Decision Support Systems xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Decision Support Systems

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

Press accept to update now: Individual differences in susceptibility to malevolent interruptions

Emma J. Williams [a], Phillip L. Morgan [b,*], Adam N. Joinson [a]

[a] School of Management, University of Bath, Claverton Down, Bath, UK
[b] Department of Health and Social Sciences, Psychological Sciences Research Group, University of the West of England (UWE) - Bristol, Frenchay Campus, Bristol, UK

ARTICLE INFO

Article history:
Received 5 May 2016
Received in revised form 12 January 2017
Accepted 24 February 2017
Available online xxxx

Keywords:
Influence techniques
Malevolent interruptions
Fraudulent messages
Individual differences
Cyber-security

ABSTRACT

Increasingly, connected communication technologies have resulted in people being exposed to fraudulent communications by scammers and hackers attempting to gain access to computer systems for malicious purposes. Common influence techniques, such as mimicking authority figures or instilling a sense of urgency, are used to persuade people to respond to malevolent messages by, for example, accepting urgent updates. An 'accept' response to a malevolent influence message can result in severe negative consequences for the user and for others, including the organisations they work for. This paper undertakes exploratory research to examine individual differences in susceptibility to fraudulent computer messages when they masquerade as interruptions during a demanding memory recall primary task compared to when they are presented in a post-task phase. A mixed-methods approach was adopted to examine when and why people choose to accept or decline three types of interrupting computer update message (*genuine*, *mimicked*, and *low authority*) and the relative impact of such interruptions on performance of a serial recall memory primary task. Results suggest that fraudulent communications are more likely to be accepted by users when they interrupt a demanding memory-based primary task, that this relationship is impacted by the content of the fraudulent message, and that influence techniques used in fraudulent communications can over-ride authenticity cues when individuals decide to accept an update message. Implications for theories, such as the recently proposed Suspicion, Cognition and Automaticity Model and the Integrated Information Processing Model of Phishing Susceptibility, are discussed.

1. Introduction

Due to the burgeoning proliferation of communicative and network-enabled technology, the likelihood of being interrupted by a computer-based update, advertisement or message has never been so high (e.g., [27,38]). We often take it for granted that such updates will occur and their common use in software update processes means that the majority of these communications are likely to be considered legitimate [4]. However, fraudulent computer-based messages continue to proliferate, exploiting common influence techniques to increase the likelihood that people will click on malicious links or downloads [1]. These techniques include instilling a sense of urgency in recipients, mimicking reputable institutions or familiar communications, and using the threat of loss in their communications, such as account closure or system shut down [6,36,44].

The majority of computer-based influence techniques rely on well-documented heuristics and biases present in human decision-making [19], such as the tendency to consider communications to be truthful rather than deceptive [21] and to make judgements based on emotional responses such as fear or panic (known as the *affect heuristic*). However, the extent that such forms of heuristic processing impact response behaviour across individuals and contexts remains uncertain. Understanding the contextual and individual factors that enhance vulnerability to fraudulent communications, including how these factors may interact, is the primary aim of this paper and is vital if targeted mitigations, such as training programmes, organisational procedures, and decision-support systems, are to be developed.

In the current paper, we consider recent theories and models regarding online trust and decision making to examine factors that impact on response behaviour to interruptive computer updates of varying degrees of malevolence. This includes the recently proposed Suspicion, Cognition and Automaticity Model relating to judgements of phishing e-mails (SCAM; [42]), the Staged Model of Trust [34], and the Integrated Information Processing Model of Phishing Susceptibility [41]. Specifically, we extend concepts within these models to judgements of computer update messages when they occur as interruptions during a demanding serial recall task compared to judgements made in a post-task questionnaire phase. By requiring participants to make judgements in two

* Corresponding author at: Department of Health and Social Sciences, University of the West of England (UWE) - Bristol, Frenchay Campus, Bristol, UK.
E-mail address: Phil.Morgan@uwe.ac.uk (P.L. Morgan).

http://dx.doi.org/10.1016/j.dss.2017.02.014
0167-9236/

different contexts where heuristic and systematic processing styles are likely to be differentially invoked, the relationship between message content, individual differences and processing strategy can be explored.

1.1. Theoretical background

Research examining what makes people susceptible to malicious influence in online environments has focused primarily on phishing e-mails and e-commerce environments [36,41]. Overall, findings suggest that people use particular informational cues, such as the message source and inaccurate spelling or grammar to determine message legitimacy [8,17], with factors such as degree of understanding of the internet contributing to individual differences in susceptibility [8,16]. Attempts to understand how fraudulent online messages affect response behaviour have led to the development of phishing susceptibility models such as the SCAM [42], which are based on existing theories regarding how information is processed within cognitive systems (e.g., Heuristic-Systematic Model, [9,39]). The SCAM suggests that heuristic processing is a crucial factor in susceptibility to fraudulent e-mails, with people who engage in more 'automatic' forms of message processing being less likely to notice errors or inconsistencies within the message and instead responding to other aspects of message content such as the influence techniques used [41]. These heuristic forms of processing are considered to be the predominant processing mechanism, due to the relative ease with which they are invoked [46]. Within the SCAM, individuals are considered as more likely to engage in such heuristic forms of processing if they are less suspicious of received messages, potentially due to a lack of knowledge or erroneous beliefs regarding risks of online environments, whereas increased suspicion is reflected in more systematic processing styles. However, the extent that such processing styles may be differentially relied upon across individuals, contexts and message types has yet to be systematically examined and elucidating this is a primary aim of the current study.

The failure of decision-makers to adequately direct attentional resources and elaborate sufficiently on inconsistencies in the message source is addressed in the Integrated Information Processing Model of Phishing Susceptibility [41]. This proposes that the majority of phishing e-mails are peripherally processed, with individuals focusing on the presence of influence cues such as urgency within the message content, at the expense of authenticity information, such as sender details. Parameters of this model were investigated using a simulated phishing attack on 325 university students and it was found that direction of attention was a primary factor in responding to phishing e-mails, with attention to the e-mail source, and spelling and grammar, suggested to reduce susceptibility. However, attention to urgency cues increased it due to such information monopolising available cognitive resources [41]. Similarly, when considering the credibility of e-health websites, Sillence et al. [34] found that individuals rely on an initial heuristic screening of relatively superficial factors, such as design appeal, when making decisions, reflecting an initial trust of information. The extent that such message factors impact response behaviour across different individuals, and the cognitive contexts they are operating in, however, is currently unclear.

The depth of processing that an individual engages in when faced with a fraudulent message may be impacted by individual differences in factors such as trust and self-control. People have been found to generally trust information in their surrounding environment unless they have a specific reason to doubt its legitimacy [3,6]. This 'truth bias' is well established in the inter-personal deception literature and has recently been expanded in Truth Default Theory [21], which suggests that for this default truth state to be temporarily abandoned, trigger events, such as a projected motive for deception, incoherent message content, or cues associated with dishonesty, are required. In order for this to occur, however, individuals must first notice these trigger events, a process that may not necessarily occur (a) when individuals are operating under cognitive pressure, due to a reliance on heuristic processing, or (b) to an equal extent across individuals. This possibility is addressed within the current paper.

People have been found to vary in their propensity to trust others [22], with dispositional trust tentatively linked with the ability to accurately differentiate legitimate and phishing e-mails [14]. In line with phishing susceptibility models (e.g., [41,42]), it is possible that trust decreases the likelihood of identifying inconsistencies within messages due to a failure to direct attention to authenticity information, or, because of inconsistencies attributed to other causes. For instance, when presented with photographs of faces, older adults have been found to be less adept at identifying cues of dishonesty; a finding that is suggested to account for their increased trust and resultant susceptibility to fraud [4]. Although such findings suggest that dispositional trust will influence susceptibility to fraudulent computer messages, scenario-specific trust related to online communication may have a greater impact on actual susceptibility [43] and these relationships are explored in the current study.

Resisting influence attempts is also deemed to be a difficult task requiring a degree of cognitive effort in regulating behaviour. As a result, traits associated with compulsive behaviours, such as low self-control and impulsivity, have been suggested to enhance susceptibility to influence techniques, due to a lack of systematic processing of message content and a failure to consider potential consequences prior to responding [12,40]. Understanding the impact of individual differences in self-control, impulsivity and trust on response behaviour, and the extent that these impacts may differ according to the cognitive context experienced, is a further aim of this paper.

Finally, it is fundamental to note that the appearance of a fraudulent computer update is likely to interrupt individuals who are already engaged in a primary task. Yet, the relative impact of being interrupted by fraudulent messages on subsequent response behaviour has, to our knowledge, not yet been examined. Although current research suggests that heuristic processing increases susceptibility to such communications, the extent that these processes are invoked across individuals, messages and contexts is less clear. Therefore, systematic investigation of the impact of cognitive context and individual differences on response behaviour to various fraudulent messages is required in order to further develop current theoretical approaches. Since the majority of fraudulent messages mimic existing organisations in order to appear more persuasive [5,36], the current study focuses on exploring the potential interaction between both cognitive context (i.e., the likely information processing strategy invoked) and individual differences, and so-called 'authority' influence techniques, namely whether the absence of authority information is more likely to trigger suspicion in users than the presence of errors within such information.

2. The current study

In the current study, we undertake exploratory research that aims to extend previous theory by examining the relationship between individual differences, cognitive context, and message factors to further understand response behaviour to fraudulent computer updates. Specifically, we utilise a task interruption approach that is known to be cognitively demanding (e.g., [28]), whereby participants complete a serial recall working memory task and are interrupted during this task by computer updates of varying degrees of authenticity purporting to require critical action. Serial recall tasks typically involve trying to remember a sequence of items, usually six to nine numbers, letters, or both and place a high demand on verbal phonological working memory (see [2]). During the task, participants must respond to occasional interruptions by computer update messages that contain either (a) genuine authority cues (i.e., designed to appear to be from a genuine authority source and do not contain any errors or inconsistencies), (b) mimicked authority cues (i.e., mimic an authority source but contain errors) or (c) no authority information (i.e., no details regarding the message source). In addition to response behaviour, this design also provides a unique

Please cite this article as: E.J. Williams, et al., Press accept to update now: Individual differences in susceptibility to malevolent interruptions, Decision Support Systems (2017), http://dx.doi.org/10.1016/j.dss.2017.02.014