. (2009). Optimistic bias: What you think, what you know, or whom you know? *North American Journal of Psychology, 11*(1).

Christin, N., Egelman, S., Vidas, T., & Grossklags, J. (2012). It's all about the Benjamins: An empirical study on incentivizing users to ignore security advice. In

Financial cryptography and data security (pp. 16–30). Springer.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology, 55*(1), 591–621. http://dx.doi.org/10.1146/annurev.psych.55.090902.142015.

Cialdini, R. B., Martin, S. J., & Goldstein, N. J. (2015). Small behavioral science-informed changes can produce large policyrelevant effects. *Behavioral Science & Policy, 1*, 21–27.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences (rev. ed.)*. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Cozby, P. C. (1973). Self-disclosure: A literature review. *Psychological bulletin, 79*(2), 73.

Cranor, L. F. (2016). Informing (public) policy. Paper presented at the Symposium On Usable Privacy and Security (SOUPS 2016), https://www.usenix.org/conference/soups2016/presentation/cranor.

Das, S., Kramer, A. D., Dabbish, L. A., & Hong, J. I. (2014). Increasing security sensitivity with social proof: A large-scale experimental confirmation. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*.

Davinson, N., & Sillence, E. (2010). It won't happen to me: Promoting secure behaviour among internet users. *Computers in Human Behavior, 26*(6), 1739–1747. http://dx.doi.org/10.1016/j.chb.2010.06.023.

Dawkins, R. (1993). Viruses of the mind. *Dennett and his critics: Demystifying mind*, 13–27.

Dimkov, T. (2012). *Alignment of organizational security policies: Theory and practice* (Ph. D.). Enschede: University of Twente.

Dindia, K., Allen, M., Preiss, R., Gayle, B., & Burrell, N. (2002). Self-disclosure research: Knowledge through meta-analysis. *Interpersonal Communication Research: Advances Through Meta-analysis*, 169–185.

Dodge, R., Coronges, K., & Rovira, E. (2012). Empirical benefits of training to phishing susceptibility. *IFIP Advances in Information and Communication Technology, 376 AICT*, 457–464.

Dodge, R., Rovira, E., Radwick, Z., & Shevchik, J. (2011). Phishing awareness exercises. In *Proceedings of the 15th colloquium for information systems security education, june* (pp. 120–125).

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2012). The intergenerational transmission of risk and trust attitudes. *The Review of Economic Studies, 79*(2), 645–677. http://dx.doi.org/10.1093/restud/rdr027.

Dolan, P., Hallsworth, M., Halpern, D., King, D., & Vlaev, D. (2010). *MINDSPACE: Influencing behaviour through public policy*. London, UK: Cabinet Office and Institute for Government.

Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PloS One, 7*(1), e29081.

Egelman, S., Cranor, L. F., & Hong, J. (2008). You've been warned: An empirical study of the effectiveness of web browser phishing warnings. In *Conference on human factors in computing systems - Proceedings*.

Egelman, S., & Schechter, S. (2013). The importance of being earnest [in security warnings]. In *LNCS. Lecture notes in computer science* (Vol. 7859, pp. 52–59) (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).

Fehr, E., Fischbacher, U., Von Rosenbladt, B., Schupp, J., & Wagner, G. G. (2003). *A nationwide laboratory examining trust and trustworthiness by integrating behavioural experiments into representative surveys*. Retrieved from CESifo Working Paper Series No. 866; IZA Discussion Paper No. 715. Available at SSRN: http://ssrn.com/abstract=385120. or http://dx.doi.org/10.2139/ssrn.385120.

Ferguson, A. J. (2005). Fostering e-mail security awareness: The West Point carronade. *EDUCASE Quarterly, 28*(1), 54–57.

Fetchenhauer, D., & Dunning, D. (2009). Do people trust too much or too little? *Journal of Economic Psychology, 30*(3), 263–276.

Förster, J., Liberman, N., & Friedman, R. S. (2007). Seven principles of goal activation: A systematic approach to distinguishing goal priming from priming of non-goal constructs. *Personality and Social Psychology Review, 11*(3), 211–233. http://dx.doi.org/10.1177/1088868307303029.

Fransen, M. L., Smit, E. G., & Verlegh, P. W. (2015). Strategies and motives for resistance to persuasion: An integrative framework. *Frontiers in psychology, 6*.

Frattaroli, J. (2006). Experimental disclosure and its moderators: A meta-analysis. *Psychological bulletin, 132*(6), 823–865. http://dx.doi.org/10.1037/0033-2909.132.6.823.

Frisina, P. G., Borod, J. C., & Lepore, S. J. (2004). A meta-analysis of the effects of written emotional disclosure on the health outcomes of clinical populations. *The Journal of Nervous and Mental Disease, 192*(9), 629–634.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *Quarterly Journal of Economics*, 811–846.

Glanville, J. L., & Paxton, P. (2007). How do we learn to trust? A confirmatory tetrad analysis of the sources of generalized trust. *Social Psychology Quarterly, 70*(3), 230–242.

Goldstein, N. J., Martin, S. J., & Cialdini, R. B. (2008). *Yes!: 50 scientifically proven ways to be persuasive*. New York: Simon & Schuster.

Grazioli, S. (2004). Where did they go wrong? An analysis of the failure of knowledgeable internet. consumers to detect deception over the internet. *Group Decision and Negotiation, 13*, 149–172.

Grazioli, S., & Wang, A. (2001). Looking without seeing: Understanding unsophisticated consumers' success and failure to detect internet deception. In *ICIS 2001 proceedings* (p. 23).

Greis, N. P., Nogueira, M. L., & Kellogg, S. (2012). *The millennial cybersecurity project. Improving awareness of and modifying risky behavior in cyberspace* (Retrieved from Chapel Hill, USA).

Gross, R., & Acquisti, A. (2005). Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on privacy in the electronic society*.

Hann, I.-H., Hui, K.-L., Lee, T., & Png, I. (2002). Online information privacy: Measuring the cost-benefit trade-off. In *ICIS 2002 proceedings*, 1.

Harris, C. R., Coburn, N., Rohrer, D., & Pashler, H. E. (2013). Two failures to replicate high-performance-goal priming effects. *PloS One, 8*(8), e72467.

Harris, P. L., & Corriveau, K. H. (2011). Young children's selective trust in informants. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1567), 1179–1187.

Harris, P. L., Corriveau, K., Pasquini, E. S., Koenig, M., Fusaro, M., & Clément, F. (2012). Credulity and the development of selective trust in early childhood. In M. J. Beran, J. Brandl, J. Perner, & J. Proust (Eds.), *Foundations of metacognition* (p. 193). Oxford, UK: Oxford University Press.

Herley, C. (2009). So long, and no thanks for the externalities: The rational rejection of security advice by users. In *Proceedings new security paradigms workshop*.

Hong, J. (2012). The state of phishing attacks. *Communications of the ACM, 55*(1), 74–81. http://dx.doi.org/10.1145/2063176.2063197.

Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Communications of the ACM, 50*(10), 94–100. http://dx.doi.org/10.1145/1290958.1290968.

Jansson, K., & von Solms, R. (2011). Phishing for phishing awareness. *Behaviour & Information Technology, 32*(6), 584–593. http://dx.doi.org/10.1080/0144929X.2011.632650.

John, L. K., Acquisti, A., & Loewenstein, G. (2011). Strangers on a plane: Context-dependent willingness to divulge sensitive information. *Journal of consumer research, 37*(5), 858–873.

Joinson, A. N., Reips, U.-D., Buchanan, T., & Schofield, C. B. P. (2010). Privacy, trust, and self-disclosure online. *Human–Computer Interaction, 25*(1), 1–24.

Kahneman, D. (2012). A proposal to deal with questions about priming effects. *Nature, 490*.

Kearney, W. D., & Kruger, H. A. (2014). Considering the influence of human trust in practical social engineering exercises. Paper presented at the Information Security for South Africa (ISSA), 2014.

Kenrick, D. T., Neuberg, S. L., & Cialdini, R. B. (2005). *Social psychology: Unraveling the mystery*. New Zealand: Pearson Education.

Kirlappos, I., & Sasse, M. A. (2012). Security education against phishing: A modest proposal for a major rethink. *Security & Privacy, IEEE, 10*(2), 24–32.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature, 435*(7042), 673–676.

Kritzinger, E., & von Solms, S. H. (2010). Cyber security for home users: A new way of protection through awareness enforcement. *Computers & Security, 29*(8), 840–847. http://dx.doi.org/10.1016/j.cose.2010.08.001.

Kritzinger, E., & von Solms, S. H. (2013). Home user security-from thick security-oriented home users to thin security-oriented home users. Paper presented at the Science and Information Conference (SAI), 2013.

Krol, K., Moroz, M., & Sasse, M. A. (2012, 10–12 Oct. 2012). Don't work. Can't work? Why it's time to rethink security warnings. Paper presented at the Risk and Security of Internet and Systems (CRiSIS), 2012 7th International Conference on.

Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M. A., et al. (2009). School of phish: A real-world evaluation of anti-phishing training. In *SOUPS 2009-Proceedings of the 5th symposium on usable privacy and security*.

Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007). Protecting people from phishing: The design and evaluation of an embedded training email system. In *Conference on human factors in computing systems - Proceedings*.

Kumaraguru, P., Rhee, Y., Sheng, S., Hasan, S., Acquisti, A., Cranor, L. F., et al. (2007). Getting users to pay attention to anti-phishing education: Evaluation of retention and transfer. In *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit*.

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2008). Lessons from a real world evaluation of anti-phishing training. In *eCrime researchers summit, 2008*, 15–16 Oct. 2008.

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2010). Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology, 10*(2). http://dx.doi.org/10.1145/1754393.1754396.

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current directions in psychological science, 19*(3), 143–148. http://dx.doi.org/10.1177/0963721410370295.

Leon, P. G., Cranshaw, J., Cranor, L. F., Graves, J., Hastak, M., Ur, B., et al. (2012). What do online behavioral advertising privacy disclosures communicate to users?. In *Proceedings of the 2012 ACM workshop on Privacy in the electronic society* (Raleigh, North Carolina, USA).

Mayhorn, C. B., Murphy-Hill, E., Zielinska, O. A., & Welk, A. K. (2015). The social engineering behind phishing. *The next wave, 21*, 24–31.

May, C. P., Kane, M. J., & Hasher, L. (1995). Determinants of negative priming. *Psychological bulletin, 118*(1), 35.

Mesch, G. S. (2012). Is online trust and trust in social institutions associated with online disclosure of identifiable information online? *Computers in Human Behavior, 28*(4), 1471–1477.

Mills, C. M. (2013). Knowing when to doubt: Developing a critical stance when learning from others. *Developmental Psychology, 49*(3), 404–418. http://dx.doi.org/10.1037/a0029500.

Montoya, L., Junger, M., & Hartel, P. (2013). How 'Digital' is traditional crime? *European Intelligence and Security Informatics Conference (EISIC), 2013*, 31–37.

Retrieved from http://ieeexplore.ieee.org/search/searchresult.jsp?newsearch=true&queryText=how+digital+is+traditional+crime%2C+montoya&x=-1280&y=-331.

Morgan, T. J. H., & Laland, K. N. (2012). The biological bases of conformity. *Frontiers in Neuroscience, 6*, 87. http://dx.doi.org/10.3389/fnins.2012.00087.

Mouton, F., Leenen, L., Malan, M. M., & Venter, H. (2014). Towards an ontological model defining the social engineering domain. In *ICT and society* (pp. 266–279). Springer.

National initiative for cybersecurity education (NICE). (2009). *NICE: Creating a cybersecurity workforce and aware public*. Retrieved from United States Department of Commerce (DoC), National Institute of Standards and Technology (NIST) http://csrc.nist.gov/nice/documents/nicestratplan/nice-strategic-plan_sep2012.pdf.

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2008). Normative social influence is underdetected. *Personality and social psychology bulletin, 34*(7), 913–923.

Olson, J. S., Grudin, J., & Horvitz, E. (2005). A study of preferences for sharing and privacy. In *CHI'05 extended abstracts on human factors in computing systems*.

Omarzu, J. (2000). A disclosure decision model: Determining how and when individuals will self-disclose. *Personality and Social Psychology Review, 4*(2), 174–185.

Ostrom, E. (1998). A behavioral approach to the rational choice theory of collective action: Presidential address, American political science association, 1997. *American Political Science Review, 92*(01), 1–22.

Parsons, K., McCormac, A., Pattinson, M., Butavicius, M., & Jerram, C. (2015). The design of phishing studies: Challenges for researchers. *Computers and Security, 52*, 194–206. http://dx.doi.org/10.1016/j.cose.2015.02.008.

Pashler, H. E. (1998). Attentional limitations in dual task performance. In H. E. Pashler, & J. C. Johnston (Eds.), *Attention* (pp. 155–190). Hove, UK: Psychology Press.

Pashler, H. E., Coburn, N., & Harris, C. R. (2012). Priming of social distance? Failure to replicate effects on social and food judgments. *PloS One, 7*(8), e42510.

Pashler, H. E., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*(6), 528–530. http://dx.doi.org/10.1177/1745691612465253.

Paulsen, C., McDuffie, E., Newhouse, W., & Toth, P. (2012). NICE: Creating a cybersecurity workforce and aware public. *IEEE Security & Privacy, 10*(3), 0076–0079.

Pennebaker, J. W., & Seagal, J. D. (1999). Forming a story: The health benefits of narrative. *Journal of clinical psychology, 55*(10), 1243–1254.

Penner, L. A., Dovidio, J. F., Piliavin, J. A., & Schroeder, D. A. (2005). Prosocial behavior: Multilevel perspectives. *Annual Review of Psychology, 56*, 365–392.

Purkait, S. (2012). Phishing counter measures and their effectiveness - Literature review. *Information Management and Computer Security, 20*(5), 382–420. http://dx.doi.org/10.1108/09685221211286548.

Rocha Flores, W., Holm, H., Nohlberg, M., & Ekstedt, M. (2015). Investigating personal determinants of phishing and the effect of national culture. *Information & Computer Security, 23*(2), 178–199.

Rouse, S. V. (2015). A reliability analysis of Mechanical Turk data. *Computers in Human Behavior, 43*, 304–307. http://dx.doi.org/10.1016/j.chb.2014.11.004.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*(3), 393–404.

Sagarin, B. J., Cialdini, R. B., Rice, W. E., & Serna, S. B. (2002). Dispelling the illusion of invulnerability: The motivations and mechanisms of resistance to persuasion. *Journal of personality and social psychology, 83*(3), 526.

v. d. Schors, A., & v. d.Werf, M. (2014). *Jongeren & geld. De financiële situatie en hulpbehoefte van 12- tot en met 24-jarigen*. Retrieved from Den Haag, Nl http://www.nibud.nl/wp-content/uploads/Rapport-2014-Jongeren-en-geld.pdf.

Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., Cenac, Z., et al. (2013). Priming intelligent behavior: An elusive phenomenon. *PloS One, 8*(4), e56515.

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J. (2010). Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. In *Conference on human factors in computing systems - Proceedings*.

Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., et al. (2007). Anti-Phishing Phil: The design and evaluation of a game that teaches people not to fall for phish. In *Proceedings of the 3rd symposium on usable privacy and security* (pp. 88–99). New York, NY, USA: ACM.

Singer, E. (2004). Confidentiality, risk perception, and survey participation. *Chance, 17*(3), 30–34.

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: An interdisciplinary review. *MIS Quarterly, 35*(4), 989–1016.

Smith, R. G., & Hutchings, A. (2014). *Identity crime and misuse in Australia: Results of the 2013 online survey* (Retrieved from Canberra, Australia).

Sprecher, S., Treger, S., & Wondra, J. D. (2013). Effects of self-disclosure role on liking, closeness, and other impressions in get-acquainted interactions. *Journal of Social and Personal Relationships, 30*(4), 497–514.

Stajano, F., & Wilson, P. (2011). Understanding scam victims: Seven principles for systems security. *Communications of the ACM, 54*(3), 70–75.

Stewart, D. W., & Martin, I. M. (1994). Intended and unintended consequences of warning messages: A review and synthesis of empirical research. *Journal of Public Policy & Marketing*, 1–19.

Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science, 7*(6), 670–688.

Sundar, S. S., Kang, H., Wu, M., Go, E., & Zhang, B. (2013). Unlocking the privacy paradox: Do cognitive heuristics hold the key?. In *CHI'13 extended abstracts on human factors in computing systems*.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of personality and social psychology, 39*(5), 806.

Weinstein, N. D. (1987). Unrealistic optimism about susceptibility to health problems: Conclusions from a community-wide sample. *Journal of Behavioral Medicine, 10*(5), 481–500.

Weinstein, N. D., & Klein, W. M. (1995). Resistance of personal risk perceptions to debiasing interventions. *Health Psychology, 14*(2), 132.

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95–116).

West, R. (2008). The psychology of security. *Communications of the ACM, 51*(4), 34–40.

Wogalter, M. S., Laughery, K. R., & Mayhorn, C. B. (2012). Warnings and hazard communications. In *Handbook of human factors and ergonomics* (4 ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Wolff, J. (2016). Perverse Effects in Defense of Computer Systems: When More Is Less. Paper presented at the 2016 49th Hawaii International Conference on System Sciences, Hawaii, US.

Worthy, M., Gary, A. L., & Kahn, G. M. (1969). Self-disclosure as an exchange process. *Journal of personality and social psychology, 13*(1), 59–63. http://dx.doi.org/10.1037/h0027990.

Wueest, C. (2014). Targeted attacks against the energy sector. *Symantec Security Response* (Mountain View, CA).

Wu, M., Miller, R. C., & Garfinkel, S. L. (2006). Do security toolbars actually prevent phishing attacks?. In *Proceedings of the SIGCHI conference on human factors in computing systems, Montréal, Québec, Canada*.

Xiao, B., & Benbasat, I. (2015). Designing warning messages for detecting biased online product recommendations: An empirical investigation. *Information Systems Research, 26*(4), 793–811. http://dx.doi.org/10.1287/isre.2015.0592.

Yang, C. C., Tseng, S. S., Lee, T. J., Weng, J. F., & Chen, K. (2012). Building an anti-phishing game to enhance network security literacy learning. In *Proceedings of the 12th IEEE international conference on advanced learning technologies, ICALT 2012*.

Zhang, B., Wu, M., Kang, H., Go, E., & Sundar, S. S. (2014). Effects of security warnings and instant gratification cues on attitudes toward mobile websites. In *Proceedings of the 32nd annual ACM conference on human factors in computing systems*.

Zhang, B., & Xu, H. (2016). Privacy nudges for mobile applications: Effects on the creepiness emotion and privacy attitudes. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. San Francisco: California, USA.