Think before you click: The effects of systematic processing on phishing susceptibility

Meredith Lillie

This report is submitted in partial fulfilment

of the degree

of Master of Psychology (Organisational and Human Factors)

School of Psychology

University of Adelaide

October 2017

Word count: 12,614

Table of Contents

Declaration

This report contains no material which has been accepted for the award of any other degree or diploma in any University, and, to the best of my knowledge, this report contains no materials previously published except where due reference is made.

Meredith Ellen Lillie

October 2017

Acknowledgements

I am enormously grateful for the support and guidance of several people throughout this research project. To my supervisors Katie Parsons and Marcus Butavicius, I cannot thank you enough for your expertise, your diligence and your approachability. I've learned so much from you both. I would also like to thank the other members of the Human Aspects of Cyber Security (HACS) team: Dragana Calic and Agata McCormac (from DST Group), and Malcolm Pattinson (from The University of Adelaide). Your encouragement and humour has made all the difference. I can only to hope to work again in such a passionate and collaborative team. Thank you to Neil Kirby for your assistance and attentiveness as Masters of Psychology (OHF) Program Coordinator, and to Aspa Sarris for your advice on conducting a thesis in the first year of the program. Lastly, it's hard to find the words to express my gratitude for my family. I could never have come this far without you, and I'm honoured to know I have made you proud.

List of Tables

Literature Review

Research Report

List of Figures

Literature Review

Word count: 4,723

Abstract

Researchers have identified the use of social influence in phishing emails and have found greater cognitive impulsivity to predict phishing susceptibility. These findings suggest that relying on predominantly heuristic (rather than systematic) information processing strategies when managing emails could be a key contributor to users' susceptibility. Accordingly, it is proposed that the effects of systematic processing on phishing susceptibility should be investigated. Specifically, research should determine whether manipulating systematic processing affects users' judgements of the legitimacy of phishing and genuine emails. The outcomes of this research would have potential implications for cyber security training.

Introduction

Organisations suffer both direct and indirect costs associated with cyber security incidents. Between 2014 and 2015, Australian organisations experienced the second highest financial loss worldwide at $3.27 million (PricewaterhouseCoopers [PWC] Australia, 2015). Other costs indicated by Australian organisations include loss of intellectual property, reputational loss, corrupted data, productivity loss, distrust from customers or partners, loss of customers, lawsuits and psychological stress to employees (Telstra Corporation, 2017).

Despite the implementation of technical safeguards to defend against cyber-attacks, an organisation's information security systems can be compromised by a single click in response to a phishing email. Cyber criminals not only see the organisation as a single entity with safeguards to overcome, but also as collections of individuals with psychological vulnerabilities to exploit (International Business Machines Corporation [IBM] Global Technology Services, 2014, 2014). As a form of *social engineering*, phishing emails use deception to exploit these psychological vulnerabilities (Muscanell, Guadagno, & Murphy, 2014). Phishers typically use the identities of well-known and trusted companies. Subsequently, email users are deceived into clicking on an embedded link or opening an email attachment before providing the phisher with confidential information (e.g., passwords) or access to computer systems through the inadvertent installation of malware (Butavicius, Parsons, Pattinson, & McCormac, 2015). Thus, by targeting the user, phishing emails can successfully compromise organisations' information security systems. Consistent with this, human error has been reported as a factor in 95 percent of cyber security incidents, wherein clicking on an infected attachment or unsafe URL was the most prevalent contributing human factor (IBM Global Technology Services, 2014). In 2016, approximately one third of Australian organisations experienced phishing attacks on a weekly or monthly basis, making them the most frequently occurring cyber security incident affecting Australian organisations (Telstra Corporation, 2017).

While technical safeguards are often implemented to defend against the threat of phishing (Purkait, 2012), these are not guaranteed solutions. For instance, although email filters can be successful in preventing phishing attacks from reaching the user’s inbox, these filters rely on updates in the wake of new attacks to maintain their effectiveness. Hence, there is a window of vulnerability in the time between when new attacks are instigated and email filters are updated to defend against them. This is especially concerning given that 255,065 unique phishing attacks worldwide were recorded in the year 2016 alone (Anti-Phishing Working Group [APWG], 2017). Exposure to phishing attacks therefore remains a possibility for almost all email users. For this reason, it is vitally important to recognise and understand the human element of susceptibility to phishing. In other words, researchers must seek to understand why users fall victim to phishing attacks and, in turn, how to reduce the phishing susceptibility of users and their organisations.

Methodologies for Studying Phishing Susceptibility

The methodologies used by phishing researchers can generally be categorised as either ‘real phishing’, phishing ‘IQ test’ or ‘scenario-based’(also known as ‘role-play’), and all three have both advantages and disadvantages for the study of phishing susceptibility. In real phishing studies, simulated phishing attacks are sent directly to participants’ normal email inboxes, and participants are not aware that their responses (or lack of responses) to these simulations are recorded by researchers. These studies provide a useful indication of the real-world response rates of members of specific institutions (e.g., students of a university, employees of an organisation) for phishing emails that can be either generic or targeted to members of that institution (Goel, Williams, & Dincelli, 2017; Jagatic, Johnson, Jakobsson,

& Menczer, 2007; Ferguson, 2005; Rocha Flores, Holm, Nohlberg, & Ekstedt, 2015). They can also provide the opportunity to educate participants who respond to the simulated attacks.

On the other hand, real phishing studies have been criticised for a number of ethical issues, including lack of informed consent for research participants, deception and the risk of negative reactions by participants (Finn & Jakobsson, 2007; Jagatic, Johnson, Jakobsson, & Menczer, 2007). Another issue specific to conducting real phishing studies in organisations is the risk of creating a negative security culture in which employees may perceive themselves to be under an undue level of scrutiny and may fear punishment as a consequence of being ‘tricked’ by their own employers. Real phishing studies are also limited in what they can tell researchers about user susceptibility because they do not observe participants’ responses to genuine emails. They can observe correct phishing judgements for phishing emails, but not incorrect phishing judgements for genuine emails. This prevents real phishing studies from being able to examine users’ ability to discriminate between the two email types.

Unlike the real phishing methodology, studies using the phishing IQ test or scenario-based methodologies require participants’ informed consent and observe participants’ responses to both phishing and genuine emails (Parsons, McCormac, Pattinson, Butavicius, & Jerram, 2013). In this way, phishing researchers can apply the performance measurement approach known as Signal Detection Theory (SDT; Green & Swets, 1966). This approach can be applied to any circumstance where two possible stimulus types must be discriminated from one another (Stanislaw & Todorov, 1999). SDT provides two performance measures. In the context of phishing, *discrimination* measures how well individuals can distinguish between genuine and phishing emails, and *bias* measures the overall tendency to judge emails as either ‘genuine’ or ‘phishing’. This approach reconceptualises the problem of phishing susceptibility as one of decision-making under uncertainty. Users make ‘phishing’ or ‘genuine’ judgements not only for phishing emails, but genuine emails as well. The SDT approach acknowledges that users make these judgements, correctly or incorrectly, for both email types. Therefore, phishing studies that do not accommodate the SDT approach are only capable of telling half of the story.

Despite their similarities, there is an important difference between the phishing IQ test and scenario-based methodologies. Unlike scenario-based studies, participants of phishing IQ test studies are aware that their ability to detect phishing emails is being measured. Accordingly, these participants are primed for signs of phishing and have been shown to exhibit the subject expectancy effect (Anandpara, Dingman, Jakobsson, Liu, & Roinestad, 2007; Parsons, McCormac, Pattinson, Butavicius, & Jerram, 2013). In particular, Anandpara, Dingman, Jakobsson, Liu, and Roinestad (2007) found evidence to suggest that these participants are more are more biased towards making 'phishing' judgements. This was speculated to be because they are more suspicious of the legitimacy of emails compared to everyday email use. Conversely, researchers who use the scenario-based methodology incorporate a role play paradigm into their studies. Participants are not informed they are participating in a phishing study. Rather, they are instructed to assume the role of a fictitious person and to make judgements on emails received by that person.

Parsons McCormac, Pattinson, Butavicius, and Jerram (2013) conducted a phishing study to validate the use of scenarios as a method for overcoming the subject expectancy effect. All participants were aware that they were participating in a study about email management, but only half were informed that they were participating in a study specifically about phishing. Rather than resulting in a bias towards 'phishing' judgements as indicated by Anandpara et al. (2007), priming participants with the notion of phishing was found to improve their ability to discriminate between genuine and phishing emails. The authors speculated that these participants engaged in more diligent decision-making, and this is supported by the finding that they took significantly longer to complete the experiment than participants who were not primed for signs of phishing.

While the phishing IQ test methodology can be a more direct way to conduct early evaluations of interventions (Robila & Ragucci, 2006) and investigate the cues relied upon by users to detect phishing (Furnell, 2007), the findings of these studies cannot be generalised to actual user behaviour because of the subject expectancy effect. The phishing IQ test methodology is therefore criticised for its lack of real-world validity. For this reason, many researchers prefer the scenario-based methodology, as it enables phishing susceptibility to be studied in a manner that is still relatively high in real-word validity. Nevertheless, the scenario-based methodology still suffers from a number of limitations. These limitations stem from the challenge of eliciting meaningful responses without alerting participants to the true purpose of the study. For example, researchers must try to ensure that the inexplicit response options provided can be meaningfully encoded as judgements of phishing or genuine (e.g., Parsons et al. (2013) used response options such as ‘delete the email’). In order to prevent participants from responding according to judgements of relevance rather than legitimacy, participants can be instructed to assume that the emails are relevant to the fictitious person.

Given the different advantages and disadvantages of the three phishing study methodologies, researchers must carefully consider which methodology is most appropriate for their research aims and attempt to minimise its disadvantages as much as possible.

Cues Relied Upon by Users

Phishers typically use company identities when constructing phishing emails, and are capable of fabricating a visual presentation that corresponds with this identity (e.g., logo, design). The intended effect is for the user’s trust in this company to be instilled in the email itself, so that the user does not question its legitimacy. Consistent with this, research has shown that users’ susceptibility is influenced by an inherent trust in the company that appears to be the email sender (Egelman, Cranor, & Hong, 2008; Parsons et al., 2015). Furthermore, a range of studies (Jagatic, Johnson, Jakobssen, & Menczer, 2007; Karakasiliotis, Furnell, & Papadaki, 2006; Parsons et al., 2015) have indicated that many users underestimate the extent to which email components can be manipulated by a phisher to appear legitimate.

Parsons et al. (2015) developed a comprehensive list of cues that have been identified as important (either by users or researchers) for discriminating between phishing and genuine emails and websites. An analysis of expert ratings on the presence of each cue in a set of 50 emails (half genuine, half phishing) resulted in the identification of five effective cues for discrimination; genuine emails were more likely to contain a legitimate URL (shown when the cursor is hovered over an embedded link), a sender's address that appeared legitimate, message consistency and personalisation, whereas phishing emails were more likely to contain spelling and grammatical errors. A further analysis of the relationship between the presence of each cue and participants' legitimacy judgements indicated that the only effective cues influencing these judgements were personalisation and spelling and grammatical errors. Participants were also influenced by five ineffective cues; visual presentation, copyright information or legal disclaimers, importance, urgency and positive consequences. Although a limitation of the study is that it cannot be determined how the cues interacted with one another to inform legitimacy judgements, it nevertheless provides insight into why some users are deceived by phishing emails. Other researchers have reported that even when users do attend to the URL as an effective cue, they do not always correctly judge its legitimacy (Egelman, Cranor, & Hong, 2008).

It is important to note that the effective cues for discrimination and those cues relied upon by users may change over time (Parsons et al., 2015). In particular, phishers are already capable of spoofing (i.e., forging) the sender's address of an email so that, for example, a phishing email using the identity of *PayPal* can appear to have been sent from the address ‘services@paypal.com’ (Furnell, 2007). Hence, it is possible for enough phishers to begin spoofing the sender’s address so that its appearance of legitimacy is no longer an effective cue. This is in addition to the increasingly common practice of spear-phishing, where phishers research their victims and personalise the email in order to increase the likelihood of responding. Currently, the only infallible cue for discrimination is the legitimacy of the URL. Even so, phishers can create an illegitimate URL that has the resemblance of a legitimate one (e.g., “www.paypa1.com/au/signin”, where the number “1” is substituted for the letter “l”).

Social Influence in Phishing

Many of the phishing cues mentioned above allude to a specific type of social engineering utilised by phishers. *Social influence* refers to overt forces occurring in our interactions with others that have the power to cause a change in our attitudes or behaviours (Cialdini & Goldstein, 2004; Muscanell, Guadagno, & Murphy, 2014). In the context of face-to-face interactions, Cialdini (2009) established six social influence techniques that serve as heuristics for decision-making: authority, scarcity, consistency, reciprocation, liking and social proof. When exploited by social engineers, these techniques are experienced as external social pressure to agree or comply with a request. In this way, social influence is used to persuade individuals and gain their compliance. Notably, social influence attempts are increasingly occurring in online contexts, and researchers have identified the use of all six social influence techniques by phishers (Akbar, 2014; Ferreira & Lenzini, 2015). Table 1 provides a description of the six social influence techniques and examples from phishing research.

The effectiveness of a given social influence technique can vary depending on the communication modality (Chaiken & Eagly, 1983), and phishing emails differ from face-to-face persuasion contexts in important ways.

Technique	Description[a]	Examples from Phishing Research
Authority	We are inclined to obey people in positions of authority.	Copyright information and/or legal disclaimers (Parsons et al., 2015) Authoritative language (Butavicius, Parsons, Pattinson, & McCormac, 2015) The sender is a person or institution of authority (Butavicius, Parsons, Pattinson, & McCormac, 2015; Wright, Jensen, Thatcher, Dinger, & Marett, 2014)
Scarcity	Opportunities seem more valuable to us when their availability is limited.	Urgent language (Parsons et al., 2015; Wang et al., 2012) Reference to an offer that is limited by a deadline and/or a restricted number of participants (Butavicius, Parsons, Pattinson, & McCormac, 2015; Wright, Jensen, Thatcher, Dinger, & Marett, 2014)
Consistency	We desire to be (and to appear) consistent with our previous actions and commitments.	Reference to a user's action(s) and/or commitment(s) (Wright, Jensen, Thatcher, Dinger, & Marett, 2014)
Reciprocation	We try to repay, in kind, what another person has provided us.	Reference to action(s) by the sender (Wright, Jensen, Thatcher, Dinger, & Marett, 2014)
Liking	We prefer to say yes to those we know and like.	Using the identity of a well-known and trusted company as the sender (Parsons et al., 2015) Humour (Wright, Jensen, Thatcher, Dinger, & Marett, 2014)
Social Proof	We are more likely to perform the actions that we see others performing	Reference to figures indicating a large number of people have already responded (Butavicius, Parsons, Pattinson, & McCormac, 2015; Wright, Jensen, Thatcher, Dinger, & Marett, 2014)

Table 1

Cialdini's (2009) social influence techniques and examples from phishing research.

[a] Cialdini (2009)

Phishing emails are usually non-interactive, meaning the social engineer has only a single opportunity per phishing email to persuade the user to respond (Hong, 2012). Additionally, the use of a text-based mediated channel may undermine the phisher's persuasion attempt by enabling the user to re-read, and hence reprocess the message, which has been shown to facilitate the discovery of deception (George, Carlson, & Valacich, 2013).

Accordingly, researchers have investigated whether some techniques are more persuasive in phishing emails than others (Butavicius, Parsons, Pattinson, & McCormac, 2015; Wright, Jensen, Thatcher, Dinger, & Marett, 2014). Wright, Jensen, Thatcher, Dinger, and Marett (2014) found that liking had the largest positive effect on likelihood to respond, followed by scarcity, social proof and reciprocity, whereas consistency was ineffective and authority had an unexpected negative effect. Butavicius, Parsons, Pattinson, and McCormac (2015) investigated the persuasiveness of authority, scarcity and social proof. In contrast to Wright et al., authority was found to be the most effective social influence technique. Unexpectedly, the absence of social influence in a phishing email was found to be more effective than any of the techniques examined. Both Wright et al. and Butavicius et al. speculated that their unexpected results could be attributed to users' increasing familiarity with social influence attempts in generic phishing emails, and hence their development of resistance to such attempts.

Heuristics and Information Processing

Research has found that higher levels of cognitive impulsivity may increase users' susceptibility to phishing (Butavicius, Parsons, Pattinson, & McCormac, 2015; Parsons, McCormac, Pattinson, Butavicius, & Jerram, 2013; Welk et al., 2015). This finding is relevant to the use of social influence by phishers because cognitive impulsivity can affect how a persuasive message is processed by the receiver. Cognitive impulsivity refers to a person's tendency to make decisions with little or no conscious thought. People who have low levels of cognitive impulsivity are more likely to monitor and correct their cognitive impulses (Primi, Morsanyi, Chiesi, Donati, & Hamilton, 2016). Although two other phishing studies measured cognitive impulsivity and did not find this relationship, this is likely because Kumaraguru et al. (2007) incorporated only three phishing emails and a small participant sample, and Mayhorn and Nyeste (2012) provided feedback to participants on their ability to detect phishing during a training intervention.

All of these studies except for Welk et al. (2015) used the Cognitive Reflection Test (CRT; Frederick, 2005) to measure cognitive impulsivity. The CRT is a set of three problems, each designed to have an immediately intuitive but incorrect response. This response must be suppressed and overridden by more careful analytic reasoning in order to yield the correct answer. For example, the following is the first problem: “A bat and a ball together cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost?” (Frederick, 2005, p. 27). Initially, there is a strong tendency for respondents to consider the answer of “10 cents”, even by those who then use analytic reasoning to give the correct answer of “5 cents”. Butavicius et al. (2015) suggested that the reason cognitive impulsivity, and the CRT in particular, is associated with phishing susceptibility is that it may indicate whether users tend to rely on heuristic cues to make judgements of email legitimacy. Social influence takes advantage of our tendency to rely on heuristic cues. Each technique has the “ability to produce a distinct kind of automatic, mindless compliance from people, that is, a willingness to say yes without thinking first” (Cialdini, 2009, p. xiv). In this way, Butavicius et al. refer to *dual-system information processing models* as one explanation for why some people are persuaded to respond to phishing emails.

Dual-system information processing models have been proposed to account for how individuals process persuasive messages, including the Elaboration Likelihood Model (ELM; Petty & Cacioppo, 1986) and the Heuristic-Systematic Model (HSM; Chen & Chaiken, 1999). According to both models, individuals rely on two separate strategies for processing persuasive messages; whereas *heuristic processing* is automatic, quick, effortless and intuitive, *systematic processing* is controlled, slow, effortful and analytic. While systematic processing generally yields better-quality judgement, heuristic processing is more economic given our limited cognitive resources and the immense volume of information that we must continually process. For this reason, heuristic processing is understood to be the default strategy for processing information. Nevertheless, the judgements formed on the basis of heuristic processing can be overridden by systematic processing.

When heuristic processing is used, only the superficially persuasive, i.e., heuristic, cues of a persuasive message inform judgement, such as the likeability of the message's source. This means that social influence is likely to be more effective for gaining compliance when the message is processed heuristically (Kaptein, Markopoulos, de Ruyter, & Aarts, 2015). Conversely, when systematic processing is used, judgement becomes informed by evidentiary cues, such as the reliability of the message's source. Hence, activating systematic processing over and above heuristic processing may be an effective strategy for resisting social influence attempts (including those occurring in phishing emails), as it changes the way the persuasive message is processed and potentially the individual's response (Sagarin & Cialdini, 2004). Thus, the use of systematic processing may reduce susceptibility to phishing.

In addition, research has indicated that individuals routinely rely on heuristic cues when processing information online (Guadagno & Cialdini, 2005; Guadagno, Muscanell, Rice, & Roberts, 2013; Metzger, Flanagin, & Medders, 2010). Muscanell, Guadagno, and Murphy (2014) argue this is because the online context imposes a high cognitive load on users; not only do users have access to a vast amount of information, they often multitask, making them more prone to relying on heuristic processing as a way of conserving cognitive resources. Similarly, Vishwanath, Herath, Chen, Wang, and Rao (2011) argue that users eventually learn to manage this high cognitive load by developing heuristic, habitual response patterns. This in turn is argued to make individuals more likely to inattentively respond to emails, and hence more susceptible to phishing. This was supported by research that found greater heuristic processing and email habit strength significantly increases phishing susceptibility, whereas greater systematic processing significantly reduces it (Vishwanath, 2015; Vishwanath, Harrison, & Ng, 2016).

Wang, Herath, Chen, Vishwanath, and Rao (2012) did not find that expending greater cognitive effort (i.e., systematic processing) whilst processing a phishing email reduced the likelihood of responding. However, Wang et al. used a single email that incorporated several effective indicators of phishing, including multiple grammatical errors and an illegitimate sender's address (Parsons et al., 2015). While grammatical errors are an important illegitimacy cue relied upon by users, many phishing emails do not contain these errors and, as mentioned previously, many contain a spoofed sender's address. The high availability of phishing cues may have meant that participants did not expend a great deal of cognitive effort to reach a phishing decision. Therefore, it would be interesting to investigate the effects of manipulating systematic processing on users' legitimacy judgements for both phishing and genuine emails where multiple, fallible cues are not present. Nevertheless, Wang et al.'s finding that attention to a social influence technique increased the likelihood to respond provides further evidence to support dual-system information processing as an explanation of phishing susceptibility.

Phishing researchers have tended to prefer the HSM to account for the relationship between information processing strategy and phishing susceptibility (Goel, Williams, and Dincelli, 2017; Luo, Zhang, Burd, & Seazzu, 2013; Vishwanath, 2015). The HSM incorporates the concept of the *sufficiency threshold*; the level of confidence an individual desires to reach before they will consider their judgement to be 'good enough' and so discontinue processing of the message. The extent of information processing required to reach the sufficiency threshold determines whether systematic processing will be activated over and above initial heuristic processing (Chen & Chaiken, 1999).

Luo, Zhang, Burd, and Seazzu (2013) suggested that the success of a phishing attack depends on whether the phisher's message can either increase the recipients' reliance on heuristic processing (through embedding heuristic cues), lower the sufficiency threshold to prevent the activation of systematic processing, or else withstand the scrutiny of at least minimal systematic processing. Similarly, Goel, Williams, and Dincelli (2017) suggested that contextualising an email message so that it is more likely to be perceived by the user as personally relevant (e.g., spear-phishing), works to lower the sufficiency threshold (thus preventing systematic processing activation) and also to help the email to withstand scrutiny should heuristic processing give way to more systematic processing.

In summary, evidence from phishing research supports dual-system information processing as an explanation for why some users respond to phishing emails. This has important implications. It suggests that increasing users' reliance on systematic processing when managing emails could reduce their susceptibility to phishing. A review of the literature determined that the experimental effects of manipulating users' information processing strategy (i.e., heuristic vs. systematic processing) on their judgements of the legitimacy of phishing and genuine emails is yet to be investigated.

Manipulations of Information Processing Strategy

Research has shown that individuals' information processing strategy can be manipulated, at least in the short term. The only study identified to have manipulated information processing strategy in the context of emails was conducted by Yan and Gozu (2012). In a repeated measures design, participants were instructed to read the email content of 36 spam emails either 'quickly and casually' (heuristic processing strategy) or 'slowly and carefully' (systematic processing strategy). Participants' decisions were recorded as correct only if they decided to delete the spam email rather than open or respond to it. It was found that participants were more likely to delete the emails when relying on systematic processing. While the results are consistent with a dual-system information processing explanation for phishing susceptibility, the study did not examine the manipulation effects on genuine emails. Therefore, it could not determine whether the information processing manipulation merely produced a greater bias towards illegitimacy judgements for both spam and genuine emails, rather than an improved ability to discriminate between the two email types. This prevented the study from providing a comprehensive examination of the manipulation effects. Also, the study did not incorporate a role-play to encourage responses that reflect judgements of legitimacy rather than personal relevance.

Several authors have experimented with the effects of information processing manipulations outside of the email context. The CRT is commonly used as an outcome variable to demonstrate these effects. Alter, Oppenheimer, Epley, and Eyre (2007) found that when the CRT is presented in a difficult-to-read font rather than an easy-to-read font, individuals assume that greater cognitive effort is required to complete the task and are subsequently more likely to provide correct responses (consistent with a systematic processing strategy). Attridge and Inglis (2015) examined the effects of completing one or more problems from the Raven's Advanced Progressive Matrices (RAPM; Raven, Raven, & Court, 1998). Respondents are presented with an incomplete matrix of visual designs, and they are asked to identify the missing design that completes the pattern from the response options provided. Merely completing a single RAPM problem was found to significantly improve participants' CRT performance. Hauser and Schwarz (2015) investigated the effects of an instructional manipulation check (Oppenheimer, Meyvis, & Davidenko, 2009). Instructional manipulation checks present a lure question (e.g., "Which of these activities do you engage in regularly? (click on all that apply).") with a smaller block of text that instructs respondents to ignore the lure responses (e.g., "running") and instead click "other" and enter "I read the instructions" in the corresponding text box. Participants who followed the instructional manipulation check instructions subsequently scored higher on the CRT.

In addition to being an outcome variable, the CRT itself has been used as an information processing manipulation. Pinillos, Smith, Nair, Marchetto, and Mun (2011) found that answering one or more of the three CRT problems correctly subsequently produced a pattern of philosophical judgements that had greater similarity to those of philosophers. The authors argued that systematic processing activation was the effect of experiencing the realisation that one's first response was incorrect, hence prompting more careful reasoning.

These findings support the possibility that users' information processing strategy can be manipulated in a way that might affect their judgements of emails. The effects of such a manipulation could be similar to the findings observed by Parsons et al. (2013), where priming participants with the notion of phishing was found to improve their ability to discriminate between genuine emails and phishing attacks.

Conclusions and Future Directions

Exposure to phishing attacks is a possibility for almost all email users. It is thus important to recognise the human element of phishing susceptibility and seek to understand why users fall victim to phishing emails. Researchers have identified the use of social influence in phishing emails and have found greater cognitive impulsivity to predict phishing susceptibility. These findings suggest that a reliance on predominantly heuristic (rather than systematic) information processing strategies when managing emails could be a key contributor to users' susceptibility. Research has shown that individuals' information processing strategy can be manipulated, at least in the short term, including in the context of emails (Yan & Gozu, 2012). However, the experimental effects of manipulating users' information processing strategy on their judgements of emails is yet to be investigated.

Of the three main methodologies used in phishing research, the scenario-based methodology would be the most appropriate for this investigation, because responses to both phishing and genuine emails, assumed to be relevant to the recipient, can be observed.

If it is the case that systematic processing activation reduces phishing susceptibility, then this result would have two important implications. First, it would provide stronger evidence for dual-system information processing as an explanation for why some users respond to phishing attacks. Second, it would provide a new focus for user training against these attacks. Rather than merely warning users about the threat posed by phishing or even providing instructions for recognising the attacks, users could be trained to activate systematic processing as an effective strategy for defending against the threat of phishing.

References

Akbar, N. (2014). *Analysing persuasion principles in phishing emails.* (Masters thesis), University of Twente.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*(4), 569-576. doi:10.1037/0096-3445.136.4.569

Anandpara, V., Dingman, A., Jakobsson, M., Liu, D., & Roinestad, H. (2007). *Phishing IQ tests measure fear, not ability.* Paper presented at the 11th International Conference on Financial Cryptography and Data Security, Scarborough, Trinidad and Tobago.

Anti-Phishing Working Group [APWG]. (2017). Global phishing survey: Trends and domain name use in 2016. http://docs.apwg.org/reports/APWG_Global_Phishing_Report_2015-2016.pdf

Attridge, N., & Inglis, M. (2015). Increasing cognitive inhibition with a difficult prior task: Implications for mathematical thinking. *ZDM Mathematics Education, 47*(5), 723-734. doi:http://dx.doi.org/10.1007/s11858-014-0656-1

Butavicius, M., Parsons, K., Pattinson, M., & McCormac, A. (2015). *Breaching the human firewall: Social engineering in phishing and spear-phishing emails.* Paper presented at the 26th Australasian Conference on Information Systems, Adelaide, Australia.

Chaiken, S., & Eagly, A. H. (1983). Communication modality as a determinant of persuasion: The role of communicator salience. *Journal of Personality and Social Psychology, 45*(2), 241-256. doi:10.1037/0022-3514.45.2.241

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73-96). New York, NY: Guilford Press.

Cialdini, R. B. (2009). *Influence: Science and practice* (5th ed.). Boston, MA: Pearson Education, Inc.

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

Egelman, S., Cranor, L. F., & Hong, J. (2008). You've been warned: An empirical study of the effectiveness of web browser phishing warnings. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1065-1074.

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

Ferreira, A., & Lenzini, G. (2015). An analysis of social engineering principles in effective phishing. *2015 Workshop on Socio-Technical Aspects in Security and Trust*, 9-16. doi:10.1109/STAST.2015.10

Finn, P., & Jakobsson, M. (2007). Designing and conducting phishing experiments. *IEEE Technology and Society Magazine, Special Issue on Usability and Security, 26*(1), 46-58.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19*(4), 25-42.

Furnell, S. (2007). Phishing: Can we spot the signs? *Computer Fraud & Security, 2007*(3), 10-15. doi:10.1016/S1361-3723(07)70035-0

George, J. F., Carlson, J. R., & Valacich, J. S. (2013). Media selection as a strategic component of communication. *MIS Quarterly, 37*(4).

Goel, S., Williams, K., & Dincelli, E. (2017). Got phished? Internet security and human vulnerability. *Journal of the Association for Information Systems, 18*(1), 22-44.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Guadagno, R. E., & Cialdini, R. B. (2005). Online persuasion and compliance: Social influence on the internet and beyond. In Y. Amichai-Hamburger (Ed.), *The social net: The social psychology of the internet* (pp. 91-113). Oxford, UK: Oxford University Press.

Guadagno, R. E., Muscanell, N. L., Rice, L. M., & Roberts, N. (2013). Social influence online: The impact of social validation and likability on compliance. *Psychology of Popular Media Culture, 2*(1), 51-60. doi:10.1037/a0030592

Hauser, D. J., & Schwarz, N. (2015). It's a trap! Instructional manipulation checks prompt systematic thinking on "tricky" tasks. *Sage Open, 5*(2), 1-6. doi:10.1177/ 2158244015584617

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

International Business Machines Corporation [IBM] Global Technology Services. (2014). *IBM Security Services 2014 cyber security intelligence index: Analysis of cyber attack and incident data from IBM's worldwide security operations*. Retreived from http://www.ibm.com/developerworks/library/se-cyberindex2014/index.html.

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

Kaptein, M., Markopoulos, P., de Ruyter, B., & Aarts, E. (2015). Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. *International Journal of Human-Computer Studies, 77*, 38-51. doi:10.1016/ j.ijhcs.2015.01.004

Karakasiliotis, A., Furnell, S., & Papadaki, M. (2006). Assessing end-user awareness of social engineering and phishing. *Proceedings of the 7th Australian Information Warfare and Security Conference*. doi:10.4225/75/57a80e47aa0cb

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. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 905-914.

Luo, X. R., Zhang, W., Burd, S., & Seazzu, A. (2013). Investigating phishing victimization with the Heuristic–Systematic Model: A theoretical framework and an exploration. *Computers & Security, 38*, 28-38.

Mayhorn, C. B., & Nyeste, P. G. (2012). Training users to counteract phishing. *Work, 41*, 3549-3552.

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication, 60*(3), 413-439.

Muscanell, N. L., Guadagno, R. E., & Murphy, S. (2014). Weapons of influence misused: A social influence analysis of why people fall prey to internet scams. *Social and Personality Psychology Compass, 8*(7), 388-396.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*(4), 867-872. doi:10.1016/j.jesp.2009.03.009

Parsons, K., Butavicius, M., Pattinson, M., Calic, D., McCormac, A., & Jerram, C. (2015). *Do users focus on the correct cues to differentiate between phishing and genuine emails?* Paper presented at the Australasian Conference on Information Systems, Adelaide, Australia.

Parsons, K., McCormac, A., Pattinson, M., Butavicius, M., & Jerram, C. (2013). *Phishing for the truth: A scenario-based experiment of users' behavioural response to emails.* Paper presented at the IFIP International Information Security Conference, Auckland, New Zealand.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology, 19,* 123-205.

Pinillos, N. Á., Smith, N., Nair, G. S., Marchetto, P., & Mun, C. (2011). Philosophy's new challenge: Experiments and intentional action. *Mind & Language, 26*(1), 115-139.

PricewaterhouseCoopers [PWC] Australia. (2015). *Australia tops Asian region for cyber security risks*: Retrieved from http://www.pwc.com.au/press-room/2015/cyber-security-risks-oct15.html

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *Journal of Behavioral Decision Making, 29*(5), 453-469. doi:10.1002/bdm.1883

Purkait, S. (2012). Phishing counter measures and their effectiveness – literature review. *Information Management & Computer Security, 20*(5), 382-420.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. San Antonio, TX: Harcourt Assessment.

Robila, S. A., & Ragucci, J. W. (2006). *Don't be a phish: Steps in user education.* Paper presented at the ACM Conference on Innovation and Technology in Computer Science Education, Bologna, Italy.

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. doi:10.1108/ICS-05-2014-0029

Sagarin, B. J., & Cialdini, R. B. (2004). Creating critical consumers: Motivating receptivity by teaching resistance. In E. S. Knowles & J. A. Linn (Eds.), *Resistance and persuasion* (pp. 259-282). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149.

Telstra Corporation. (2017). *Telstra Cyber Security Report 2017*. Retrieved from https://www.telstraglobal.com/images/assets/insights/resources/Telstra_Cyber_Security_Report_2017_-_Whitepaper.pdf

Vishwanath, A. (2015). Examining the distinct antecedents of e-mail habits and its influence on the outcomes of a phishing attack. *Journal of Computer‐Mediated Communication, 20*(5), 570-584. doi:10.1111/jcc4.12126

Vishwanath, A., Harrison, B., & Ng, Y. J. (2016). Suspicion, cognition, and automaticity model of phishing susceptibility. *Communication Research*, 1-21. doi:http://dx.doi.org/10.1177/0093650215627483

Vishwanath, A., Herath, T., Chen, R., Wang, J., & Rao, H. R. (2011). Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems, 51*(3), 576-586.

Wang, J., Herath, T., Chen, R., Vishwanath, A., & Rao, H. R. (2012). Phishing susceptibility: An investigation into the processing of a targeted spear phishing email. *IEEE Transactions on Professional Communication, 55*(4), 345-362. doi:10.1109/TPC.2012.2208392

Welk, A. K., Hong, K. W., Zielinska, O. A., Tembe, R., Murphy-Hill, E., & Mayhorn, C. B. (2015). Will the “phisher-men” reel you in? Assessing individual differences in a phishing detection task. *International Journal of Cyber Behavior, Psychology and Learning, 5*(4), 1-17.

Wright, R. T., Jensen, M. L., Thatcher, J. B., Dinger, M., & Marett, K. (2014). Influence techniques in phishing attacks: An examination of vulnerability and resistance. *Information Systems Research, 25*(2), 385-400.

Yan, Z., & Gozu, H. Y. (2012). Online decision-making in receiving spam emails among college students. *International Journal of Cyber Behavior, Psychology and Learning, 2*(1), 1-12. doi:10.4018/ijcbpl.2012010101

Research Report

Word count: 7,891

Think before you click: The effects of systematic processing on phishing susceptibility

Meredith Lillie[1]

[1] Affiliation: The University of Adelaide

Postal Address: The University of Adelaide
SA 5005
AUSTRALIA

Email Address: [redacted]

Abstract

The present study investigated the effects of systematic processing on phishing susceptibility. A total of 1,037 participants were randomly allocated to one of four conditions. Participants either completed one of three information processing manipulations (IPMs) or did not complete an IPM before responding to email stimuli. Participants' scores on the IPMs were used as an indication of their information processing strategy (heuristic vs. systematic) when responding to email stimuli. Participants who completed an IPM and attained a high score were better able to discriminate between phishing and genuine emails. Of the three IPMs, discrimination performance was only found to be improved by attaining higher scores on the Matrix Reasoning Task. The outcomes of this research have implications for cyber security training. Directions for future research on phishing susceptibility are discussed.

Keywords: phishing susceptibility, systematic processing, information processing, cyber security, human-computer interaction

1. Introduction

Despite the implementation of technical safeguards to defend against cyber-attacks, an organisation's information security systems can be compromised by a single click in response to a phishing email. Cyber criminals not only see the organisation as a single entity with safeguards to overcome, but also as collections of individuals with psychological vulnerabilities to exploit (International Business Machines Corporation [IBM] Global Technology Services, 2014). As a form of *social engineering*, phishing emails use deception to exploit these psychological vulnerabilities (Muscanell, Guadagno, & Murphy, 2014). Phishers typically use the identities of well-known and trusted companies. Subsequently, email users are deceived into clicking on an embedded link or opening an email attachment before providing the phisher with confidential information (e.g., passwords) or access to computer systems through the inadvertent installation of malware (Butavicius, Parsons, Pattinson, & McCormac, 2015). Thus, by targeting the user, phishing emails can successfully compromise organisations' information security systems.

Phishing emails are the most prevalent cyber security incident affecting Australian organisations. In 2016, approximately one third of those surveyed experienced phishing attacks on a weekly or monthly basis (Telstra Corporation, 2017).While technical safeguards are often implemented to defend against the threat of phishing (Purkait, 2012), these are not guaranteed solutions. For instance, although email filters can be successful in preventing phishing attacks from reaching the user's inbox, these filters rely on updates in the wake of new attacks to maintain their effectiveness. Hence, there is a window of vulnerability in the time between when new attacks are instigated and email filters are updated to defend against them. This is especially concerning given that 255,065 unique phishing attacks worldwide were recorded in the year 2016 alone (Anti-Phishing Working Group [APWG], 2017). Exposure to phishing attacks therefore remains a possibility for almost all email users. For this reason, it is important to recognise and understand the human element of susceptibility to phishing. In other words, researchers must seek to understand why users fall victim to phishing attacks and, in turn, how to reduce the phishing susceptibility of users and their organisations.

Accordingly, the present study aims to investigate users' information processing strategy as a key contributor to their susceptibility. Specifically, it aims to determine whether manipulating systematic processing can affect their judgements of the legitimacy of phishing and genuine emails. The outcomes of this research have potential implications for cyber security training.

1.1 The Design of Phishing Studies

The methodologies used by phishing researchers can generally be categorised as either 'real phishing', phishing 'IQ test' or 'scenario-based' (also known as 'role-play'). In real phishing studies, simulated phishing attacks are carried out using participants' normal email inboxes, and participants are not aware that their responses (or lack of responses) to these simulations are recorded by researchers (Finn & Jakobsson, 2007). Although real-phishing studies have high real-world face validity, they are limited in what they can tell researchers about user susceptibility. This is mainly because they do not observe participants' responses to genuine emails (Butavicius, Parsons, Pattinson, & McCormac, 2015), i.e., they observe correct phishing judgements for phishing emails, but not incorrect phishing judgements for genuine emails.

Unlike the real phishing methodology, studies using the phishing IQ test or scenario-based methodologies observe participants' responses to both genuine and phishing emails (Parsons, McCormac, Pattinson, Butavicius, & Jerram, 2013). In this way, phishing researchers can apply the performance measurement approach known as Signal Detection Theory (SDT; Green & Swets, 1966). This approach can be applied to any task that involves discriminating between two possible stimulus types (Stanislaw & Todorov, 1999). SDT provides two performance measures. In the context of phishing, *discrimination* measures how well individuals can distinguish between genuine and phishing emails, and *bias* measures the overall tendency to judge emails as either genuine or phishing. This approach reconceptualises the problem of phishing susceptibility as one of decision-making under uncertainty. Users make ‘phishing’ or ‘genuine’ judgements not only for phishing emails, but genuine emails as well. The SDT approach acknowledges that users make these judgements, correctly or incorrectly, for both email types. Therefore, phishing studies that do not accommodate the SDT approach are only capable of telling half of the story.

Despite their similarities, there is an important difference between the phishing IQ test and scenario-based methodologies. Unlike scenario-based studies, participants of phishing IQ test studies are aware that their ability to detect phishing emails is being measured. Accordingly, these participants are primed for signs of phishing and have been shown to exhibit the *subject expectancy effect* (Anandpara, Dingman, Jakobsson, Liu, & Roinestad, 2007; Parsons, McCormac, Pattinson, Butavicius, & Jerram, 2013). Parsons McCormac, Pattinson, Butavicius, and Jerram (2013) conducted a study to test the prediction by Anandpara, Dingman, Jakobsson, Liu, and Roinestad (2007) that participants exhibiting the subject expectancy effect are more biased towards making ‘phishing’ judgements. The prediction was based on the assertion that these participants are more suspicious of the legitimacy of emails compared to everyday email use. In contrast, priming participants with the notion of phishing was actually found to improve their ability to discriminate between genuine and phishing emails. Parsons et al. (2013) speculated that these participants engaged in more diligent decision-making, and this is supported by the finding that they took significantly longer to complete the experiment than participants who were not primed for signs of phishing. Due to the subject expectancy effect, the findings of phishing IQ test studies cannot be generalised to actual user behaviour.

Researchers who use the scenario-based methodology incorporate a role play into their study design. Participants are not informed they are participating in a phishing study. Rather, they are instructed to assume the role of a fictitious person and to make judgements on emails received by that person. Parsons et al. (2013) observed the subject expectancy effect in participants who were informed the study was about phishing, but not in participants who were informed the study was about email management. In this way, Parsons et al. (2013) validated the use of scenarios as a method for overcoming the subject expectancy effect. The scenario-based methodology was used in the present study.

1.2 User Susceptibility to Phishing

Research has shown that phishers utilise *social influence* in their construction of phishing emails to persuade users to respond (Akbar, 2014). Social influence refers to overt forces occurring in our interactions with others that have the power to cause a change in our attitudes or behaviours (Cialdini & Goldstein, 2004; Muscanell, Guadagno, & Murphy, 2014). In the context of face-to-face interactions, Cialdini (2009) established six social influence techniques that serve as heuristics for decision-making: authority, scarcity, consistency, reciprocation, liking and social proof. When exploited by social engineers, these strategies are experienced as external social pressure to agree or comply with a request. In this way, social influence is used to persuade individuals and gain their compliance. Notably, social influence attempts are increasingly occurring in online contexts, and researchers have identified the use of all six social influence strategies by phishers (Akbar, 2014; Ferreira & Lenzini, 2015).

Researchers have further investigated whether some techniques are more persuasive in phishing emails than others (Butavicius, Parsons, Pattinson, & McCormac, 2015; Wright, Jensen, Thatcher, Dinger, & Marett, 2014). Wright, Jensen, Thatcher, Dinger, and Marett (2014) found that liking had the largest positive effect on likelihood to respond, followed by scarcity, social proof and reciprocation, whereas consistency was ineffective and authority had an unexpected negative effect. Butavicius, Parsons, Pattinson, and McCormac (2015) investigated the persuasiveness of authority, scarcity and social proof. In contrast to Wright et al., authority was found to be the most effective social influence technique. Unexpectedly, the absence of social influence in a phishing email was found to be more effective than any of the techniques examined. Both Wright et al. and Butavicius et al. speculated that their unexpected results could be attributed to users' increasing familiarity with social influence attempts in generic phishing emails, and hence their development of resistance to such attempts.

Research has found that higher levels of cognitive impulsivity may increase users' susceptibility to phishing (Butavicius, Parsons, Pattinson, & McCormac, 2015; Parsons, McCormac, Pattinson, Butavicius, & Jerram, 2013; Welk et al., 2015). This finding is relevant to the use of social influence by phishers because cognitive impulsivity can affect how a persuasive message is processed by the receiver. Cognitive impulsivity refers to a person's tendency to make decisions with little or no conscious thought. People who have low levels of cognitive impulsivity are more likely to monitor and correct their cognitive impulses (Primi, Morsanyi, Chiesi, Donati, & Hamilton, 2016). Although two phishing studies that measured cognitive impulsivity did not find this relationship with phishing susceptibility, this is perhaps because Kumaraguru et al. (2007) incorporated only three phishing emails and a small participant sample, and Mayhorn and Nyeste (2012) provided feedback to participants on their ability to detect phishing during a training intervention.

All of these studies except for Welk et al. (2015) used the Cognitive Reflection Test (CRT; Frederick, 2005) to measure cognitive impulsivity. The CRT is a set of three problems, each designed to invoke an immediately intuitive but incorrect response. This intuitive response must be suppressed and overridden by more careful analytic reasoning in order to yield the correct answer. For example, the following is the first problem: “A bat and a ball together cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost?” (Frederick, 2005, p. 27). Initially, there is a strong tendency for respondents to consider the answer of “10 cents”, even by those who then use analytic reasoning to give the correct answer of “5 cents”. Butavicius et al. (2015) suggested that the reason cognitive impulsivity, and the CRT in particular, is associated with phishing susceptibility is that it may indicate whether users tend to rely on heuristic cues to make judgements of email legitimacy. Social influence takes advantage of our tendency to rely on heuristic cues. Each technique has the “ability to produce a distinct kind of automatic, mindless compliance from people, that is, a willingness to say yes without thinking first” (Cialdini, 2009, p. xiv). In this way, Butavicius et al. refer to *dual-system information processing models* as one explanation for why some people are persuaded to respond to phishing emails.

1.3 Heuristics and User Decision-making

Dual-system information processing models have been proposed to account for how individuals process persuasive messages, including the Elaboration Likelihood Model (Petty & Cacioppo, 1986) and the Heuristic-Systematic Model (Chen & Chaiken, 1999). According to these models, individuals rely on two separate strategies for processing persuasive messages; whereas *heuristic processing* is automatic, quick, effortless and intuitive, *systematic processing* is controlled, slow, effortful and analytic. While systematic processing generally yields better-quality judgements, heuristic processing is more economic given our limited cognitive resources and the immense volume of information that we must continually process. For this reason, heuristic processing is understood to be the default strategy for processing information. Nevertheless, the judgements formed on the basis of heuristic processing can be overridden by systematic processing.

When heuristic processing is used, only the superficially persuasive, i.e., heuristic, cues inform a person's judgement (e.g., likeability of the message's source). This means that social influence is likely to be more effective when the message is processed heuristically (Kaptein, Markopoulos, de Ruyter, & Aarts, 2015). Conversely, when systematic processing is used, judgement becomes informed by evidentiary, i.e., systematic, cues (e.g., reliability of the message's source). Hence, activating systematic processing over and above heuristic processing may be an effective strategy for resisting social influence attempts (including those occurring in phishing emails), as it changes the way the persuasive message is processed and potentially the individual's response (Sagarin & Cialdini, 2004). Thus, the use of systematic processing may reduce susceptibility to phishing.

In addition, research has indicated that individuals routinely rely on heuristic cues when processing information online (Guadagno & Cialdini, 2005; Guadagno, Muscanell, Rice, & Roberts, 2013; Metzger, Flanagin, & Medders, 2010). Muscanell, Guadagno, and Murphy (2014) argue this is because the online context imposes a high cognitive load on users; not only do users have access to a vast amount of information, they often multitask, making them more prone to relying on heuristic processing as a way of conserving cognitive resources. Similarly, Vishwanath, Herath, Chen, Wang, and Rao (2011) argue that users eventually learn to manage this high cognitive load by developing heuristic, habitual response patterns. This, in turn, is argued to make individuals more likely to inattentively respond to emails, and hence more susceptible to phishing. This was supported by research that found greater heuristic processing and email habit strength significantly increases phishing susceptibility, whereas greater systematic processing significantly reduces it (Vishwanath, 2015; Vishwanath, Harrison, & Ng, 2016).

Wang, Herath, Chen, Vishwanath, and Rao (2012) did not find that expending greater cognitive effort (i.e., systematic processing) whilst processing a phishing email reduced the likelihood of responding. However, Wang et al. used a single email that incorporated several indicators of phishing, including multiple grammatical errors and an illegitimate sender's address (Parsons et al., 2015). Parsons et al. (2015) showed that spelling and grammatical errors in particular is an important illegitimacy cue relied upon by users. However, many phishing emails do not contain these errors, and many additionally contain a 'spoofed' sender's address (Furnell, 2007) so that, for example, a phishing email using the identity of *PayPal* can appear to have been sent from the address "services@paypal.com". The high availability of phishing cues may have meant that participants did not expend a great deal of cognitive effort to reach a phishing decision. Therefore, it would be interesting to investigate the effects of systematic processing on users' legitimacy judgements for both phishing and genuine emails where multiple, fallible cues are not present. Nevertheless, Wang et al.'s finding that attention to a social influence technique increased the likelihood to respond provides further evidence to support dual-system information processing as an explanation of phishing susceptibility.

In summary, evidence from phishing research supports dual-system information processing as an explanation for why some users respond to phishing emails. This has important implications. It suggests that increasing users' reliance on systematic processing when managing emails could reduce their susceptibility to phishing. Our review of the literature determined that the experimental effects of manipulating users' information processing strategy (i.e., heuristic vs. systematic processing) on their judgements of the legitimacy of phishing and genuine emails is yet to be investigated.

1.4 Manipulations of Information Processing Strategy

Research has shown that individuals' information processing strategy can be manipulated, at least in the short term. The only study identified to have manipulated information processing strategy in the context of emails was conducted by Yan and Gozu (2012). In a repeated measures design, participants were instructed to read the email content of spam emails either 'quickly and casually' (heuristic processing strategy) or 'slowly and carefully' (systematic processing strategy). It was found that participants were more likely to delete the emails, rather than open or respond to them, when relying on systematic processing. While the results are consistent with a dual-system information processing explanation for phishing susceptibility, the study did not examine the manipulation effects on genuine emails. Therefore, it could not determine whether the information processing manipulation merely produced a greater bias towards 'phishing' judgements for both spam and genuine emails, rather than an improved ability to discriminate between the two email types.

Several authors have experimented with information processing manipulations outside of the email context to examine the effects on various judgments. Attridge and Inglis (2015) examined the effects of completing one or more problems from the Raven's Advanced Progressive Matrices (RAPM; Raven, Raven, & Court, 1998). These problems are explicitly difficult tasks requiring the use of systematic processing (Attridge & Inglis, 2015). Respondents are presented with an incomplete matrix of visual designs, and they are asked to identify the missing design that completes the pattern from the response options provided. The authors found that merely completing a single RAPM problem significantly improved participants' performance on the CRT.

In addition to being an outcome variable, the CRT itself has been used as an information processing manipulation. Pinillos, Smith, Nair, Marchetto, and Mun (2011) found that answering one or more of the three CRT problems correctly subsequently produced a pattern of philosophical judgements that had greater similarity to those of philosophers. The authors argued that increased systematic processing was the effect of experiencing the realisation that one's first response was incorrect, hence prompting more careful reasoning.

These findings support the possibility that users' information processing strategy can be manipulated in a way that might affect their judgements of emails. The effects of such a manipulation could be similar to the findings observed by Parsons et al. (2013), where priming participants with the notion of phishing was found to improve their ability to discriminate between phishing and genuine emails. Research has found a significant relationship between phishing susceptibility and performance on the CRT (Butavicius, Parsons, Pattinson, & McCormac, 2015; Parsons, McCormac, Pattinson, Butavicius, & Jerram, 2013), and Pinillos et al. (2011) demonstrated that the CRT is an effective information processing manipulation. Accordingly, it is important to investigate the effects of the CRT as an information processing manipulation on users' email judgement.

However, research has also identified a number of limitations associated with the CRT. First, male respondents tend to score higher than female respondents (Thomson & Oppenheimer, 2016). Thomson and Oppenheimer (2016) argue that the difference between genders can be explained by the results of studies indicating that CRT performance shares a positive correlation with numerical reasoning ability (Weller et al., 2013; Welsh, Burns, & Delfabbro, 2013). This suggests that respondents can reject the heuristic answer but still answer incorrectly due to low numerical reasoning ability. Second, Thomson and Oppenheimer found evidence to suggest that the CRT suffers from over-exposure amongst research participants. This is in addition to the finding by Chandler, Mueller, and Paolacci

(2014) that CRT performance shares a positive correlation with research participation experience. These findings potentially undermine the validity of the CRT.

Consequently, Thomson and Oppenheimer (2016) developed and validated the CRT-2. The CRT-2 is a set of four problems designed to provide an extension of, or an alternative to, the original CRT. It is found to share a weaker correlation with numerical reasoning ability, to have no gender differences and to elicit a higher proportion of systematic, rather than heuristic, responses than the CRT. Otherwise, the CRT-2 shares the same characteristics as the CRT. For these reasons, the effects of attaining a high CRT-2 score on email judgement may be different to the effects of attaining a high score on the CRT.

Like the CRT, matrix reasoning problems have also been shown to be an effective information processing manipulation (Attridge & Inglis, 2015). These problems are multiple choice and are not designed to have intuitive responses. Hence, the effects of attaining a high matrix reasoning task (MRT) score on email judgement may be different to the effects of attaining a high score on the CRT and CRT-2. A notable advantage of a MRT is that, unlike the CRT and CRT-2, it is easy to generate new problems that individuals have never been exposed to before. Another advantage is that an MRT is less culturally biased. Therefore, the present study also sought to investigate the effects of completing a Matrix Reasoning Task (MRT) on users’ judgements of the legitimacy of phishing and genuine emails.

1.5 The Present Study

The present study used a scenario-based methodology to investigate the effects of systematic processing on phishing susceptibility. The study incorporated 14 images of emails, half of which were genuine and the other half phishing, created in a way that the only cue for legitimacy was the URL. Three different tasks were utilised to manipulate information processing strategy: the CRT, CRT-2 and a MRT. The main aims of this study are summarised below:

1. To determine the effect of systematic processing on users' ability to discriminate between phishing and genuine emails.
2. To determine the effect of systematic processing on users' bias towards judging an email as either 'phishing' or 'genuine'.
3. To compare the CRT, CRT-2 and MRT as manipulations of information processing strategy in the context of email judgement.

2. Method

2.1 Participants

Participants were recruited between May and July 2017 using the Qualtrics online survey platform. They were required to be working adults (18 years or older) living in Australia and to spend at least some proportion of their time at work using a computer or portable device (e.g., laptop, tablet, smartphone). Participants received an incentive via Qualtrics to participate in the study.

A total of 1,082 participants met the inclusion criteria and finished the survey with a completion time that did not indicate inattentive responding. An additional 45 participants were excluded from analysis based on inattentive responding (e.g., responding with the same response option for 90% or more of the items within a scale). This resulted in a total sample of 1,037 participants (551 males, 485 females and 1 gender unspecified). Participants were evenly distributed across age categories, with 18.6% of them aged between 18-29, 22.7% aged between 30-39, 20.6% aged between 40-49, 21.8% aged between 50-59 and 16.3% aged 60 and over.

2.2 Materials

2.2.1 Email Stimuli.

The main study incorporated 14 images of emails, half of which were genuine emails and the other half phishing emails. These images were created by using actual genuine and phishing emails, either received by the authors or found online, as templates.

The emails were created in a way that the only cue for legitimacy was the URL (i.e., all emails contained logos and a legitimate sender's address, and they lacked personalisation and spelling and grammatical errors). The genuine emails contained URLs that were consistent with actual URLs of the claimed sender. The phishing emails contained actual phishing URLs, modified by a single character (see Appendix C for examples of the email stimuli). At the beginning of the Email Task, participants were advised that if they were to hover over a link (or hold their finger down on a link if on a mobile device), they would be shown where it would take them (i.e., the URL was displayed to the user).

The 14 emails were described as having been taken from the inbox of 'Alex Jones', and participants were instructed to assume that all emails had been sent to Alex deliberately (i.e., "Alex has not received them by mistake") and that the topics in the emails were relevant to Alex (i.e., "if the email mentions a piece of software, assume that Alex is interested in that software"). This role play methodology is intended to prevent participants from giving responses that reflect judgements of personal relevance rather than legitimacy. A range of email topics were utilised (see Table 1 for a description of each email). Participants were advised that the names and contact details in the emails were fictitious.

2.2.2 Information Processing Manipulations.

Three different tasks were utilised as information processing manipulations (IPMs) (see Appendix D).

2.2.2.1 CRT.

	Sender	
Email Topic	Phishing Email	Genuine Email
Police matter	Australian Federal Police	National Crime Check
Enter competition	Coles	Velocity Frequent Flyer
Upgrade for an improved user experience	Microsoft Outlook	Dropbox
Customer reward program voucher	Amazon Prime	Athlete's Foot
Donate to charity	Australian Red Cross	Ronald McDonald House Charity
Customer satisfaction survey	McDonald's	Virgin Mobile
Reset password	Apple	LinkedIn

Table 1

Description of email stimuli.

The three CRT items were presented on the same page, in a set order, and participants provided text responses. Responses were categorised as 'correct' or 'incorrect' and participants received a score between 0 and 3.

2.2.2.2 CRT-2.

The four CRT-2 items were presented on the same page, in a set order, and participants provided text responses. Responses were categorised as 'correct' or 'incorrect' and participants received a score between 0 and 4.

2.2.2.3 MRT.

Three matrix reasoning items were selected from the pool of items by the International Cognitive Ability Resource Team (ICAR, 2014). They are similar to RAPM items (Raven, Raven, & Court, 1998). The three items were presented on the same page, in a set order, and participants selected their responses from the six options provided. Responses were categorised as 'correct' or 'incorrect' and participants received a score between 0 and 3.

2.2.2.4 Task Difficulty.

Participants rated the difficulty of each IPM after its completion by responding to the item “Overall, how difficult do you think it was to solve those problems?” on a 5-point likert scale ranging from *very easy* to *very difficult*.

2.2.3 Email Judgement Measures.

2.2.3.1 Link Safety Rating.

The 14 images of emails were presented separately in a random order, and participants were asked to respond to the statement “It is okay to click on the link in this email” on a 5-point likert scale ranging from strongly disagree to strongly agree.

2.2.3.2 Discrimination and Bias.

The SDT non-parametric measures of discrimination and bias (Stanislaw & Todorov, 1999), represented as *A’* and *B”* respectively, were calculated using participants responses to the Link Safety Rating items. A score of 1 for *A’* means that discrimination performance is perfect, while a score of .5 means that phishing emails cannot be distinguished from genuine emails. *B”* scores range from -1 (phishing responses only) to 1 (genuine responses only), while a score of 0 indicates no response bias.

2.3 Procedure

Ethics approval was granted by the Human Research Ethics Subcommittee of The University of Adelaide School of Psychology. The present study was part of a larger data collection project, and so does not report on all of the materials that were incorporated in the survey.

Participants were invited to participate via email received from Qualtrics. They were informed the study as an investigation of how people use email and social media, and the factors that may affect how people use email and social media. Participants were required to confirm that they had read the Information Sheet and given their consent to participate in order to continue with the survey.

Participants first responded to inclusion criteria items and demographic items before completing a 13-item self-construal measure not analysed in the present study. This measure, together with the aforementioned items, acted as a buffer to ensure that the activities participants were engaged in just prior to beginning the survey did not have an effect on their information processing strategy during subsequent sections of the survey (Hauser & Schwarz, 2015).

Participants were then randomly allocated to one of four conditions, as depicted in Figure 1. Participants in the three experimental conditions completed an IPM (i.e., the CRT, CRT-2 or MRT). Only those participants in the Control condition did not complete an IPM prior to the Email Task. All participants completed the Email Task, followed by those IPMs that they had not completed earlier. This was to ensure the length of the study was consistent across conditions, and to determine that there were no differences in scores on the IPMs between participants assigned to separate conditions. Throughout the survey, participants rated the difficulty of each IPM immediately after its completion.

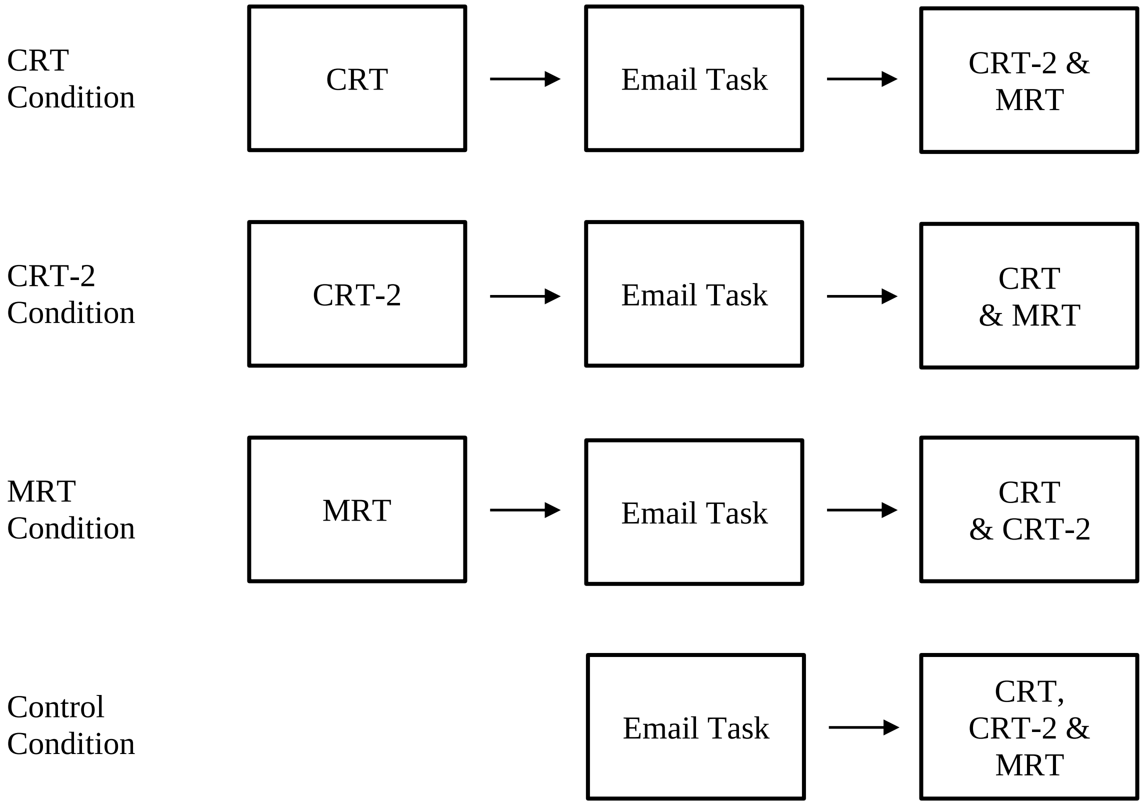


Figure 1. Experimental design with four conditions.

3. Results

3.1 Information Processing Manipulation Characteristics

Table 2 presents IPM descriptive statistics across the sample and by condition. A series of one-way between groups ANOVAs determined that there were no differences between the four conditions regarding performance on the CRT, *F*(3, 1033) = 0.45, *p* = .72, CRT-2, *F*(3, 1033) = 1.10, *p* = .35, and MRT, *F*(3, 1033) = 0.36, *p* = .78.

An independent samples *t*-test indicated that male participants (*M* = 0.92, *SD* = 1.06) scored significantly higher on the CRT than female participants (*M* = 0.60, *SD* = 0.91), *t*(1033.25) = 5.10, *p* < .001.

		CRT				CRT-2				MRT			
		Score (%)		Task difficulty		Score (%)		Task difficulty		Score (%)		Task difficulty	
Condition	*n*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
CRT	253	25.82	33.74	2.66	1.18	40.32	31.97	2.41	1.10	48.22	32.57	3.44	1.10
CRT-2	261	25.93	34.04	2.84	1.19	39.75	31.82	1.96	0.89	48.66	33.00	3.53	1.07
MRT	265	27.30	34.53	2.80	1.22	43.77	32.91	2.34	1.08	48.43	32.67	3.45	1.09
Control	258	23.90	31.97	2.78	1.25	39.15	31.78	2.32	1.13	45.99	33.71	3.45	1.12

Table 2

Information processing manipulation descriptive statistics.

Note. CRT = Cognitive Reflection Test; MRT = Matrix Reasoning Task.

Conversely, no differences were observed between male and female participants on the CRT-2 (*M* = 1.64, *SD* = 1.30 and *M* = 1.62, *SD* = 1.27, respectively), *t*(1034) = 0.25, *p* < .80, and the MRT (*M* = 1.43, *SD* = 0.97 and *M* = 1.44, *SD* = 1.01, respectively), *t*(1034) = 0.24, *p* = .81.

Furthermore, a one-way repeated measures ANOVA found significant differences between participants' scores (percentage correct) across the three IPMs, Wilks' Lambda = .73, *F*(2, 1035) = 196.26, *p* < .001, η_p^2 = .28. Post-hoc pairwise comparisons indicated significantly higher scores on the MRT than both the CRT (*p* < .001, *d* = 0.66) and CRT-2 (*p* < .001, *d* = 0.22), and significantly higher scores on the CRT-2 than the CRT (*p* < .001, *d* = 0.46). There were also significant differences found by a one-way repeated measures ANOVA between participants' task difficulty ratings across the IPMs, Wilks' Lambda = .50, *F*(2, 1035) = 521.50, *p* < .001, η_p^2 = .50. Despite attaining higher scores on the MRT, participants rated the MRT as significantly more difficult to complete than both the CRT (*p* < .001, *d* = 0.61) and CRT-2 (*p* < .001, *d* = 1.12). The CRT was also rated as significantly more difficult than the CRT-2 (*p* < .001, *d* = 0.45).

3.2 Effects of Systematic Processing

To calculate SDT measures, Link Safety Rating was recoded into a binary variable (*Email Judgement*). Ratings of 4 (*agree*) and 5 (*strongly agree*) were classified as 'correct' for genuine emails and 'incorrect' for phishing emails. Ratings of 1 (*strongly disagree*) and 2 (*disagree*) were classified as 'incorrect' for genuine emails and 'correct' for phishing emails. Ratings of 3 (*unsure*) were classified as incorrect for both email types.

Table 3 presents the descriptive statistics for Email Judgement and SDT measures by condition. When calculated for *all emails*, Email Judgement scores have a minimum of 0 and a maximum of 14. The scores for *A'* and *B"* across the sample indicate poor discrimination between phishing and genuine emails and a bias towards phishing decisions.

	Email Judgement						Signal Detection Theory			
	All emails		Genuine		Phishing		*A’*		*B”*	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
CRT										
High	7.00	2.63	2.57	2.33	4.43	2.49	.49	.30	-.25	.67
Low	6.18	2.47	2.49	2.17	3.69	2.43	.40	.27	-.13	.60
CRT-2										
High	6.65	2.43	2.66	2.22	3.99	2.38	.44	.27	-.12	.64
Low	5.98	2.23	2.85	2.16	3.13	2.42	.37	.25	-.10	.65
MRT										
High	6.95	2.50	2.51	2.22	4.44	2.33	.48	.28	-.21	.58
Low	5.99	1.97	2.47	2.15	3.52	2.32	.37	.22	-.07	.55
Control	5.95	2.32	2.39	2.18	3.57	2.43	.37	.25	-.13	.64

Table 3

Email Judgement and Signal Detection Theory descriptive statistics.

Note. CRT = Cognitive Reflection Test; MRT = Matrix Reasoning Task. The *n* descriptive statistics for the Email Judgement variables are as presented in Table 2. The *n* descriptive statistics for Signal Detection Theory (SDT) measures are different due to error scores (136 error scores produced by *A’* and 149 error scores produced by *B”*).

In line with Kumaraguru et al. (2007), participants who completed an IPM and scored 0 or 1 were categorised as ‘low’ (n = 464), whereas participants who attained a score greater than 1 were categorised as ‘high’ (n = 315). In this way, only those participants who had repeatedly used systematic processing during the IPM were categorised as high (see Table 3 for Email Judgement and SDT descrptive statistics by IPM performance).

3.2.1 Effect on Discrimination

A one-way, between subjects ANOVA was used to determine whether *A’* scores differed across each condition and whether participants attained a low or high score. Since the assumption of homogeneity of variance was not met, the Welch’s adjusted value was used. This analysis was significant $F(6, 276.81) = 2.32$, $p = .03$, $\eta^2 = .016$. Figure 2 shows the effect of condition (CRT, CRT-2, MRT or Control) and IPM performance (high or low) on *A’*. Planned comparisons were then conducted to examine the differences between groups. A Bonferroni adjustment was applied to the alpha level used to determine statistical significance. The conventional alpha level of .05 was divided by three to produce a new alpha level of .017. The results indicated that participants who completed an IPM and attained a high score ($M = 0.47$, $SD = 0.28$) had significantly better discrimination performance compared to participants who attained a low score ($M = 0.38$, $SD = 0.25$), $F(1, 271.68) = 17.36$, $p < .001$, $d = 0.34$, and participants who did not complete an IPM (i.e., the Control condition), $F(1, 326.91) = 17.90$, $p < .001$, $d = 0.23$. Although the effect sizes for the ANOVA and planned comparisons are small, as shown in Table 2, the mean difference between the Control condition and the CRT High group was .12. The small effect sizes could be explained by the large variation in *A’* scores within each of the groups (see Table 2). There was no difference between participants who attained a low score and participants who did not complete an IPM, $F(1, 462.98) = 0.37$, $p = .52$.

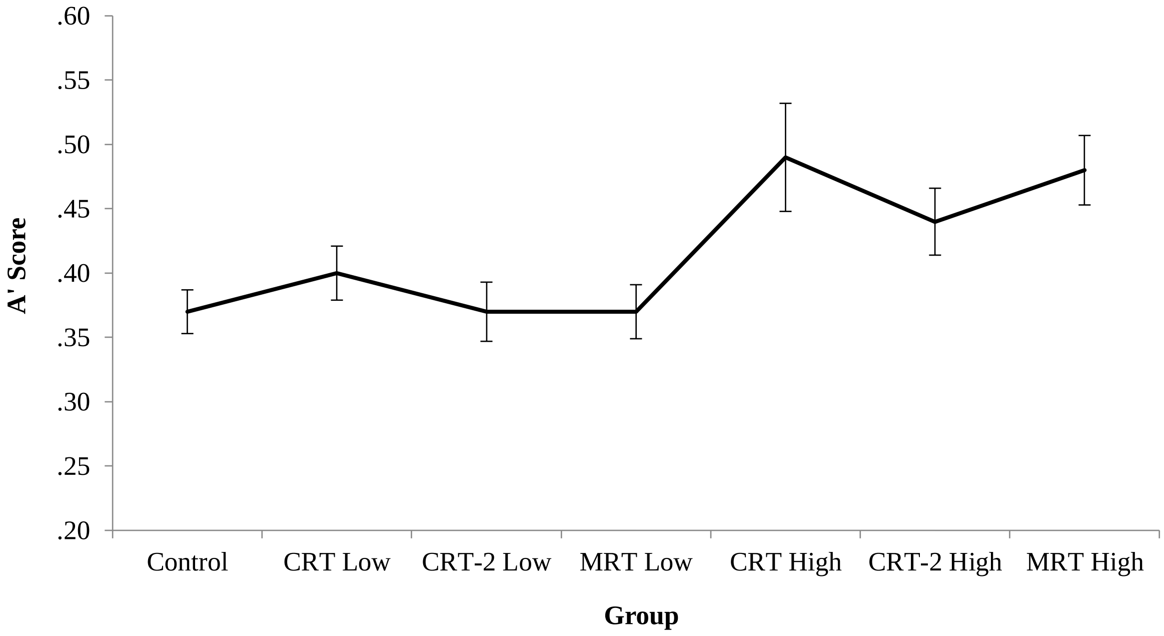


Figure 2. A' scores according to condition (CRT, CRT-2, MRT or Control) and information processing manipulation (IPM) score (low or high).

Pearson's correlation analyses were conducted to determine whether participants' judgements during the Email Task were manipulated by attaining a high score on the IPM, and to compare the CRT, CRT-2 and MRT as manipulations of information processing. If, for example, the correlation between CRT scores and *A'* scores is larger for the CRT condition than the Control condition, then this would indicate that participants' discrimination performance was improved by attaining higher CRT scores prior to the Email Task. As shown in Table 4, only the MRT condition has a stronger correlation between IPM score and *A'* than the Control condition.

3.2.2 Effect on Bias

A one-way, between subjects ANOVA was used to determine whether *B"* scores differed across each condition and whether participants attained a low or high score. This analysis was non-significant, *F*(6, 881) = .78, *p* = .59. Therefore, there was no effect of IPM on participants' bias and no further analyses were conducted.

	Pearson's *r* correlation with *A'*	
IPM	IPM conditions[a]	Control condition
CRT	.14*	.23**
CRT-2	.17**	.25**
MRT	.19**	.14*

Table 4

Correlation between information processing manipulation scores and A'.

Note. IPM = information processing manipulation; CRT = Cognitive Reflection Test; MRT = matrix reasoning task.

[a]The correlation reported is respective to condition, e.g., the correlation between CRT performance and *A'* is reported only for the CRT condition.

*p < .05. **p < .001

4. Discussion

The present study investigated the effects of systematic processing on phishing susceptibility. Participants either completed one of three IPMs or did not complete an IPM before responding to email stimuli. Their scores (low or high) on these IPMs were then used as an indication of their information processing strategy (heuristic vs. systematic) when responding to email stimuli.

The first aim of the study was to determine the effect of systematic processing on users' ability to discriminate between phishing and genuine emails. Consistent with previous research (Butavicius, Parsons, Pattinson, & McCormac, 2015; Parsons, McCormac, Pattinson, Butavicius, & Jerram, 2013; Welk et al., 2015), participants who completed an IPM and attained a high score were better able to discriminate between phishing and genuine emails. Although the effect sizes of this analysis were small, this can be explained by the large variation within each of the groups. This variation was potentially caused by individual differences in information processing (Hamilton, Shih, & Mohammed, 2016; Cacioppo &

Petty, 1982). Furthermore, discrimination was only found to be improved by attaining higher scores on the MRT, where the correlation between MRT performance and discrimination performance was higher for the MRT condition than the Control condition. Unexpectedly, the correlation between IPM performance and discrimination performance was higher for the Control condition than the CRT and CRT-2 conditions.

There are several potential explanations for these inconsistent findings between the three IPMs. First, the difference in characteristics between the MRT and the CRT and CRT-2 may have interacted with individual differences in information processing. It is possible that participants who had a relatively strong preference for systematic processing were less likely to be affected by the manipulation; whether or not they completed the IPM before the Email Task, they may have been more likely to process the email stimuli systematically. Similarly, participants who had a relatively strong preference for heuristic processing were perhaps less likely to answer the IPM problems correctly and more likely to process the email stimuli heuristically. However, for those individuals who have no strong preference for either of the information processing strategies, it is possible that the CRT and CRT-2 were more likely to encourage the use of heuristic processing, whereas the MRT was more likely to activate systematic processing. This could be because, unlike the MRT, both the CRT and CRT-2 are designed to immediately invoke a heuristic response. While heuristic processing is our default information processing strategy (Chen & Chaiken, 1999), a heuristic answer is not necessarily immediately available when responding to the MRT. It is perhaps far more obvious than the CRT and CRT-2 that systematic processing is required to solve the MRT problems correctly. This could explain why only the MRT was found to activate systematic processing, and is supported by the findings that the MRT was significantly more likely to be solved correctly whilst also being perceived as more difficult to complete.

Second, participants' performance on the IPMs used in the study may not be a reliable indication of their information processing strategies. The fact that an individual answered an IPM problem incorrectly does not necessarily mean that they weren't using systematic processing to arrive at their answer. In addition to systematic processing, all three IPMs require respondents to rely on some form of reasoning ability. Hence, individual differences in participants' reasoning ability may have distorted the data and produced the inconsistent findings.

Third, the Email Task itself may have acted as a manipulation of information processing strategy. Several authors in the literature have indicated that, when processing information online, users are more reliant on heuristic processing as a way of managing the high cognitive load associated with this context (Guadagno & Cialdini, 2005; Guadagno, Muscanell, Rice, & Roberts, 2013; Metzger, Flanagin, & Medders, 2010; Muscanell, Guadagno, & Murphy, 2014). Furthermore, a neuroimaging study by Neupane, Saxena, Kuruvilla, Georgescu, and Kana (2014) observed participants expending "considerable effort" (p. 13) whilst attempting to detect phishing. In the current study, the cognitive demands of the Email Task may have caused participants to become more reliant on heuristic processing. This was perhaps especially the case for participants in the Control condition who had no opportunity to activate systematic processing prior to the Email Task. This would explain the unexpected higher correlations observed between performance on the IPMs and discrimination performance in the Control condition. It is also supported by the observation that the Control condition's scores on all three IPMs were slightly, albeit non-significantly, lower than the IPM scores of the other conditions.

All of these potential explanations should be investigated by future research. In particular, researchers should seek to manipulate information processing strategy in a way that does not depend on participants' performance on the IPM (i.e., participants' categorisation as ‘systematic’ vs. ‘heuristic’ should only depend on their assignment to a condition). An example of such an IPM could be the use of a difficult-to-read font, as opposed to an easy-to-read font. Alter, Oppenheimer, Epley, and Eyre (2007) demonstrated that presenting the CRT in a difficult-to-read font significantly improves individuals’ CRT performance. Researchers should also measure and control for individual differences in information processing. Two measures that could potentially be used in future studies are Hamilton, Shih, and Mohammed’s (2016) Decision Styles Scale and Cacioppo and Petty’s (1982) Need for Cognition scale.

With regards to the second aim of the study, the results indicated there was no effect of systematic processing on users’ bias towards judging an email as either ‘phishing’ or ‘genuine’. This is consistent with findings by Parsons et al. (2013) where priming participants with the notion of phishing was only found to only improve their discrimination performance, and not affect their bias. Parsons et al.’s (2013) interpretation of this priming effect is applicable to the present study. Rather than causing participants to be more suspicious of the emails and hence more biased towards ‘phishing’ judgements, participants were speculated to have engaged in more diligent decision-making, thus improving their discrimination performance.

The third aim of the study was to compare the CRT, CRT-2 and MRT as manipulations of information processing strategy in the context of email judgement. As discussed above, the MRT was the only IPM found to activate systematic processing during the Email Task. In addition, the MRT was more likely to be solved correctly whilst also being perceived as more difficult to complete. Altogether, these findings suggest that the MRT is a more effective manipulation of information processing than CRT and CRT-2 in the context of email judgement. Two advantages of an MRT over the CRT and CRT-2 are that it is easy to generate new problems that individuals have never been exposed to before and the task itself is less culturally biased. In addition, the finding that the CRT is the only IPM to have gender differences is consistent with Thomson and Oppenheimer (2016). Consequently, an MRT should be preferred over the CRT and CRT-2 by phishing researchers in future studies that seek to manipulate information processing, and the CRT-2 should be preferred over the CRT as a measure of cognitive impulsivity.

In summary, the findings of the study are consistent with the body of literature that suggests reliance on predominantly heuristic (rather than systematic) information processing strategies when managing emails is a key contributor to users' susceptibility. Furthermore, these results suggest that increasing users' reliance on systematic processing when managing emails can reduce their phishing susceptibility to some extent. The outcomes of this research have implications for cyber security training. Rather than merely warning users about the threat posed by phishing or even providing instructions for recognising the attacks, users could be trained to activate systematic processing as an effective strategy for defending against the threat of phishing. This could be achieved by requiring users to solve new matrix reasoning problems prior to being allowed to access their email accounts. Future research should seek to examine the effects of completing a MRT in a research setting that is more representative of email management in the real-world.

4.1 Limitations

There are a number of limitations associated with this study. First, as mentioned above, the study did not directly measure the information processing strategies used during the Email Task. Once again, it is suggested that future studies incorporate a measure of individual differences in information processing so as to more reliably investigate the effects on phishing susceptibility.

Second, despite the theoretical relationship between the use of social influence in phishing emails and users' information processing strategy, the study did not investigate the effectiveness of systematic processing against different social influence techniques (Cialdini, 2009). For example, systematic processing activation may have a strong effect on users' susceptibility to the scarcity technique, but little or no effect on their susceptibility to the authority technique. This possibility is consistent with studies that show certain social influence techniques in phishing emails are more effective than others (Butavicius, Parsons, Pattinson, & McCormac, 2015; Wright, Jensen, Thatcher, Dinger, & Marett, 2014). Therefore, the relative effectiveness of systematic processing against each of Cialdini's (2009) six social influence techniques should also be explored by future studies.

Third, the study did not measure actual phishing susceptibility. Participants were not required to click on any of the links or provide personal information, and it is therefore possible that, in a real-world situation, participants may have become suspicious before responding to any of the phishing emails. Given that research has found strong email habits increase phishing susceptibility (Vishwanath, 2015; Vishwanath, Harrison, & Ng, 2016), it also possible that participants may have been less susceptible in the present study because they could not respond habitually, such as automatically clicking on links. Similarly, participants were responding to emails that were received by the inbox of a fictitious person, rather than their own personal inboxes. Participants may respond differently to actual emails received by their personal inboxes compared to the present study.

Furthermore, participants' email judgements were characterised by poor discrimination and a bias towards phishing decisions. This might reflect a lack of engagement with the role-play, so that participants' responses were influenced by judgements of personal relevance. A solution to this might be to provide participants with more contextual information, such as a list of the fictitious person's online accounts. A second explanation is that participants' judgements were influenced by the lack of personalisation in all emails (emails contained generic greetings or no greetings). This is consistent with research that shows personalisation is both an effective cue for detecting phishing and a cue relied upon by users (Parsons, McCormac, Pattinson, Butavicius, & Jerram, 2015). Hence, the effects of systematic processing should be explored using phishing and genuine emails that contain personalised greetings, as is the case in some spear-phishing emails (Butavicius, Parsons, Pattinson, & McCormac, 2015). A third explanation is that participants became biased towards ‘phishing’ decisions due to the high proportion of phishing emails used in the present study compared to the proportion encountered in the real-world. A solution to this might be to incorporate a larger proportion of genuine emails in future research. Eliminating these potential causes of poor discrimination and bias in future studies may result in a larger effect of systematic processing on the ability to discriminate between phishing and genuine emails.

By seeking to address the limitations outlined above, future studies can more accurately determine the effects of systematic processing activation on phishing susceptibility.

4.2 Conclusions

Exposure to phishing attacks is a possibility for almost all email users. For this reason, researchers must seek to understand why users fall victim to phishing attacks and, in turn, how to reduce susceptibility to phishing. The findings of the present study suggest that increasing users’ reliance on systematic processing when managing emails can reduce their phishing susceptibility. Matrix reasoning problems in particular were found to affect users’ judgements of the legitimacy of emails, such that correctly solving these problems improved their ability to discriminate between phishing and genuine emails. This effect is ascribed to the activation of systematic processing.

These findings have implications for cyber security training. It suggests users could be trained to activate systematic processing as an effective strategy for defending against the threat of phishing. Future research should seek to address the limitations of the current study in order to more accurately determine the effects of systematic processing activation on phishing susceptibility.

5. References

Akbar, N. (2014). *Analysing persuasion principles in phishing emails.* (Masters thesis), University of Twente.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*(4), 569-576. doi:10.1037/0096-3445.136.4.569

Anandpara, V., Dingman, A., Jakobsson, M., Liu, D., & Roinestad, H. (2007). *Phishing IQ tests measure fear, not ability*. Paper presented at the 11th International Conference on Financial Cryptography and Data Security, Scarborough, Trinidad and Tobago.

Anti-Phishing Working Group [APWG]. (2017). *Global phishing survey: Trends and domain name use in 2016*. Retrieved from http://docs.apwg.org/reports/APWG_Global_Phishing_Report_2015-2016.pdf

Attridge, N., & Inglis, M. (2015). Increasing cognitive inhibition with a difficult prior task: Implications for mathematical thinking. *ZDM Mathematics Education, 47*(5), 723-734. doi:http://dx.doi.org/10.1007/s11858-014-0656-1

Butavicius, M., Parsons, K., Pattinson, M., & McCormac, A. (2015). *Breaching the human firewall: Social engineering in phishing and spear-phishing emails.* Paper presented at the 26th Australasian Conference on Information Systems, Adelaide, Australia.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116-131. doi:10.1037/0022-3514.42.1.116

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior research methods, 46*(1), 112-130.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73-96). New York, NY: Guilford Press.

Cialdini, R. B. (2009). *Influence: Science and practice* (5th ed.). Boston, MA: Pearson Education, Inc.

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

Ferreira, A., & Lenzini, G. (2015). An analysis of social engineering principles in effective phishing. *2015 Workshop on Socio-Technical Aspects in Security and Trust*, 9-16. doi:10.1109/STAST.2015.10

Finn, P., & Jakobsson, M. (2007). Designing and conducting phishing experiments. *IEEE Technology and Society Magazine, Special Issue on Usability and Security, 26*(1), 46-58.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19*(4), 25-42.

Furnell, S. (2007). Phishing: Can we spot the signs? *Computer Fraud & Security, 2007*(3), 10-15. doi:10.1016/S1361-3723(07)70035-0

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Guadagno, R. E., & Cialdini, R. B. (2005). Online persuasion and compliance: Social influence on the internet and beyond. In Y. Amichai-Hamburger (Ed.), *The social net: The social psychology of the internet* (pp. 91-113). Oxford, UK: Oxford University Press.

Guadagno, R. E., Muscanell, N. L., Rice, L. M., & Roberts, N. (2013). Social influence online: The impact of social validation and likability on compliance. *Psychology of Popular Media Culture, 2*(1), 51-60. doi:10.1037/a0030592

Hamilton, K., Shih, S. I., & Mohammed, S. (2016). The development and validation of the rational and intuitive decision styles scale. *Journal of Personality Assessment, 98*(5), 523-535. doi:10.1080/00223891.2015.1132426

Hauser, D. J., & Schwarz, N. (2015). It's a trap! Instructional manipulation checks prompt systematic thinking on "tricky" tasks. *Sage Open, 5*(2), 1-6. doi:10.1177/ 2158244015584617

International Business Machines Corporation [IBM] Global Technology Services. (2014). *IBM Security Services 2014 cyber security intelligence index: Analysis of cyber attack and incident data from IBM's worldwide security operations*. Retreived from http://www.ibm.com/developerworks/library/se-cyberindex2014/index.html.

Kaptein, M., Markopoulos, P., de Ruyter, B., & Aarts, E. (2015). Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. *International Journal of Human-Computer Studies, 77*, 38-51. doi:10.1016/ j.ijhcs.2015.01.004

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. *Proceedings of the SIGCHI conference on Human factors in computing systems, 905-914.*

Mayhorn, C. B., & Nyeste, P. G. (2012). Training users to counteract phishing. *Work, 41*, 3549-3552.

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication, 60*(3), 413-439.

Muscanell, N. L., Guadagno, R. E., & Murphy, S. (2014). Weapons of influence misused: A social influence analysis of why people fall prey to internet scams. *Social and Personality Psychology Compass, 8*(7), 388-396.

Neupane, A., Saxena, N., Kuruvilla, K., Georgescu, M., & Kana, R. K. (2014). *Neural signatures of user-centered security: An fMRI study of phishing, and malware warnings.* Paper presented at the Network and Distributed System Security Symposium, San Diego, CA.

Parsons, K., Butavicius, M., Pattinson, M., Calic, D., McCormac, A., & Jerram, C. (2015). *Do users focus on the correct cues to differentiate between phishing and genuine emails?* Paper presented at the Australasian Conference on Information Systems, Adelaide, Australia.

Parsons, K., McCormac, A., Pattinson, M., Butavicius, M., & Jerram, C. (2013). *Phishing for the truth: A scenario-based experiment of users' behavioural response to emails.* Paper presented at the IFIP International Information Security Conference, Auckland, New Zealand.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology, 19*, 123-205.

Pinillos, N. Á., Smith, N., Nair, G. S., Marchetto, P., & Mun, C. (2011). Philosophy's new challenge: Experiments and intentional action. *Mind & Language, 26*(1), 115-139.

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *Journal of Behavioral Decision Making, 29*(5), 453-469. doi:10.1002/ bdm.1883

Purkait, S. (2012). Phishing counter measures and their effectiveness – literature review. *Information Management & Computer Security, 20*(5), 382-420.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. San Antonio, TX: Harcourt Assessment.

Sagarin, B. J., & Cialdini, R. B. (2004). Creating critical consumers: Motivating receptivity by teaching resistance. In E. S. Knowles & J. A. Linn (Eds.), *Resistance and persuasion* (pp. 259-282). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149.

Telstra Corporation. (2017). *Telstra Cyber Security Report 2017*. Retrieved from https://www.telstraglobal.com/images/assets/insights/resources/Telstra_Cyber_Security_Report_2017_-_Whitepaper.pdf

The International Cognitive Ability Resource Team [ICAR]. (2014). Matrix reasoning [Items and scoring key]. Retrieved from https://icar-project.com/documents/10

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making, 11*(1), 99-113.

Vishwanath, A. (2015). Examining the distinct antecedents of e-mail habits and its influence on the outcomes of a phishing attack. *Journal of Computer-Mediated Communication, 20*(5), 570-584. doi:10.1111/jcc4.12126

Vishwanath, A., Harrison, B., & Ng, Y. J. (2016). Suspicion, cognition, and automaticity model of phishing susceptibility. *Communication Research*, 1-21. doi:http:// dx.doi.org/10.1177/0093650215627483

Vishwanath, A., Herath, T., Chen, R., Wang, J., & Rao, H. R. (2011). Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems, 51*(3), 576-586.

Wang, J., Herath, T., Chen, R., Vishwanath, A., & Rao, H. R. (2012). Phishing susceptibility: An investigation into the processing of a targeted spear phishing email. *IEEE*

Transactions on Professional Communication, 55(4), 345-362. doi:10.1109/TPC.2012.2208392

Welk, A. K., Hong, K. W., Zielinska, O. A., Tembe, R., Murphy-Hill, E., & Mayhorn, C. B. (2015). Will the "phisher-men" reel you in?: Assessing individual differences in a phishing detection task. *International Journal of Cyber Behavior, Psychology and Learning, 5*(4), 1-17.

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making, 26*(2), 198-212. doi:10.1002/bdm.1751

Welsh, M., Burns, N., & Delfabbro, P. H. (2013). *The cognitive reflection test: How much more than numerical ability?* Paper presented at the 35th Annual Meeting of the Cognitive Science Society, Berlin, Germany.

Wright, R. T., Jensen, M. L., Thatcher, J. B., Dinger, M., & Marett, K. (2014). Influence techniques in phishing attacks: An examination of vulnerability and resistance. *Information Systems Research, 25*(2), 385-400.

Yan, Z., & Gozu, H. Y. (2012). Online decision-making in receiving spam emails among college students. *International Journal of Cyber Behavior, Psychology and Learning, 2*(1), 1-12. doi:10.4018/ijcbpl.2012010101

Appendices

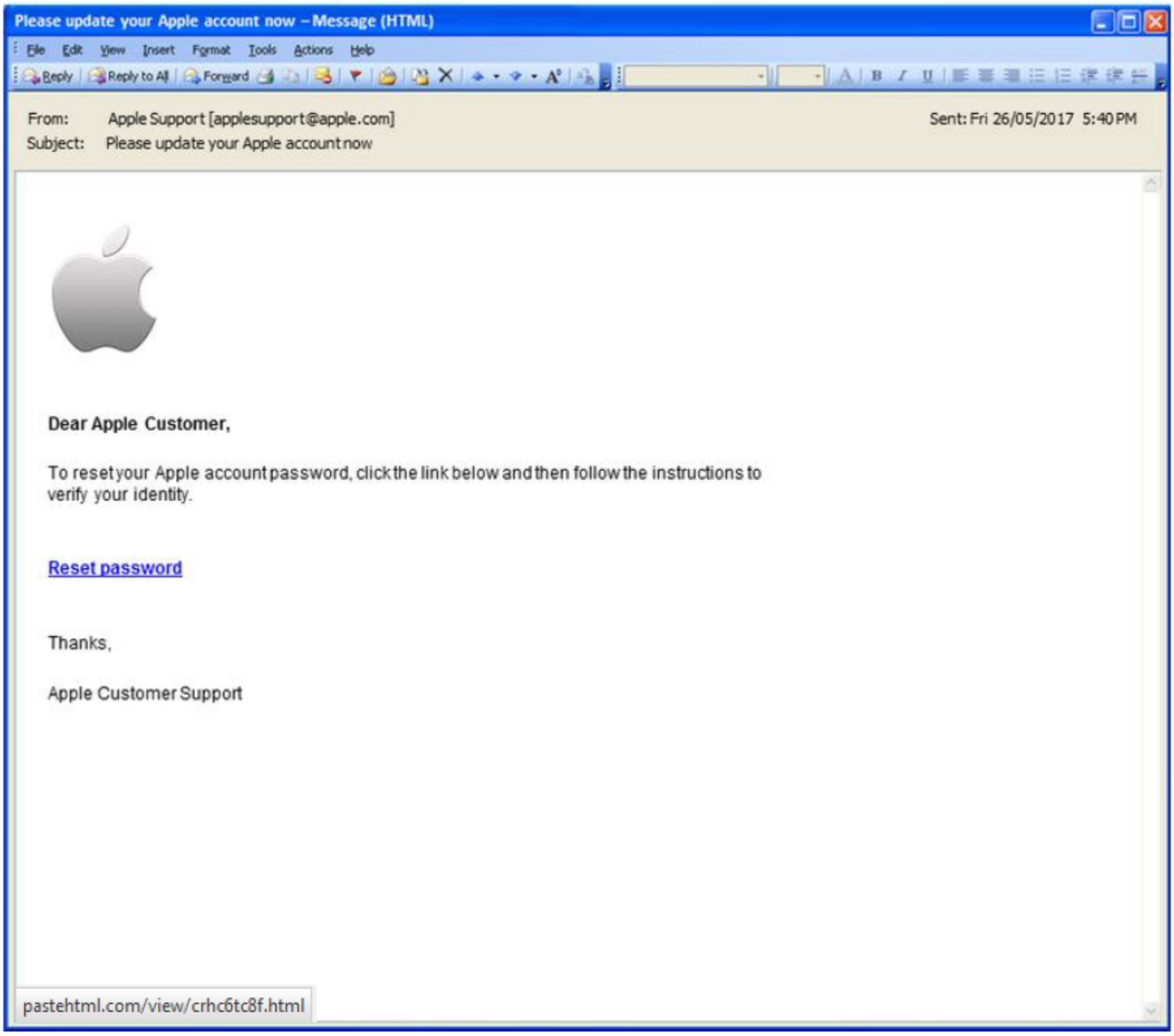
Appendix A: Journal Guidelines for Submission

(See attached).

Appendix B: Acknowledgements (as required by Journal Guidelines)

(See attached).

Appendix C: Examples of Email Stimuli


Please update your Apple account now – Message (HTML)
File Edit View Insert Format Tools Actions Help
Reply Reply to All Forward

From: Apple Support [applesupport@apple.com]

Subject: Please update your Apple account now

Sent: Fri 26/05/2017 5:40 PM

Dear Apple Customer,

To reset your Apple account password, click the link below and then follow the instructions to verify your identity.

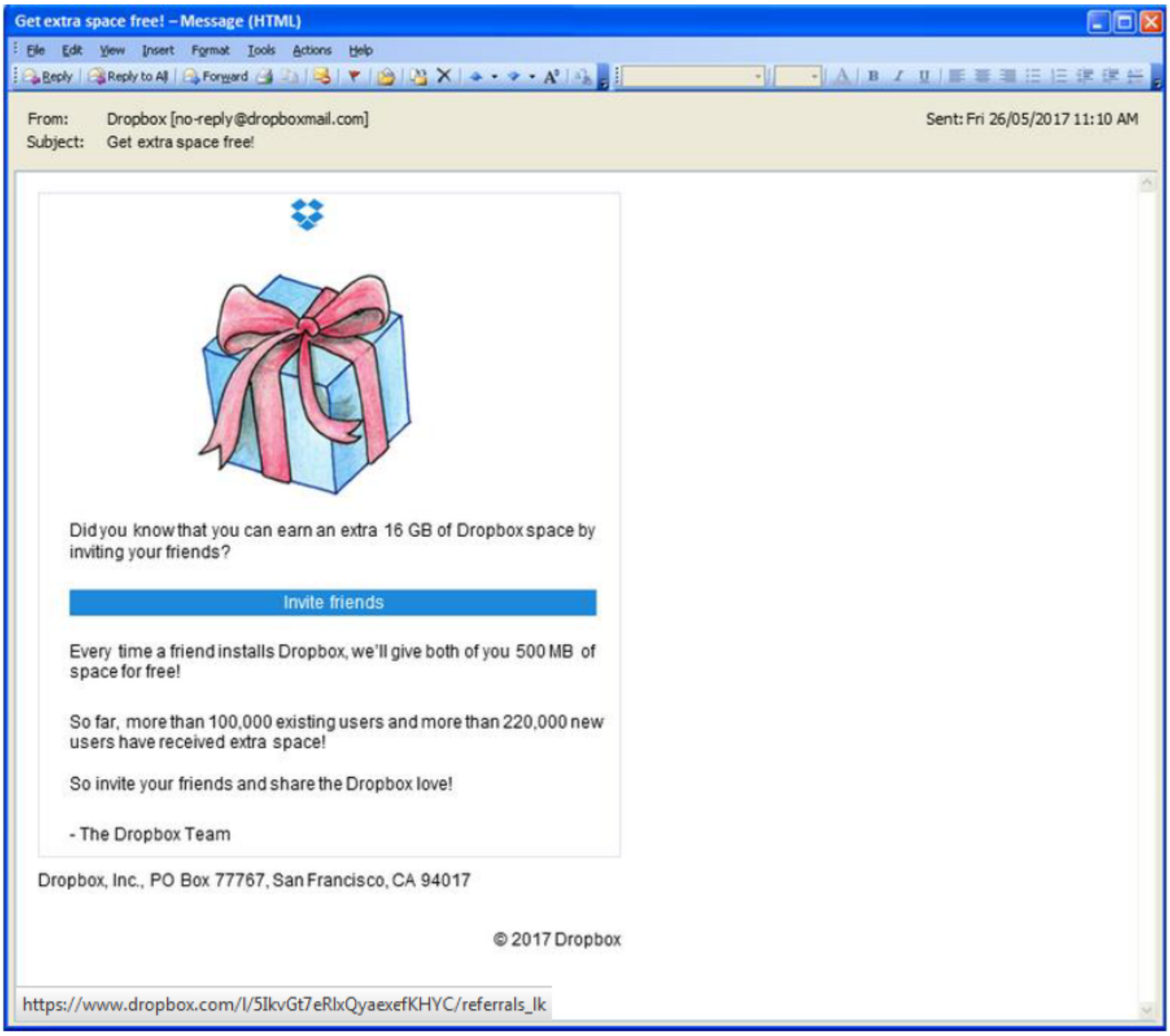
Reset password

Thanks,

Apple Customer Support

pastehtml.com/view/crhc6tc8f.html

Figure 1. Example of an image of a phishing email used in the study.



Get extra space free! – Message (HTML)
File Edit View Insert Format Tools Actions Help
Reply Reply to All Forward

From: Dropbox [no-reply@dropboxmail.com]

Subject: Get extra space free!

Sent: Fri 26/05/2017 11:10 AM

Did you know that you can earn an extra 16 GB of Dropbox space by inviting your friends?

Every time a friend installs Dropbox, we'll give both of you 500 MB of space for free!

So far, more than 100,000 existing users and more than 220,000 new users have received extra space!

So invite your friends and share the Dropbox love!

- The Dropbox Team

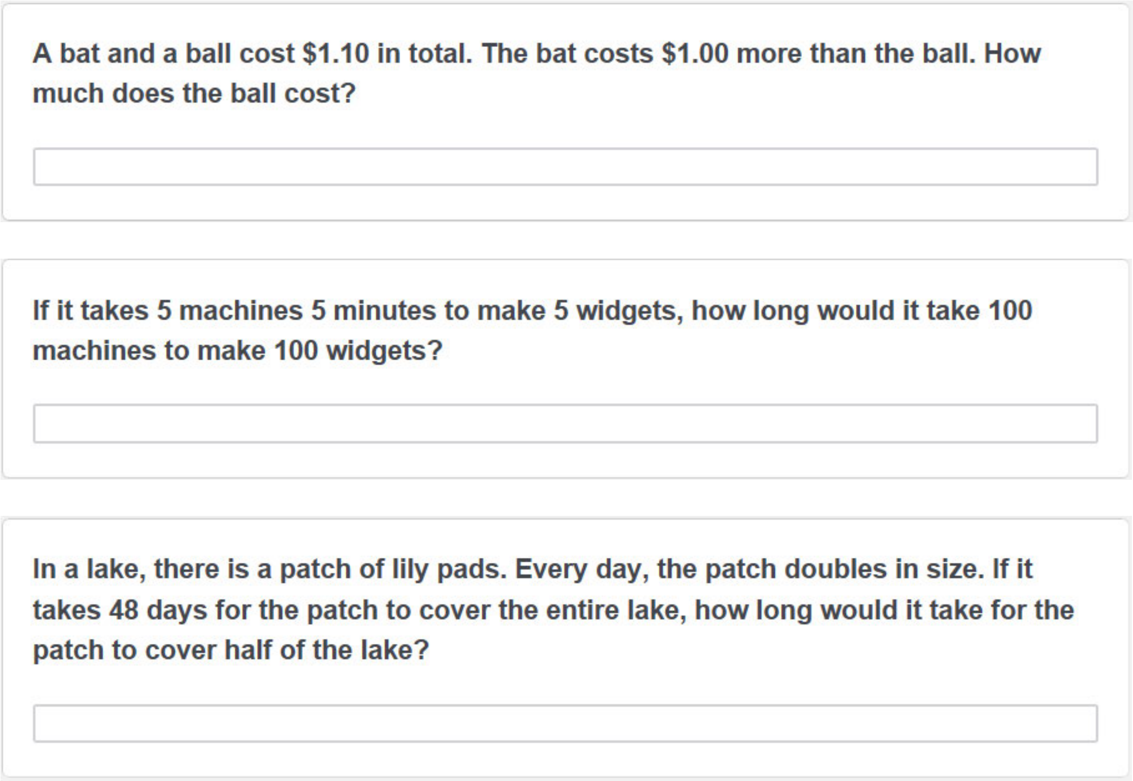
Dropbox, Inc., PO Box 77767, San Francisco, CA 94017

© 2017 Dropbox

https://www.dropbox.com/l/5IkvGt7eRlxQyaexefKHYC/referrals_lk

Figure 2. Example of an image of a genuine email used in the study.

Appendix D: Information Processing Manipulations

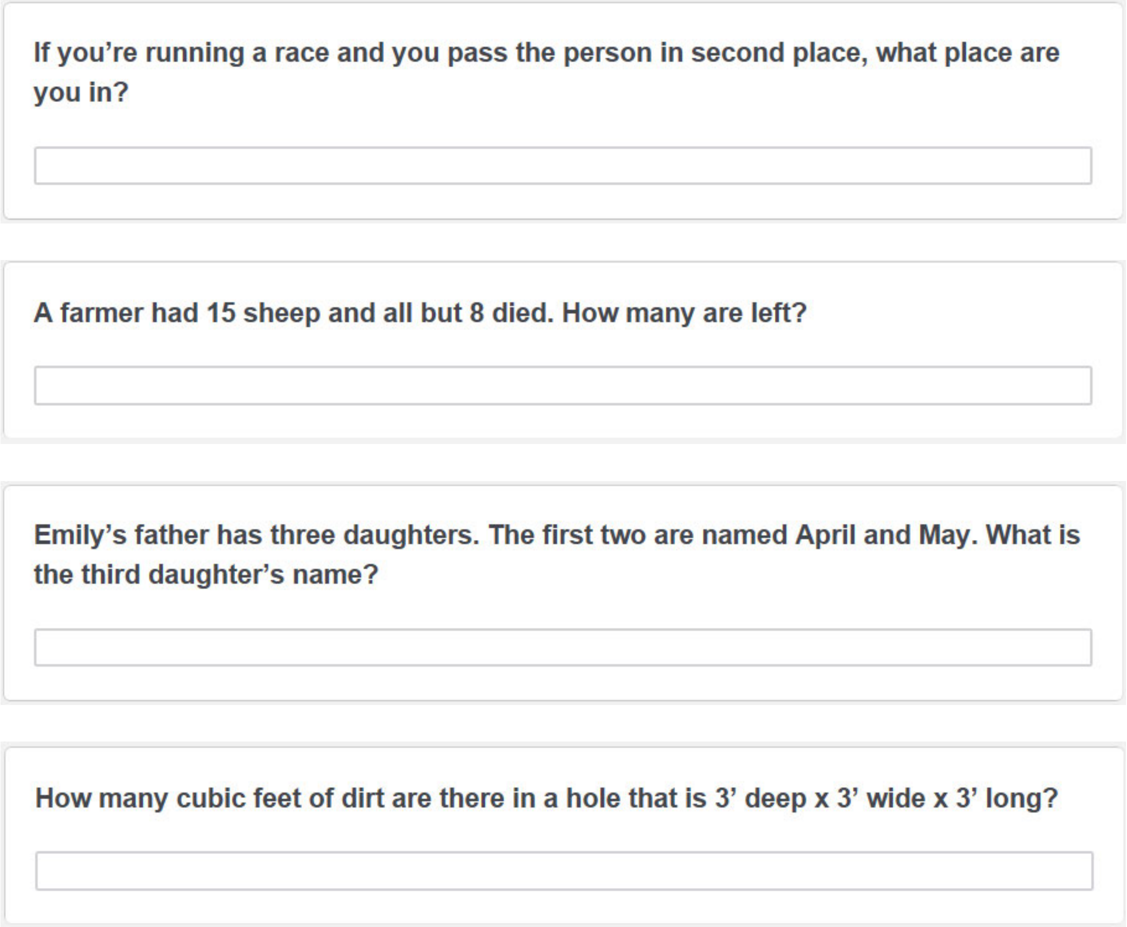


A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?

If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Figure 1. The Cognitive Reflection Test (CRT).



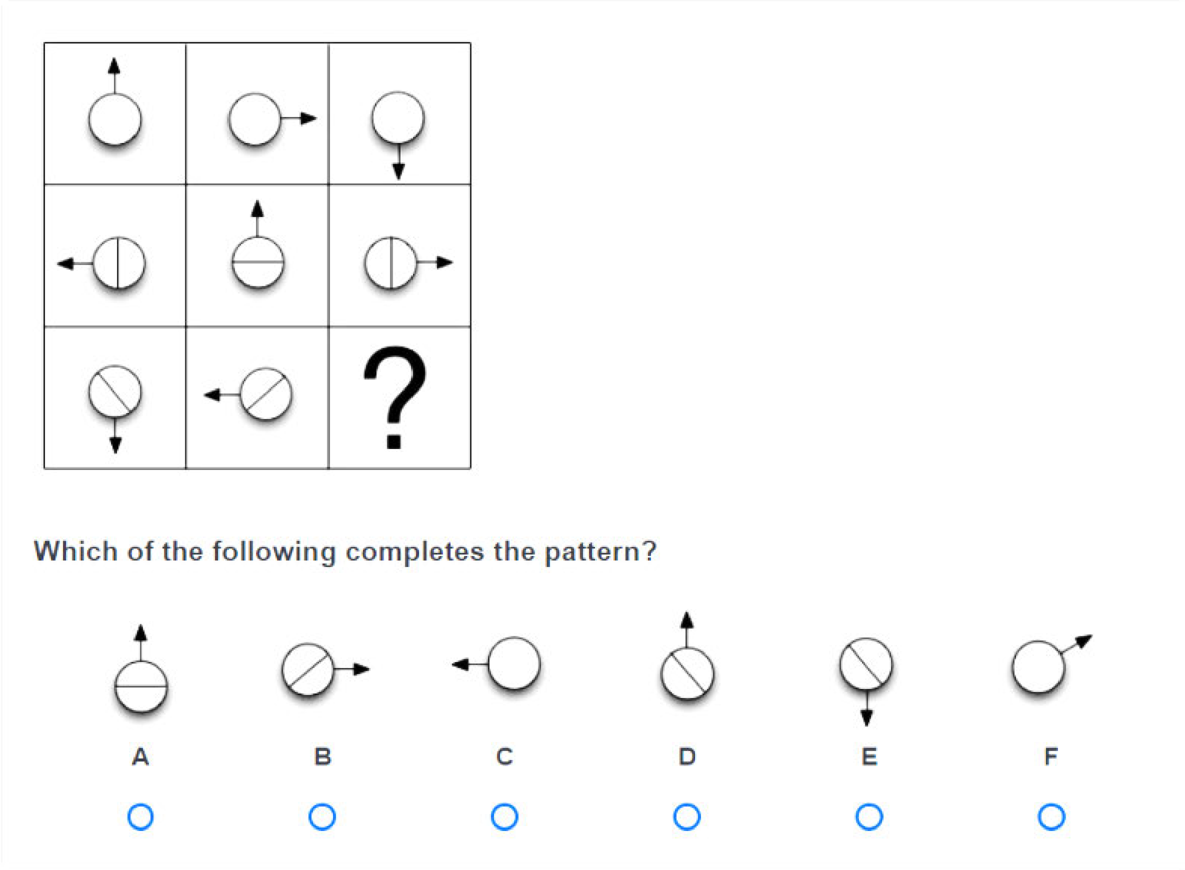
If you’re running a race and you pass the person in second place, what place are you in?

A farmer had 15 sheep and all but 8 died. How many are left?

Emily’s father has three daughters. The first two are named April and May. What is the third daughter’s name?

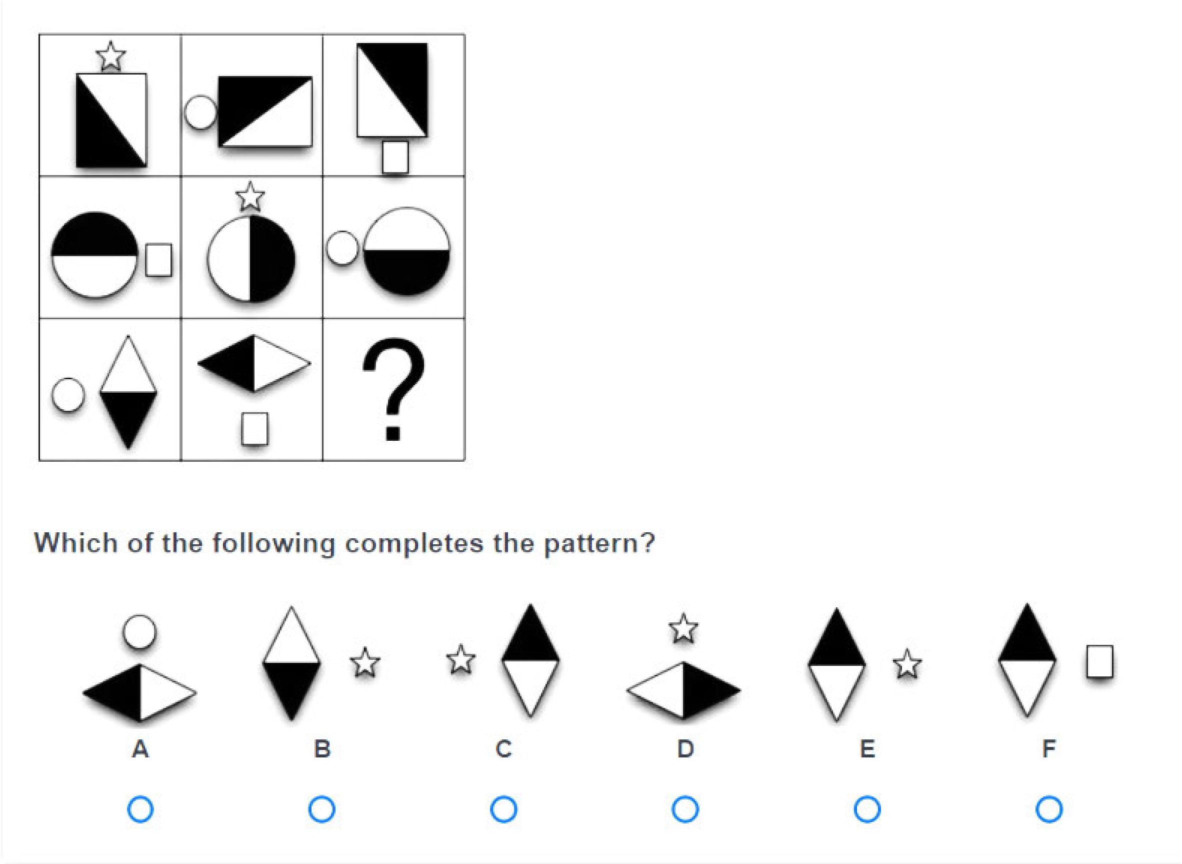
How many cubic feet of dirt are there in a hole that is 3’ deep x 3’ wide x 3’ long?

Figure 2. The CRT-2.



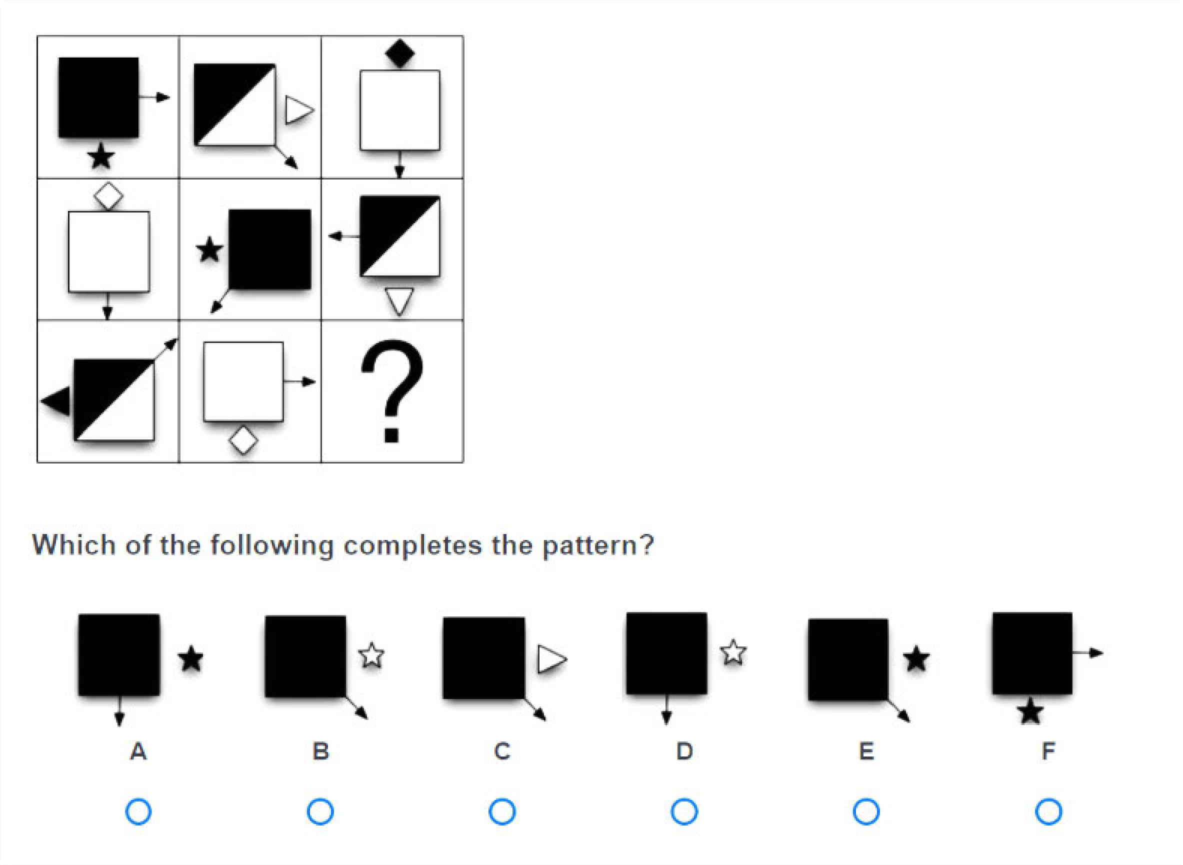
Which of the following completes the pattern?

A ○ B ○ C ○ D ○ E ○ F ○



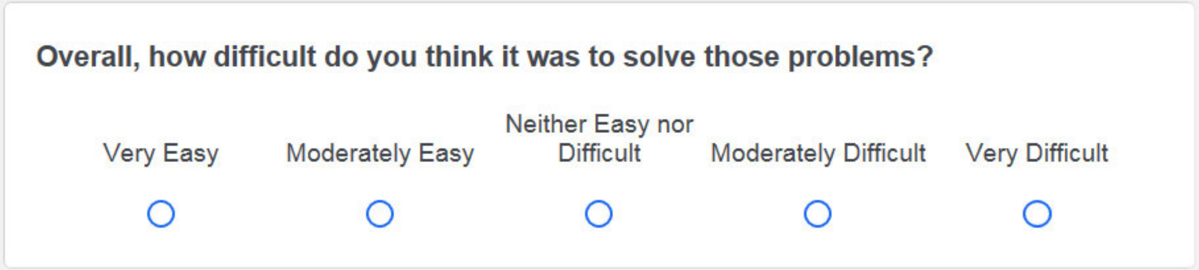
Which of the following completes the pattern?

A ○ B ○ C ○ D ○ E ○ F ○



Which of the following completes the pattern?

Figure 3. The Matrix Reasoning Task (MRT).



Overall, how difficult do you think it was to solve those problems?

Very Easy	Moderately Easy	Neither Easy nor Difficult	Moderately Difficult	Very Difficult
○	○	○	○	○

Figure 4. Measure of task difficulty.