Decision Support Systems 96 (2017) 119–129

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 27 February 2017

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

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 mechanism to identify the relative disruption to task performance of different fraudulent message cues, thus aiding understanding of the cognitive mechanisms involved. Following the serial recall phase, participants respond to the same messages in a questionnaire phase, whereby there are no additional cognitive demands and therefore participants will likely have more cognitive resource available to process information. This design allowed us to directly examine the relative impact of authority information on the response behaviour of individuals when they are operating in an environment that induces cognitive pressure, and therefore are more likely to use heuristic processing strategies, compared to when no additional cognitive demands are placed on them.

2.1. Hypotheses

According to previous theoretical work regarding susceptibility to deceptive messages [41,42], the presence of urgency and loss influence techniques within computer update messages, combined with the pressure of returning to, and continuing with, a demanding primary task, should result in recipients responding to messages without considering the possibility that they are fraudulent. Instead, they should be more likely to engage in simple heuristic processing that defaults to a trusting stance [21,42] and thus will likely fail to notice differences in authenticity cues. When individuals have more time to process messages and are under less cognitive pressure, however, potential inconsistencies may be more likely to be noticed, leading to illegitimate messages being declined. However, if the presence of authority information is still capable of overriding the presence of errors, then fraudulent messages that mimic ‘authority sources’ should still be more difficult to identify than those that do not contain any authority information, even under optimal processing conditions. Finally, we know from previous literature that differences in time spent dealing with an interrupting task impacts on the degree of disruption shown when individuals resume the primary task [13,26]. Since fraudulent messages contain cues that may lead recipients to question the message legitimacy [8,17,41], if these cues are noticed then the time required to respond to these messages may differ from genuine messages. This could either be reflected in longer processing times of fraudulent messages to consider potential authenticity cues, or alternatively faster processing times due to such messages being disregarded. These response time differences should also be reflected in a corresponding impact on subsequent primary task performance.

H1. If increased cognitive complexity makes recipients more susceptible to fraudulent messages due to the use of heuristic processing strategies, it is predicted that:

a) There will be no difference in response choice between genuine and mimicked or low authority messages during the serial recall task, due to increased cognitive pressure leading to a failure to identify inconsistencies in message content.
b) There will be no difference in performance on the serial recall task between genuine and mimicked or low authority message interruptions, due to all messages being processed to an equal extent (i.e., heuristically) and therefore having an equal impact on the resumption of the primary task.
c) When participants have unlimited time to inspect the content of messages, mimicked and low authority messages will be declined significantly more than genuine messages, due to a higher degree of systematic processing aiding the identification of inconsistencies.

The extent that participants are susceptible to responding to fraudulent messages heuristically may also be impacted by individual differences in dispositional trust and self-control [12,23,24,32]. People who show a high propensity to trust may consider messages as more likely to be genuine [14] and therefore will be less likely to actively search for, or evaluate, inconsistencies that suggest a message is illegitimate. Previous literature in relation to influence also suggests that participants with lower levels of self-control, and higher levels of impulsivity and sensation-seeking, will be more susceptible to negative urgency-based influence techniques, due to an increased likelihood of responding before potential inconsistencies have been identified [12,40]. However, whether such differences will have an additional effect in scenarios where participants are already likely to be using heuristic processing styles (i.e., when under a high degree of cognitive pressure), is currently unknown.

H2. If individual differences in dispositional trust and self-control influence the extent that messages are processed using heuristic processing strategies, it is predicted that:

a) Dispositional trust will be positively related to accept behaviour for mimicked and low authority messages due to a lower likelihood of identifying inconsistencies.
b) Self-control will be negatively related to accept behaviour for mimicked and low authority messages, such that those low in self-control will be less able to inhibit a potential accept response when faced with persuasive messages compared to those higher in self-control.

Finally, the current paper also explores participants' self-reported reasons for choosing to accept or decline update messages using a post-task questionnaire. This will allow the role of influence techniques, authenticity cues, and factors such as routine behaviour and knowledge, to be explored using a qualitative methodology. Factors included in the SCAM [42] suggest that the presence of influence techniques may override authenticity cues in participant decision making and this may combine with themes related to participant knowledge, perception of threat (both in relation to accepting and declining updates), and their usual norms of behaviour, to impact on decision making. We anticipate that themes identified in qualitative analysis of participant reasons for their chosen response will highlight these factors as being a potential mechanism for both accepting and declining updates.

3. Method

3.1. Participants

87 participants were recruited to participate in an experiment concerning decision making during complex tasks via the Undergraduate Psychology Participant Pool operated by the University Psychology Group. For measures of the disruption caused by interruptions, this number was adequate to detect a small to medium effect size (Cohen's $f = 0.1–0.25$) with power of 0.8 (determined using G*Power 3.1.7 software; [10]). Seventy participants were women and 17 men. Participants had a mean age of 18.56 ($SD = 1.82$) and were tested in a computer laboratory and provided with both written and oral instructions presented by the experimenter.

3.2. Design

A repeated-measures design involved all participants completing the same computer-based task and post-task questionnaires. Within the main experiment, there were 36 serial recall trials, and nine of these were interrupted by messages that required a response (either accept or decline). Each of these interrupting messages was one of three types: genuine authority, mimicked authority, and low authority, and each participant had three instances of each message type. Further details regarding message type are provided in the Materials and Procedure Section. Dependent variables included the number of to-be-remembered/TBR items recalled in the correct serial order (maximum of 9 items per trial/condition) and the proportion of genuine, mimicked and low authority interrupting messages that were accepted (maximum of 3 per condition).

3.3. Materials and procedure

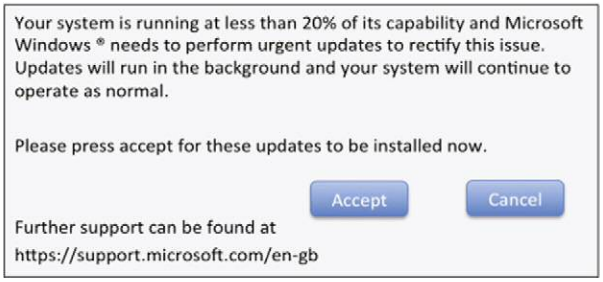
There were two experiment phases. The first involved participants completing the serial recall task whilst being exposed to computer update interruptions. The second involved completing a computer-based questionnaire that included re-evaluation of update messages as well as self-report measures of self-control, trust, impulsivity, and sensation seeking. These phases and specific materials are described below.

3.3.1. Phase 1 tasks

The primary serial recall and secondary interrupting message tasks were programmed using PsychoPy (http://www.psychopy.org: [32]), an open-source experiment generation software package. The serial recall task involved the presentation of a string of nine letters and numbers in the center of the screen simultaneously for 9-s in an Arial size 24 font presented in white on a dark grey background. Each number and letter string was designed to mimic a mock National Insurance (NI) number (equivalent to a US Social Security Number). After 9-s the letter/number string disappeared and following a 2-second retention interval a message reading 'enter code' appeared on the screen and remained there for 10-s. During this 10-second period, participants needed to write in an answer booklet as many numbers and letters that they could recall in the order in which the information was originally presented. For any numbers or letters that they could not remember, they could leave a blank space or make a guess. After the 10-second recall period, participants had to 'press spacebar to start next trial.' Once the spacebar was pressed, there was a 500-msec interval before the next trial started. A total of 36 trials were used, each using a different letter and number string. Number and letter strings were randomly generated prior to the experiment, and the same number and letter strings were used for each participant. The same format was used for all number and letter strings across all serial recall trials.

Of the 36 trials, nine contained an interruption message, representing 25% of all trials in order to reduce the frequency and predictability of interruptions for participants [29]. Interrupting messages consisted of system security-related update messages appearing in the center of the screen measuring approximately 150 mm × 60 mm. These occurred following the 2-second retention-interval (after the letter/number string had disappeared) but before participants were able to start recalling the string (i.e., before the 'enter code' instruction). Participants were instructed never to start writing responses until the enter code instruction appeared and were to stop writing when it disappeared. The interrupting message remained on the screen until the participant chose to either decline it by pressing 'c' on the keyboard or accept it by pressing 'a' on the keyboard. Following an 'a' or 'c' response, the update disappeared and was replaced by the enter code instruction. As on non-interruption trials, participants were then required to write the letter/number string within the answer booklet. The order of trials was initially randomised and this same order was then used for all participants.

Of the nine interruption trials, three update messages were designed as genuine authority update messages (Fig. 1), three as mimicked authority update messages (Fig. 2) and three as low authority update messages (Fig. 3). Genuine authority messages were designed to contain specific details linked to recognised expertise, organisations and software manufacturers (e.g., accurate programme reference, presence of a copyright symbol and genuine website link). Mimicked authority messages were designed to contain the same level of detail but provided an inaccurate reference (spelling error), an inaccurate website link and lacked a copyright symbol. These messages were designed to mimic the communications of legitimate software or organisations, as commonly found in the techniques of scams that mimic trusted institutions or entities. Low authority messages were designed to contain no specific details relating to the sender of the communication (e.g., programmes or applications) and no website link. Therefore, the message contained no references to potential authority sources.


Your system is running at less than 20% of its capability and Microsoft Windows ® needs to perform urgent updates to rectify this issue. Updates will run in the background and your system will continue to operate as normal.

Please press accept for these updates to be installed now.

Further support can be found at
https://support.microsoft.com/en-gb

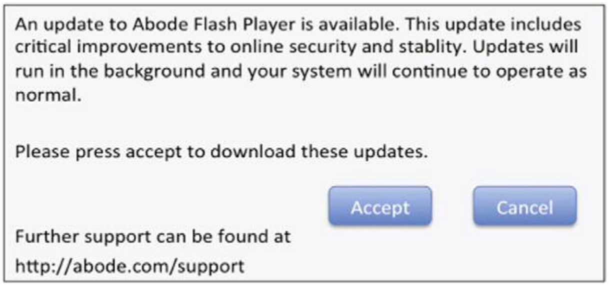
Fig. 1. Example genuine authority update.

Updates were matched as best as possible to ensure that they were of a similar length whilst retaining a sense of realism in different scenario content across each message. The mean word length within updates was 50.1 words (excluding website links), with the minimum number of words = 34 and the maximum = 71 (Genuine Authority $M = 54.7$; Low Authority $M = 53.7$; Mimicked Authority $M = 44.3$). All updates included the same influence techniques (urgency and impact), requiring an urgent response to counter a purported threat, and focused on anti-virus, spyware, programme critical fixes, and expiry of licenses.

Comprehensive written instructions were provided prior to the task and were read aloud by the experimenter. As well as instructing participants on how the tasks would function and what they needed to do to make responses and when (i.e., only following the enter code instruction), the experimenter also emphasised that the letter/number and the security update tasks were of equal importance. Participants completed two practice trials, one containing a genuine themed interrupting update message and one without an interruption. Prior to the start of the computer task, participants were also presented with a brief message displaying a generic system security message for a 15-second period. This was based on system security log-on messages that are often used by organisations to ensure that employees are aware of cyber-security policy when logging on to internal systems and networks. The message in the current experiment contained the following information:

'This system is protected by virus protection software and updates are installed on a regular basis. However, please be vigilant about the security of this system by ensuring that any attempts by applications to access system information or data are legitimate. Some applications may attempt to access system information and personal or organisational data on this device with or without authorisation. The virus protection software installed on this system will make every attempt to destroy any such threats'.

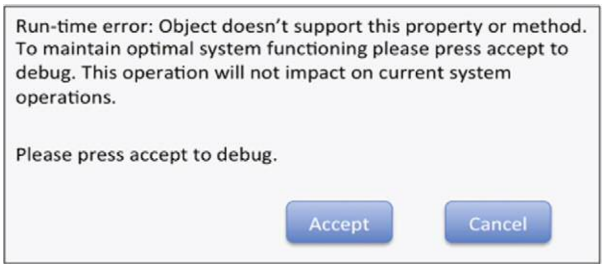
This message was presented to make participants aware that although their computer system was protected it was still vulnerable to attack via fraudulent communications or malware attempting to infiltrate the system. It was expected that whilst participants would read this message, they would likely differ in the extent that the information was processed. Following the computer task participants were tested


An update to Abode Flash Player is available. This update includes critical improvements to online security and stablity. Updates will run in the background and your system will continue to operate as normal.

Please press accept to download these updates.

Further support can be found at
http://abode.com/support

Fig. 2. Example mimicked authority update.



Run-time error: Object doesn't support this property or method. To maintain optimal system functioning please press accept to debug. This operation will not impact on current system operations.

Please press accept to debug.

Fig. 3. Example low authority update.

on: (1) whether they remembered reading the message; (2) message content; and (3) whether they felt that they took the message seriously. Following this, they were also asked to judge the degree to which they abided by the message content, i.e., 0% of the time to 100% of the time. In total, Phase 1 took approximately 15–20 min to complete.

3.3.2. Phase 2 tasks

Following Phase 1, participants were again presented with each of the 9 updates/interruptions via the Qualtrics online survey platform (www.qualtrics.com). Participants could spend as long as they wished reading the update before answering the following questions: Would you ordinarily accept or decline this message? (Accept or Decline); Why? (open-ended response). This unlimited response time, combined with providing reasons for their decision, would likely maximise the likelihood that systematic processing strategies are used, making participants more aware that some of the messages may not be trustworthy and thus impacting their likely information search strategies. They were then asked a series of questions related to cyber security awareness, which included: 'To what extent do you trust communications from your computer system, such as security updates, in general?'; 'How confident are you in your ability to differentiate genuine communications from scam communications in daily life?'; 'How would you rate your awareness of the common techniques used in scams?' and 'To what degree did you trust the system to deal with malicious attacks?' These four questions had Likert response scales ranging from 1 to 7. Participants were also asked 'Have you previously been the victim of a computer-based/online scam?' with a 'yes'/'no' response.

Next, participants completed a series of self-report questionnaire measures to examine individual differences in factors that might have contributed to how they chose to handle interrupting update messages. The Brief Self-Control Scale is a 13-item self-report measure designed to measure trait self-control and the Cronbach's alpha for this questionnaire has been reported as 0.89 (BSCS; [37]). Participants respond on a 5-point Likert Scale to questions such as 'I am good at resisting temptation'. The NEO Trust facet is a 10-item self-report measure designed to measure generic propensity to trust and the Cronbach's alpha for this questionnaire has been reported as between 0.82 and 0.88 on internet and field samples [18]. Participants respond on a 5-point Likert Scale to questions such as 'I believe that others have good intentions' (NEO-T; www.ipip.org). The Barratt Impulsiveness Scale is a 30-item self-report measure divided into 3 sub-scales: attentional impulsivity, non-planning impulsivity and motor impulsivity. It is designed to measure trait impulsivity and the Cronbach's alpha for this questionnaire has been reported as 0.82 in testing with undergraduate students (BIS; [30]). Participants respond on a 4-point Likert Scale to questions such as 'I concentrate easily'. Finally, the Brief Sensation-Seeking Scale is an 8-item self-report measure designed to measure trait sensation-seeking and the Cronbach's alpha for this questionnaire has been reported as 0.74 (BSSS; [15]). Participants respond on a 4-point Likert Scale to questions such as 'I get restless when I spend too much time at home'. In total, Phase 2 of the experiment took approximately 20 min to complete. Following this, participants were fully debriefed.

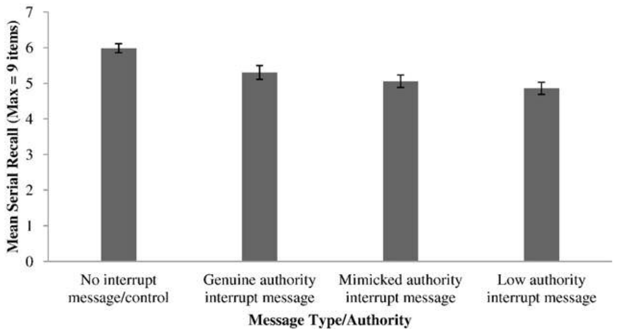


Fig. 4. Effect of interrupt message type on serial recall memory. Note. Error bars represent ± standard error.

4. Results and discussion

All tests are two-tailed with alpha levels of 0.05. Effect sizes were determined using Cohen's *d* for *t*-tests (with 0.2, 0.5 and 0.8 indicating small, medium and large effect sizes respectively) and Cohen's *f* for F tests (with 0.1, 0.25 and 0.4 indicating small, medium and large effect sizes) [45].

4.1. H1: the impact of message authority and cognitive complexity on judgements

Table 1 illustrates the number of messages that were accepted when presented during the serial recall memory trials (*Max* 3 per condition) and during the questionnaire phase of the study (*Max* 3 per condition). A 2 (serial recall phase, questionnaire phase) × 3 (high authority, mimicked authority, low authority) factorial repeated measures ANOVA revealed significant main effects of phase, $F(1, 86) = 5.11$, $MSE = 2.46$, $p = 0.026$, $f = 0.24$, with messages more likely to be accepted in the serial recall phase than the questionnaire phase, indicative of a truth bias under heuristic processing conditions, and message authority, $F(2, 172) = 61.03$, $MSE = 0.49$, $p < 0.001$, $f = 0.84$, and a significant interaction, $F(2, 172) = 25.12$, $MSE = 0.47$, $p < 0.001$, $f = 0.54$. During the serial recall phase, whilst accepts for genuine and mimicked authority messages did not differ, low authority messages were accepted significantly less than both genuine and mimicked authority messages ($ps = 0.007$ and 0.038 respectively). Thus, Hypothesis 1a was only supported for mimicked authority messages. During the questionnaire phase, accepts significantly differed between all message types, with more genuine than low authority messages accepted and more genuine authority than mimicked authority messages accepted ($ps < 0.001$), supporting Hypothesis 1c. However, the difference in message accepts between the serial recall phase and questionnaire phase was only significant for low authority messages ($p < 0.001$), with fewer low authority messages accepted in the questionnaire phase (approximately 27% of messages accepted) compared to the serial recall phase (approximately 56%). The lack of a significant difference for mimicked authority messages is particularly striking given the greater likelihood of increased suspicion in the questionnaire phase due to the requirement to describe the decision process. This suggests that the lack of authority cues present in low authority messages are more likely to be noticed than errors in mimicked authority messages in both high and low cognitive pressure

Table 1
Number of authority messages accepted when presented during serial recall phase and questionnaire phase.

Message authority	Serial recall phase		Questionnaire phase	
	M	*SD*	*M*	*SD*
Low	1.68	1.25	0.82	0.95
Mimicked	1.89	1.23	1.65	1.04
Genuine	1.98	1.25	2.15	0.92

scenarios. However, participants were still significantly more likely to decline low authority updates during the questionnaire phase of the study, suggesting that greater systematic processing aids in the identification of these particular cues [41,42].

Serial recall memory performance was also considered for each of the four conditions (no interruption, low authority updates, mimicked authority updates, and genuine authority updates). A clear trend was found such that as the authority of interrupt messages reduced (from genuine to low), so did task performance. A repeated measures ANOVA, with a Huynh-Feldt correction applied due to a violation of sphericity, revealed a significant main effect of interrupting message authority, $F(2.77, 237.93) = 18.96$, $MSE = 1.20$, $p < 0.001$, $f = 0.47$. As would be expected, Bonferroni pairwise comparisons revealed that serial recall was higher in the no interruption condition than in the low authority, mimicked authority and genuine authority message conditions (*p's* < 0.001). However, serial recall memory was also higher in the genuine authority condition than the low authority condition, although this difference was marginally significant ($p = 0.05$). The low and mimicked authority conditions and the genuine and mimicked authority conditions did not differ significantly. Thus Hypothesis 1b was supported for mimicked authority messages only. However, this reflects the finding of lower message accepts for low authority messages during the serial recall phase. This suggests that low authority messages disrupt performance to a greater degree than messages that appear genuine due to the triggering of suspicion and the resultant requirement to invoke more systematic processing of message content, enabling further consideration of the lack of authority cues and hence whether the message should be declined due to potential illegitimacy [21,42] (Fig. 4).

The finding that low authority messages disrupt processing to a greater degree is also reflected in differences in message processing time. Whilst a similar amount of time was spent viewing interrupt messages in the genuine (*M* 5.47s) and mimicked authority conditions (*M* 5.37s), participants spent approximately half a second longer viewing low authority messages (*M* 6.00s) before responding in the serial recall phase. A repeated measures ANOVA, with a Huynh-Feldt correction applied due to a violation of sphericity, revealed that this time difference was significant, $F(1.77, 152.10) = 4.19$, $MSE = 2.71$, $p = 0.021$, $f = 0.22$, with Bonferroni pairwise comparisons revealing that significantly more time was spent viewing low authority messages than genuine authority messages ($p = 0.027$). Whilst more time was spent viewing low than mimicked authority messages, this difference was marginally non-significant ($p = 0.07$).

Taken together, these findings suggest that heuristic processing strategies were utilised during the serial recall phase of the experiment, resulting in errors contained within mimicked authority messages not being identified, and such messages being processed and accepted to the same degree as genuine authority messages [41]. Interestingly, the lack of authority information contained within low authority messages seemed to invoke deeper processing strategies that were both more time-consuming and had a greater impact on resultant task performance, likely due to the triggering of suspicion [21]. This suspicion, and the resultant processing differences, was reflected in more low authority messages being declined. Since authority cues have been highlighted as a key factor in determining online credibility [11], the lack of sender details in low authority messages likely resulted in participants taking longer to consider the possibility that these updates may not be genuine. This may lead to a greater degree of suspicion being invoked, reducing the degree of initial trust otherwise experienced in online communications [34].

4.2. H2: the relationship between individual differences and response behaviour

To examine whether individual differences related to trust and self-control increased the likelihood of accepting mimicked and low authority message updates across the two phases, Pearson bivariate correlations were conducted. Table 2 presents descriptive statistics for the questionnaire measures and Table 3 presents *r* values for all correlations. In contrast to Hypothesis 2a, a trend towards a negative relationship was found between trust and message accepts for mimicked authority messages within the serial recall phase ($r = -0.195$, $p = 0.07$), such that higher propensity to trust scores were related to a lower likelihood of accepting messages. Although the relationship was not significant for low authority messages ($r = -0.141$, $p = 0.19$), this result was also in a negative direction, and a significant negative relationship was also shown with genuine authority messages ($r = -0.141$, $p = 0.017$). This suggests that the presence of authenticity cues did not alter this relationship. Although this contrasts with the proposed relationship in Hypothesis 2a, it may be that higher propensity to trust leads to a reduction in accept behaviour due to the maintenance of system security via updates not being considered a priority. This may be due to a failure to consider the malevolent intentions of others who may attempt to infiltrate the system, reducing the perceived importance of maintaining up-to-date software. For instance, individuals high in dispositional trust may not consider the potential threat of operating online (i.e., they believe the intentions of other online actors are genuine). As a result, they are less likely to consider computer security to be a priority and therefore are also less likely to accept computer updates that maintain security, regardless of their apparent authenticity. Conversely, individuals who are more suspicious may prioritise accepting such updates in order to counter any potential threats that they perceive [22].

No significant relationships were found between any of the other individual difference measures and message accepts during either the serial recall phase or the questionnaire phase, so Hypothesis 2b was also not supported. It is possible that negative urgency influence techniques are able to impact on response behaviour regardless of the degree of trait self-control that a recipient has, such that individual differences are effectively overridden by such message factors. Future work should focus on clarifying this potential relationship between individual differences, such as self-control, and the presence of various influence techniques.

To examine whether a more context-specific conceptualisation of trust had a greater impact on accept behaviour for fraudulent messages, the relationship between participant responses on scenario-specific questions (e.g., trust in computer communications and trust in the system to deal with malicious attacks) and message accepts for mimicked and low authority messages was examined using Pearson bivariate correlations. Table 4 shows descriptive statistics for the scenario specific questions and Table 5 shows *r* values for all correlations. Significant positive relationships were found between trust in computer communications and message accepts in the questionnaire phase for both mimicked and low authority messages (mimicked: $r = 0.44$, $p < 0.001$; low: $r = 0.41$, $p < 0.001$), such that increased trust in computer communications was related to an increased likelihood of accepting both mimicked and low authority messages under more systematic processing conditions. This suggests that Hypothesis 2a is supported when context-specific conceptualisations of trust are invoked [43]. When participants were under more cognitive pressure in the serial recall phase, however, and therefore more likely to respond heuristically, this effect was only found for low authority messages. This differential relationship with low and mimicked authority messages was also demonstrated in the positive relationship between trust in the system and accepts for

Table 2
Descriptive statistics for questionnaire measures.

	Trust	Sensation seeking	Self-control	Impulsivity
Min	21	12	16	21
Max	43	38	58	103
M	34.15	25.74	39.59	64.16
SD	5.02	5.35	8.22	11.14

Table 3
Pearson bivariate correlations (r values) between questionnaire measures and response choice according to message type.

		Trust	Sensation seeking	Self-control	Impulsivity
Accepts SR phase	Genuine	−0.256*	−0.010	0.085	0.053
	Mimicked	−0.195	0.084	0.014	0.083
	Low	−0.141	0.021	−0.062	0.067
Accepts Q phase	Genuine	−0.110	−0.019	−0.032	−0.125
	Mimicked	−0.174	0.059	−0.092	−0.032
	Low	0.023	0.019	−0.172	−0.006

* Correlation is significant at <0.05 level (two-tailed).

Table 5
Pearson bivariate correlations (r values) between scenario specific questions and response choice according to message type.

		Trust in computer messages	Confidence in detecting	Awareness of scam techniques	Trust in system
Accepts SR phase	Genuine	0.121	0.137	−0.118	0.020
	Mimicked	0.168	0.106	−0.151	0.106
	Low	0.215*	0.107	−0.075	0.276*
Accepts Q phase	Genuine	0.282*	0.066	−0.227*	0.122
	Mimicked	0.439**	−0.059	−0.356**	0.037
	Low	0.414**	−0.083	−0.182	0.315**

* Correlation is significant at <0.05 level (two-tailed).
** Correlation is significant at <0.01 level (two-tailed).

low authority messages only across both phases (serial recall phase: $r = 0.28$, $p = 0.01$; questionnaire phase: $r = 0.41$, $p < 0.005$). This suggests that increased trust may increase susceptibility to fraudulent cues that others identify more easily (i.e., a lack of authority cues), even in contexts where systematic processing is more likely [4]. A lack of sender information could, therefore, either (a) be less likely to be noticed by more trusting individuals due to differences in attention direction processes, in line with heuristic defaults, or (b) not be considered a threat even if it is noticed, with differential threat perceptions resulting in a failure of identified information to 'trigger' suspicion in line with the mechanisms of Truth Default Theory [21]. Future work should focus on disentangling these possibilities.

In contrast to individual differences that may increase susceptibility, a significant negative correlation was found between self-reported awareness of scam techniques and message accepts for mimicked authority messages in the questionnaire phase ($r = -0.36$, $p < 0.001$), and a negative trend for low authority messages ($r = -0.18$, $p = 0.09$), such that greater awareness of scam techniques was related to a decreased likelihood of accepting fraudulent messages. This stronger relationship with mimicked authority messages suggests that awareness aids in the identification of more fraudulent messages that are more difficult to identify. However, this only occurs when the opportunity is provided to engage in more systematic processing strategies, with awareness having no impact on the likelihood of accepting fraudulent updates during the serial recall task. This suggests that any protective impact is overridden when participants are operating under a higher degree of cognitive pressure and therefore are more likely to rely on heuristic processing strategies.

4.3. Why do people accept or decline?

Open-ended responses in relation to *why* participants chose to accept or decline a particular update were included as the data set for qualitative analysis. Thematic analysis was used to identify the presence of themes that participants reported as influencing them to accept a message or to decline it during the questionnaire phase. A primarily theoretical approach was used when analysing the data, such that general themes of interest were identified prior to data analysis based on theoretical ideas [5,36,41,42]. These included:

(a) *Use of Influence Techniques* - reference to the influence techniques that were present in messages, namely, known and trusted organisations or programmes (authority), the requirement to undertake the action immediately (urgency) or loss of some form of functionality (loss).
(b) *Use of Authenticity Cues* - reference to authenticity cues, such as spelling errors, grammatical errors, poor design or layout, message inaccuracies or inconsistencies.
(c) *Routine* - reference to behavioural habits, such as always accepting or declining certain types of update.
(d) *Degree of Knowledge* - reference to a lack of knowledge, such as not understanding computer systems sufficiently to do anything other than their chosen action.

Table 4
Descriptive statistics of scenario specific measures.

	Trust in computer messages	Confidence in detecting	Awareness of scam techniques	Trust in system
Min	1	1	1	1
Max	7	6	7	7
M	3.91	4.23	4.24	4.59
SD	1.53	1.53	1.62	1.47

Additional themes or sub-themes were also identified during the analysis process and these primarily related to:

- *Context-specific* reasons, whereby elements of the wider context influence response choice. For example, the computer not appearing to be running slowly or having any issues that would mean the update is considered necessary at that point in time.
- *'Why not?'* as a motivator for response choice. For example, individuals considering that the update would have no negative impact on their current operations, and so not identifying a reason to decline it.

Since we were interested in understanding the specific aspects of each message that may have influenced participant decision making, the qualitative data set was divided into responses relating to accepting a message and those related to declining a message for each of the nine interrupting messages. Following familiarisation with the data, initial codes were produced and considered in relation to each data item. Coded data extracts were then considered in relation to the a priori themes described above and the formation of potential sub-themes. Themes and sub-themes were then reviewed and refined and the dataset revisited in relation to these. Counts of themes according to response choice (accept vs. decline) are shown in Table 6 and according to message type in Table 7 and the most common themes are discussed in more detail below.

4.3.1. The use of influence techniques

The ability of urgency cues to attract and monopolise attentional resources has been discussed in previous theoretical models of

Table 6
Proportion of themes when respondents accept a message compared to when they decline a message.

	Accept	Decline
Authenticity cues	28.2%	34.6%
Influence cues	60.5%	–
Failed influence cues	–	41.4%
Context cues	0.9%	7.9%
Routine cues	4.6%	6.7%
Knowledge cues	0.6%	9.3%
Why not?	5.2%	–
Total	100%	100%

susceptibility [41], with heuristic processing strategies considered to increase the likelihood that greater emphasis will be placed on influence-related message content at the expense of authenticity information. Linking communications with authority figures is an established influence technique that exploits the tendency for individuals to comply with authority requests and its use has been found to impact on the relative effectiveness of genuine malware warnings [25]. In line with this, a number of responses related to messages being from recognisable and known software providers or organisations, which were considered to signify a message as likely to be both legitimate and important. For example, *'I would trust that the update needs to be done, as it is from [ ] which I recognise and trust'*.

Similar to exploiting authority references, placing individuals under a time pressure and instilling a sense of urgency when making a decision is a technique that can be found in a number of scam communications [36]. In line with this, the presence of cues related to the time-critical nature of the request, suggesting that the message should be accepted in order to address an urgent issue, was also mentioned by a number of respondents. For instance, it *'indicates something urgent'* and *'because when I hear threats I instantly think that I need to fix it'*.

The concept of loss aversion is also well documented within the psychology literature [20], with individuals motivated to avoid potential losses in the future. In line with this, accepting a message in order to avoid a potential negative impact was often highlighted, particularly in relation to avoiding a loss of functionality or security. For example, the *'risk to computer if you don't'* or *'[if I don't] computer could crash mid way through something'*.

However, although influence techniques may be effective in motivating an individual to respond to a message, they can also fail to have the desired impact on the recipient, and these 'failed influence' attempts were present in a number of the reasons that participants gave for declining a message. In particular, this related to the message failing to persuade the recipient that the update was necessary or important, e.g., *'because it does not seem urgent',* lacking in sufficient detail or the recipient not understanding the message sufficiently to act upon it, for example *'vague message, doesn't tell me what it's doing'*.

4.3.2. The use of authenticity cues

When participants considered their response decisions, this was based on the questionnaire phase whereby more systematic processing strategies were likely to be used. As such, increased suspicion may have increased the likelihood of considering authenticity cues when making decisions. In accordance with this, reasons regarding response decisions often related to whether a message was likely to be genuine and the different factors that were used to judge legitimacy. Some of these authenticity cues related to tangible features within the message, such as the presence of a website link or a trademark symbol (or lack thereof), for example *'it supplies a link below so it seems more trustworthy'* and *'the [ ] has a little circle which means it is probably safe?',* whereas others related to more intuitive and subjective feelings of legitimacy, such as *'I don't trust it'*.

Table 7
Proportion of themes when respondents accept a message compared to when they decline a message according to message authority.

	Genuine authority		Mimicked authority		Low authority	
	Accept	Decline	Accept	Decline	Accept	Decline
Authenticity cues	36.5%	32.3%	23.8%	36.8%	12.7%	34.4%
Influence cues	57.1%	–	60.7%	–	70.4%	–
Failed influence cues	–	38.7%	–	42.4%	–	42%
Context cues	0.5%	11.8%	1.1%	4%	1.4%	8.5%
Routine cues	2.3%	12.9%	5.9%	8.8%	8.4%	8%
Knowledge cues	0.9%	4.3%	0.6%	8%	–	7%
Why not?	2.7%	–	7.7%	–	7%	–
Total	100%	100%	100%	100%	100%	100%

4.3.3. Routine

A small number of respondents also referenced their usual behaviour when responding to similar messages in the past, supporting elements of the SCAM [42] related to the role of habitual and routine behaviour on judgements. For instance, defaulting to accepting a message even if unsure about its meaning because that is how such messages are usually handled. Often such behaviour was linked to a lack of technical knowledge, highlighting the role of knowledge in perceived ability to cope with situations that may arise, with some habitual response behaviours appearing to emerge in response to a lack of technical understanding. For example, *'I don't know anything about computers so I just trust the little box'* and *'Know the phrase run-time error but wasn't sure what its asking for. So I do an 'old-man' move and accept it anyway'*.

Familiarity with the message content was also raised as a reason for accepting a message, with having seen a message before resulting in it being considered "normal" and therefore less likely to be fraudulent. For example, *'I see this all the time'* or *'It seems more usual and I've encountered similar messages before'*.

Alternatively, habitual behaviours (e.g., *'I don't accept pop*-ups') and lack of familiarity (*'Not familiar with the message'*) were also cited as a reason to decline updates, such as defaulting to declining any update or considering an unusual message as more likely to be suspicious. To our knowledge, the role of familiarity with message content on response behaviour is not addressed in current models and therefore warrants further examination in the future.

4.4. Comparing accept and decline decision themes

Table 6 shows the relative proportion of themes when respondents chose to accept the message and when they chose to decline it. When participants chose to accept a message, reasons referring to the influence techniques used were the most highly cited (60.5%), with authenticity cues the second most common (28.2%), suggesting that message factors related to authority, urgency and potential loss influence self-reported decision making to a greater degree than authenticity cues. When participants declined the message, primary reasons related to not considering the update as important or necessary, which we consider to be a failure of the influence technique to effectively persuade individuals to respond, potentially due to these participants having greater resistance to such techniques (41.4%). Declining a message was also often linked to a lack of authenticity cues (34.6%).

Reasons not solely related to the message itself accounted for a higher proportion of responses when declining messages than when accepting them, suggesting that factors such as routine behaviour, the particular context and level of knowledge (whether technical or risk-related) have a greater impact on behaviour when choosing to decline a message. This could be considered in relation to a general tendency to accept requests unless there is a reason not to, such as doubts regarding message authenticity, wider impacts on behaviour, or established habits and norms.

In order to examine the potential impact of message type on decision-making, the presence of themes was further broken down according to message type (i.e., genuine, mimicked or low authority message). This showed that the influence techniques used within messages contributed to a substantial proportion of reasons provided when accepting messages across all message types. This proportion was also found to increase as the authenticity cues present within the message diminished (Genuine Authority: 57.1% influence cue; Mimicked Authority: 60.7% influence cue; Low Authority: 70.4% influence cue). This suggests that the presence of influence techniques such as threat of loss and urgency can still provide a valid reason to respond to fraudulent messages, supporting the findings of work by Vishwanath et al. [41] that influence cues may be processed to the detriment of authenticity cues.

Overall, reference to authenticity cues when accepting a message reduced with mimicked (23.8%) compared to genuine authority messages (36.5%), suggesting that the relative differences in authenticity cues present in these message types are being noticed to an extent. However, despite a reduction in references to authenticity cues in these messages, they are still highlighted in a substantial proportion of reasons for accepting. Therefore, although a proportion of responders notice the inconsistencies and errors within these messages, these cues are not noticed by all. This may be accounted for by individual differences in trust and awareness (as shown in Section 4.2), as well as varying degrees of risk awareness and technical knowledge in relation to computer updates, whereby a higher weighting is given to influence cues such as professed authority than to potential authenticity cues when making decisions. The relative increase in the proportion of knowledge-related factors when declining mimicked and low authority messages (Genuine Authority: 4.3%; Mimicked Authority: 8%; Low Authority: 7%) also supports the role of cyber risk and technical awareness in identifying and avoiding suspicious messages, with such knowledge potentially impacting the identification and use of authenticity cues when deciding to decline a message.

Finally, themes related to specific context and routine factors appeared to be more influential when choosing to decline genuine authority messages (Context: 11.8%; Routine: 12.9%) compared to mimicked (Context: 4%; Routine: 8.8%) or low authority messages (Context: 8.5%; Routine: 8.8%). This may be due to genuine messages being declined less as a result of specific suspicions and more as a result of external factors, such as never accepting updates via pop-up.

5. Conclusions

The current study investigated participant response behaviour to computer update messages containing varying degrees of fraudulent cues when they interrupted a demanding serial recall task compared to when they were presented in a questionnaire phase with no competing demands. By requiring participants to make judgements in two different contexts where heuristic and systematic processing styles were likely to be differentially invoked, the relationship between message content, individual differences and processing strategy could be explored.

5.1. Primary theoretical implications

The findings of this study provide experimental evidence that scenarios likely to evoke heuristic processing (i.e., the serial recall phase of our experiment) increase susceptibility to fraudulent messages compared to more systematic processing conditions (i.e., the questionnaire phase), supporting recent models of susceptibility to phishing-based communications [41,42]. Most importantly, we found a differential impact of these processing conditions on (a) judgements of fraudulent cues present within messages and (b) the relationship between individual differences and response behaviour, providing a basis to extend current theoretical approaches to include a wider range of factors, such as the specific influence techniques used and particular individual traits, that may influence response behaviour at any given time.

Overall, the processing strategy that is used when an individual responds to a fraudulent communication appears to be a crucial mechanism in influencing susceptibility, providing a mediating factor that links the specific context that an individual is operating within to their ultimate response decision. It was hypothesised that the increased cognitive complexity of the serial recall phase would result in the use of heuristic processing strategies, which in turn would reduce the likelihood of identifying inconsistencies within messages that suggest a message may not be genuine. In line with this, we identified that participants did not differentiate between genuine authority messages and those that contained errors in authority cues when operating under cognitive pressure during the serial recall phase. This failure to identify so-called authenticity 'trigger' points within information reduced the likelihood that individuals would move from an initial truth default position to one of suspicion. However, this process was improved during the questionnaire phase when participants operated under conditions more likely to evoke systematic processing strategies [21,41,42].

It should be noted, however, that mimicked authority messages were still accepted significantly more than messages that did not contain any authority information during the questionnaire phase. This suggests that the presence of authority information is still able to override the presence of fraudulent cues, such as spelling errors, even under optimum processing conditions, whereas a lack of authority information was significantly less persuasive in such contexts. In this way, factors within the message content appear to differentially 'trigger' suspicion across both heuristic and systematic processing conditions. As a result, the use of 'traditional' cues in determining message legitimacy that have been identified in previous studies, such as poor spelling and grammar and the message source [8,17,41], appear to differentially apply according to both (a) other aspects of the message content (i.e., a perceived known message source overriding the presence of spelling errors), and (b) the information processing environment.

Our finding that a lack of authority information appears to be weighted more heavily in triggering suspicion than the presence of errors across both experimental phases suggests that authority cues have an independent persuasive impact. That this difference was found when all messages equally included urgency and loss influence techniques is particularly striking and suggests that the presence of urgency information does not override the absence of authority information when individuals make response decisions. As such, the findings of Vishwanath et al. [41], whereby the presence of urgency cues monopolised attentional resources also appears to differentially apply to other influence techniques (i.e., authority) and be impacted to a degree, although not entirely, by the availability of cognitive resource. Current models of susceptibility to fraudulent communications (e.g., [42]) should therefore consider that (a) the threshold of potential trigger points is likely to vary according to both the cognitive context of the individual and the particular influence techniques used within the message, and (b) certain influence techniques may prove more resistant than others to systematic processing strategies. Therefore, a consideration of the relative role of message factors should be incorporated into susceptibility models [41,42].

Finally, there was tentative support for the hypothesis that individual differences in trust impact on response behaviour [14], but this related primarily to scenario-specific conceptualisations (i.e., trust in computer communications and trust in the system). However, where such effects were identified, these were found to have a differential impact on response behaviour according to the cognitive context experienced. In particular, such individual differences were related to the likelihood of accepting fraudulent updates primarily when participants were operating under optimum conditions (i.e., where more systematic processing strategies were likely to be used). This suggests that in scenarios with a higher degree of cognitive pressure, individuals are equally vulnerable to being truth-biased [21], unless fraudulent messages are particularly easy to identify, in which case a higher degree of trust may reduce the likelihood of suspicion being triggered. Although individual differences in factors such as trust and awareness should therefore be included in susceptibility models, their relative impact across contexts and message types requires further investigation in relation to primary processing mechanisms. It may be that individual differences are to an extent negated by factors related to the cognitive context and message content, such that heuristic processing strategies increase trust in all recipients to an equal level, decreasing the ability to notice inconsistencies to a similar degree as those high in trust may experience when operating under optimum conditions.

5.2. Practical implications

The current study has identified that engaging in a cognitively demanding task reduces the ability to effectively identify fraudulent messages due to an increased reliance on heuristic forms of processing. This leads to a greater trust of fraudulent messages, which combines with the use of influence techniques designed to attract attention, to reduce the likelihood that inconsistencies or other fraudulent cues within the message will be noticed and trigger suspicion. These context-induced processing biases appear to override the impact of awareness and other individual differences to encourage people to 'accept'. Since the appearance of any fraudulent computer message is likely to interrupt a current primary task(s), this is likely to further increase pressure on employees to respond to such notifications as quickly as possible, and therefore heuristically, in order to avoid lost productivity [35]. As systems become increasingly integrated and complex, it is therefore essential that a holistic approach involving supportive software design, improved digital literacy, and training opportunities within pressured scenarios, be taken to address these susceptibilities by assisting in the 'triggering' of suspicion in the user.

Although awareness of fraud and phishing attacks may benefit individuals when responding to communications in optimum conditions, the findings of the current study suggest that such approaches are not sufficient to impact on behaviour when operating under cognitive pressure. As a result, decision support systems that reduce the impact of these heuristic processes via user-centred design would benefit from further development (e.g., [33]). Ideally, such development would be responsive to the increased cognitive pressure of a user and provide assistance in the direction of attention, the identification of inconsistencies and errors within message content, and the provision of in-the-moment technical knowledge, in order to enhance threat awareness whilst also providing support to address user uncertainty and potential gaps in technical understanding.

5.3. Limitations and future work

The current study represents an important exploration of individual and message-based factors that influence people to respond to fraudulent communications across different information processing conditions. Within the serial recall task, only a small number of messages were used in order to reduce the predictability of message interruptions whilst keeping the overall serial recall task length manageable for participants (e.g., [29]). Future work examining the impact of differing frequencies of messages (i.e., low vs. high rate of interruptions) would be beneficial in order to explore the potential role of habituation to such messages on subsequent judgements. Due to the small number of interruption messages that could be used within the serial recall task, this study also focused explicitly on only one influence technique (authority cues). However, this 'mimicking' style is an increasingly common technique within phishing scenarios and allowed us to explore whether the inclusion of errors, as commonly found in phishing messages, impacted on the persuasiveness of this influence technique across two processing scenarios (i.e., whether errors in authority messages were less likely to be noticed than the absence of authority information). As a result, other influence techniques commonly used in fraudulent messages were kept constant to ensure that these effects could be reliably examined, thus limiting the ability to manipulate additional content factors. Further work is therefore required to examine the relative contribution of other influence techniques, such as urgency cues and the use of loss versus reward-based techniques, by systematically manipulating the presence of this information in fraudulent messages, both independently and in combination with other techniques.

When exploring the qualitative themes regarding why people chose to respond to the computer update messages, data was collected during the questionnaire phase of the study and therefore it should be considered that the findings might only be applicable to this context. The potential reasons and themes given may therefore differ if participants are asked about their decision making during more cognitively complex scenarios when they are likely to be responding more heuristically (e.g., the serial recall phase), for example, through a lower reference to authority cues and a higher reference to influence cues. However, due to the nature of the serial recall task, this data could not be collected following participant responses to each message type in the serial recall phase and would also likely impact on participant response choice during the questionnaire phase of the study. Furthermore, the questionnaire phase of the study was designed to maximise the likelihood of systematic processing strategies being used, with the inclusion of questions related to decision strategies likely to have increased corresponding suspicion and thus the likelihood that authenticity information would impact on response behaviour. In order to examine potential differences in the distribution of themes whilst avoiding creating the above effect in heuristic processing conditions, future work would need to use different qualitative data collection mechanisms, such as think aloud protocols.

Finally, within this study respondents were not using their home computers and were instead using university PCs under controlled laboratory testing conditions. This may limit the generalisability of the findings to home computer users and instead focuses more explicitly on computer users within organisational settings, such as a work or university context. Since the sample of this study was limited to predominantly female university students, the findings are particularly relevant to contexts where young adults are the predominant computer users, such as universities, colleges or organisations employing individuals from this group. However, more research is required to determine the extent that the approach of this group of users (in terms of both age and gender) may differ from other demographic groups when dealing with computer security in cognitively complex conditions (e.g., [16]).

Although the design of the genuine updates in this study were based on actual updates experienced by computer users (e.g., the authors and their research team members), it is also acknowledged that they were still developed by the researchers for the purposes of the experiment. During the questionnaire phase, participants were asked how they would ordinarily respond to such messages, which may differ from decisions made during laboratory tasks that are considered to be less personally relevant, and therefore effortful systematic processing strategies may not be considered necessary. Further targeted work using a range of decision scenarios will enable such potential differences to be further explored and disentangled.

Funding

This work was partly funded by the Centre for Research and Evidence on Security Threats (ESRC Award: ES/N009614/1).

References

[1] B. Atkins, W. Huang, A study of social engineering in online frauds, Open J. Soc. Sci. 1 (3) (2013) 23, http://dx.doi.org/10.4236/jss.2013.13004.

[2] A. Baddeley, Working memory, Current Biology 20 (4) (2010) R136–R140, http://dx.doi.org/10.1016/j.cub.2009.12.014.

[3] C.F. Bond, B.M. DePaulo, Accuracy of deception judgments, Personality and Social Psychology Review 10 (2006) 214–234, http://dx.doi.org/10.1207/s15327957pspr1003_2.

[4] E. Castle, N.I. Eisenberger, T.E. Seeman, W.G. Moons, I.A. Boggero, M.S. Grinblatt, S.E. Taylor, Neural and behavioral bases of age differences in perceptions of trust, PNAS 109 (51) (2012) 20848–20852, http://dx.doi.org/10.1073/pnas.1218518109.

[5] R. Cialdini, Influence: The Psychology of Persuasion, HarperCollins, New York, 2007.

[6] P.J. DePaulo, B.M. DePaulo, Can attempted deception by salespersons and customers be detected through nonverbal behavioural cues? Journal of Applied Social Psychology 19 (1989) 1552–1577, http://dx.doi.org/10.1111/j.1559-1816.1989.tb01463.x.

[7] R. Dhamija, J.D. Tygar, M. Hearst, Why phishing works, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2006, pp. 581–590 (New York) 10.1145/1124772.1124861.

[8] J. Downs, M. Holbrook, L. Cranor, Decision Strategies and Susceptibility to Phishing, Symposium on Usable Privacy and Security 2006, pp. 79–90 (Pittsburgh, PA) 10.1145/1143120.1143131.

[9] A.H. Eagly, S. Chaiken, The Psychology of Attitudes, Harcourt Brace Jovanovich College Publishers, San Diego, CA, 1993.

[10] F. Faul, E. Erdfelder, A.-G. Lang, A. Buchner, G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences, Behavior Research Methods 39 (2007) 175–191, http://dx.doi.org/10.3758/BF03193146.

[11] B.J. Fogg, J. Marshall, A. Ospiovich, C. Varma, O. Laraki, N. Fang, J. Paul, A. Rangnekar, J. Shon, P. Swani, M. Treinen, Elements that affect web credibility: early results from a self-report study, Proceedings of CHI '00 Extended Abstracts on Human Factors in Computing Systems 2000, pp. 287–288, http://dx.doi.org/10.1145/633292.633460.

[12] M.L. Fransen, B.M. Fennis, Comparing the impact of explicit and implicit resistance induction strategies on message persuasiveness, The Journal of Communication 64 (5) (2014) 915–934, http://dx.doi.org/10.1111/jcom.12118.

[13] H.M. Hodgetts, D.M. Jones, Interruption of the tower of London task: support for a goal-activation approach, Journal of Experimental Psychology. General 135 (1) (2006) 103–115, http://dx.doi.org/10.1037/0096-3445.135.1.103.

[14] K.W. Hong, C.M. Kelley, C.B. Mayhorn, E. Murphy-Hill, Keeping Up With the Joneses, 2013.

[15] R.H. Hoyle, M.T. Stephenson, P. Palmgreen, E.P. Lorch, R.L. Donohew, Reliability and validity of a brief measure of sensation seeking, Personality and Individual Differences 32 (2002) 401–414, http://dx.doi.org/10.1111/j.1360-0443.2007.01958.x.

[16] T. Jagatic, N. JohnSon, M. Jakobsson, F. Menczer, Social phishing, Communications of the ACM 50 (10) (2007) 94–100, http://dx.doi.org/10.1145/1290958.1290968.

[17] M. Jakobsson, A. Tsow, A. Shah, E. Blevis, Y.-K. Lim, What instills trust? A qualitative study of phishing, Financial Cryptography & Data Security: Lecture Notes in Computer Science, 4886, 2007, pp. 356–361, http://dx.doi.org/10.1007/978-3-540-77366-5_32.

[18] J. Johnson, Measuring thirty facets of the five factor model with a 120-item public domain inventory: development of the IPIP-NEO-120, Journal of Research in Personality 51 (2014) 78–89, http://dx.doi.org/10.1016/j.jrp.2014.05.003.

[19] D. Kahneman, Thinking, Fast and Slow, Penguin, London, UK, 2011.

[20] D. Kahneman, A. Tversky, Choices, values, and frames, The American Psychologist 39 (4) (1984) 341–350, http://dx.doi.org/10.1037/0003-066X.39.4.341.

[21] T.R. Levine, Truth-default theory: a theory of human deception and deception detection, Journal of Language and Social Psychology 33 (2014) 378–392, http://dx.doi.org/10.1177/0261927X14535916.

[22] R.C. Mayer, J.H. Davis, F.D. Schoorman, An integrative model of organisational trust, The Academy of Management Review 20 (3) (1995) 709–734, http://dx.doi.org/10.5465/AMR.1995.9508080335.

[23] D.H. McKnight, V. Choudhury, C. Kacmar, Developing and validating trust measures for e-commerce: an integrative typology, Information Systems Research 13 (3) (2002) 334–359, http://dx.doi.org/10.1287/isre.13.3.334.81.

[24] S. Mishra, Decision-making under risk: integrating perspectives from biology, economics, and psychology, Pers. Soc. Psychol. 18 (3) (2014) 280–307, http://dx.doi.org/10.1177/1088868314530517.

[25] D. Modic, R.J. Anderson, Reading this may harm your computer: the psychology of malware warnings, Computers in Human Behavior 41 (2014) 71–79, http://dx.doi.org/10.1016/j.chb.2014.09.014.

[26] C.M. Monk, J.G. Trafton, D.A. Boehm-Davis, The effect of interruption duration and demand on resuming suspended goals, Journal of Experimental Psychology. Applied 14 (4) (2008) 299–313, http://dx.doi.org/10.1037/a0014402.

[27] P.L. Morgan, J. Patrick, Designing interfaces that encourage a more effortful cognitive strategy, Proceedings of the 54th Annual Meeting of the Human Factors and Ergonomics Society 2010, pp. 408–412 (San Francisco, California, USA, 27 October - 1 September 2010) 10.1177/154193121005400429.

[28] P.L. Morgan, J. Patrick, Paying the price works: increasing goal access cost improves problem solving and mitigates the effect of interruption, The Quarterly Journal of Experimental Psychology 66 (1) (2013) 160–178, http://dx.doi.org/10.1080/17470218.2012.702117.

[29] P.L. Morgan, J. Patrick, S.M. Waldron, S.L. King, T. Patrick, Improving memory after interruption: exploiting soft constraints and manipulating information access cost, Journal of Experimental Psychology. Applied 15 (4) (2009) 291–306, http://dx.doi.org/10.1037/a0018008.

[30] J.H. Patton, M.S. Stanford, E.S. Barratt, Factor structure of Barratt impulsiveness scale, Journal of Clinical Psychology 51 (6) (1995) 768–774 (PMID: 8778124).

[31] J.W. Peirce, Generating stimuli for neuroscience using PsychoPy, Frontiers in Neuroinformatics 2 (2009) 10, http://dx.doi.org/10.3389/neuro.11.010.2008.

[32] J.A. Roberts, C. Manolis, Cooking up a recipe for self-control: the three ingredients of self-control and its impact on impulse buying, Journal of Marketing Theory and Practice 20 (2) (2012) 173–188, http://dx.doi.org/10.2753/MTP1069-6679200204.

[33] J. Sänger, N. Hänsch, B. Glass, Z. Benenson, R. Landwirth, M.A. Sasse, Look before you leap: improving the users' ability to detect fraud in electronic marketplaces, Proceedings of CHI '16 Human Factors in Computing Systems 2016, pp. 3870–3882, http://dx.doi.org/10.1145/2858036.2858555.

[34] E. Sillence, P. Briggs, P. Harris, L. Fishwick, A framework for understanding trust factors in web based health advice, International Journal of Human Computer Studies 64 (2006) 697–713, http://dx.doi.org/10.1016/j.ijhcs.2006.02.007.

[35] J.B. Spira, J.B. Feintuch, The cost of not paying attention, How Interruptions Impact Knowledge Worker Productivity, 2005 (http://iorgforum.org/wp-content/uploads/2011/06/CostOfNotPayingAttention.BasexReport1.pdf Accessed 04.05.2016).

[36] F. Stajano, P. Wilson, Understanding scam victims: seven principles for systems security, Communications of the ACM 54 (3) (2011) 70–75, http://dx.doi.org/10.1145/1897852.1897872.

[37] J.P. Tangney, R.F. Baumeister, A.L. Boone, High self-control predicts good adjustment, less pathology, better grades, and interpersonal success, Journal of Personality 72 (2) (2004) 271–324, http://dx.doi.org/10.1111/j.0022-3506.2004.00263.x.

[38] J.G. Trafton, C.M. Monk, Task interruptions, in: D.A. Boehm-Davis (Ed.), Reviews of Human Factors and Ergonomics, 3, Human Factors & Ergonomics Society, Santa Monica, CA 2008, pp. 111–126.

[39] A. Tversky, D. Kahneman, Judgment under uncertainty: heuristics and biases, Science 185 (1974) 1124–1131, http://dx.doi.org/10.1126/science.185.4157.1124.

[40] A. Vishwanath, Habitual Facebook use and its impact on getting deceived on social media, Journal of Computer-Mediated Communication 20 (2015) 83–98, http://dx.doi.org/10.1111/jcc4.12100.

[41] A. Vishwanath, T. Herath, R. Chen, J. Wang, H.R. Rao, Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model, Decis. Support. Syst. 51 (2011) 576–586, http://dx.doi.org/10.1016/j.dss.2011.03.002.

[42] A. Vishwanath, B. Harrison, Y.J. Ng, Suspicion, cognition, and automaticity model of phishing susceptibility, Communication Research (2016) 1–21 (online pre-print) 10.1177/0093650215627483.

[43] Y.D. Wang, H.H. Emurian, An overview of online trust: concepts, elements, and implications, Computers in Human Behavior 21 (2005) 105–125, http://dx.doi.org/10.1016/j.chb.2003.11.008.

[44] M. Workman, Wisecrackers: a theory-grounded investigation of phishing and pretext social engineering threats to information security, Journal of the American Society for Information Science and Technology 59 (4) (2008) 662–674, http://dx.doi.org/10.1002/asi.20779.

[45] J. Cohen, Statistical power analysis for the behavioural sciences, Lawrence Earlbaum Associates, Hillside. NJ, 1988.

[46] S. Sundar, The MAIN model: a heuristic approach to understanding technology effects on credibility, in: M. Metzger, A. Flanagin (Eds.), Digital Media, Youth, and Credibility, MIT Press, Cambridge, MA 2008, pp. 73–100.

Dr Emma Williams is a Research Associate within the School of Management at the University of Bath. She specialises in research exploring susceptibility to malicious forms of influence, with a primary focus on the interaction between individuals, influence techniques and the wider environment.

Dr Phillip L. Morgan is a Senior Lecturer in Cognitive Psychology and Human Factors at the University of the West of England - Bristol. He specialises in research involving interruptions and distractions within every day and workplace settings, as well as human-computer interaction and display design including research involving autonomous vehicles.

Professor Adam Joinson is Professor of Information Systems at the University of Bath. He conducts research on the intersection between technology and behaviour – including work on communication patterns, influence, security and privacy.