ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/123913/

This is the author’s version of a work that was submitted to / accepted for publication.

Citation for final published version:

Morgan, Phillip L. , Soteriou, Robinson, Williams, Craig and Zhang, Qiyuan 2019. Attempting to reduce susceptibility to fraudulent computer pop-ups using malevolence cue identification training. Presented at: AHFE 2019: International Conference on Applied Human Factors and Ergonomics, Washington, DC, USA, 24-28 July 2019. Published in: Ahram, Tareq and Karwowski, Waldemar eds. Advances in Human Factors in Cybersecurity: Proceedings of the AHFE 2019 International Conference on Human Factors in Cybersecurity, July 24-28, 2019, Washington D.C., USA. Advances in Intelligent Systems and Computing. , vol.960 Springer Verlag, pp. 3-15. 10.1007/978-3-030-20488-4_1

Publishers page: http://dx.doi.org/10.1007/978-3-030-20488-4_1

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher’s version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

Attempting to Reduce Susceptibility to Fraudulent Computer Pop-Ups using Malevolence Cue Identification Training

Phillip L. Morgan*, Robinson Soteriou, Craig Williams, & Qiyuan Zhang

Cardiff University, School of Psychology, Cardiff, CF10 3AT, UK
morganphil@cardiff.ac.uk (corresponding author)

Abstract. People accept a high number of computer pop-ups containing cues that indicate malevolence when they occur as interrupting tasks during a cognitively demanding memory-based task [1, 2], with younger adults spending only 5.5-6-seconds before making an accept or decline decision [2]. These findings may be explained by at least three factors: pressure to return to the suspended task to minimize forgetting; adopting non-cognitively demanding inspection strategies; and, having low levels of suspicion [3]. Consequences of such behavior could be potentially catastrophic for individuals and organizations (e.g., in the event of a successful cyber breach), and thus it is crucial to develop effective interventions to reduce susceptibility. The current experiment (N = 50) tested the effectiveness of *malevolence cue identification training* (MCIT) interventions. During phase 1, participants performed a serial recall task with some trials interrupted by pop-up messages with accept or cancel options that either contained cues (e.g., missing company name, misspelt word) to malevolence (*malevolent condition*) or no cues (*non-malevolent condition*). In phase 2, participants were allocated to one of three groups: no MCIT / Control, non-incentivized MCIT / N-IMCIT, or incentivized MCIT / IMCIT. Control group participants only had to identify category-related words (e.g., colors). Participants in intervention conditions were explicitly made aware of the malevolence cues in Phase 1 pop-ups before performing trying to identify malevolence cues within adapted passages of text. The N-IMCIT group were told that their detection accuracy was being ranked against other participants, to induce social comparison. Phase 3 was similar to phase 1, although 50% of malevolent pop-ups contained new cues. MCIT did lead to a significant reduction in the number of malevolent pop-ups accepted under some conditions. Incentivized training did not (statistically) improve performance compared to non-incentivized training. Cue novelty had no effect. Ways of further improving the MCIT training protocol used, as well as theoretical implications, are discussed.

Keywords: Cyber-Security · Susceptibility · Task Interruption · Intervention Training

1 Introduction

The prevalence of malevolent online communications (MOCs), such as phishing attempts, is growing at a rapid pace. Recent statistics indicate that 264,483 phishing reports were made in the third quarter of 2018, which is markedly higher than the same quarter in 2016 [4]. The UK Government commissioned a report with the research revealing a staggering 46% of UK businesses reporting a breach of cyber-security, including phishing attempts, in the 12-months prior to being surveyed [5]. Such MOCs are targeted at individuals and organizations. Examples include fake pop-ups claiming to be from well-known companies that if clicked/accepted result in malware infection and/or payment demands [6]. Recent large scale disruptive attacks include Sony Pictures, 2015, where employees clicked fake links resulting in login details & passwords being stolen, allowing fraudsters to hack-in [7]. The significance of this problem, together with other cyber threats, has been reflected by worldwide investments in cybersecurity with the likes of the UK Government and Bank of America committing funds to improve cybersecurity prevention and protection [8, 9]. Whilst much of this investment is being dedicated to improvements in the protection of networks, systems and software, computer users are seen as the main weakness in effective prevention of successful cyber-attack breaches [10], due to multiple fallibilities related to e.g., perception, attention, memory, decision making, and risk. Many cyber hackers are aware of these and will exploit them when developing MOCs. The current paper examines (1) susceptibility to MOCs delivered when humans are under short-term memory pressure and (2) the efficacy of an intervention training method to reduce susceptibility.

Pop-up messages occur regularly on networked computer devices, often unexpectedly and during engagement in another task(s) [11, 12]. Many contain information on and/or links to updates that are essential to maintain efficient performance of the computer system and/or software [13]. However, there are growing numbers of fake computer updates that mimic trusted companies and brands, with hackers' intent on encouraging people to clink on links which can result in cyber breaches.

Pop-ups at times will act as a distractor (e.g., if the user is able to ignore or deal with it without disengaging from an ongoing task) but are more likely to initiate an interruption (e.g., if the user is not able to ignore it and has to disengage from the ongoing task). Even short interruptions (as short as 2.8-seconds) can shift the focus of attention and memory and lead to increased errors within a suspended task [14] with factors such as interruption duration and demand exacerbating the extent of disruption [15, 16], as predicted by a leading model [17, 18]. However, few have considered how individuals choose to engage with interrupting tasks when there is no time constraint on their completion, i.e., when their response (which could be a few seconds) to the interrupting task determines when they will resume the suspended task (see [1, 2]).

In considering pop-up messages as task interruptions, how individuals allocate resources to verify authenticity will likely depend on factors outside the usual task parameters often studied, such as time costs and possible performance impairments. According to the Suspicion, Cognition, Automaticity Model / SCAM [3], whether malevolent cues are noticed within fraudulent communications depends on the depth of processing an individual engages in. The less suspicious and more trusting an individual is, the more likely they are to process the content of pop-up messages using automatic heuristic processing compared to someone who is more suspicious and less trusting who will likely engage in more cognitively effortful and time consuming processing. Similarly, those who have a higher need for cognitive stimulation [19], will be more susceptible to influence techniques used within pop-up messages such as urgency, compliance with authority and avoidance of loss; at the expense of looking for suspicious aspects, such as message authenticity cues (e.g., correct spelling and grammar, company name).

This leads to a prediction that an intervention training protocol that increases suspicion and encourages more effortful processing of pop-up message content should have carryover effects to a subsequent task performed with malevolent pop-up interruptions.

To our knowledge, only two published studies have considered human susceptibility to fraudulent pop-up interruptions occurring during a demanding memory-based task. [2] developed a paradigm where young adult participants were interrupted by one of three different types of pop-up message during a serial recall memory recall task. One third of pop-ups were designed to look genuine (*genuine* condition) and high in authority with no cues to potential malevolence. Another third (*mimicked* condition) were also high in authority but contained cues to suggest malevolence. The other third were also of a malevolent nature and *low authority* (i.e., contained no authority details relating to the source of the pop-up such as company name, logo, or website link). Participants had to decide whether to accept or decline pop-ups, at which point the primary task would be reinstated at the point of interruption. Predictions informed by parameters of SCAM [3] were supported, with an alarming 63% of mimicked pop-ups accepted compared with 66% in the genuine condition. Even more worrying was that 56% of low authority pop-ups were accepted. Participants spent on average only ~5.5-6-seconds viewing pop-up message content before committing to a response. When there were no time constraints to resume an interrupted task, participants accepted a slightly higher percentage (72%) of genuine pop-ups and slightly fewer (55%) mimicked pop-ups. This suggests that even without other cognitive and time pressures, people are still not very good at detecting malevolent cues within mimicked pop-up interruptions. [1] reported similar findings with older adults. Participants demonstrated higher levels of susceptibility to malevolent pop-ups during an interrupted memory recall phase, despite spending significant more time (~10.5-11-s) viewing them than in [1]. Fitting with SCAM-based low suspicion and automaticity predictions [3], both studies demonstrate very high levels of human susceptibility to malevolent pop-up interruptions that occur during a demanding memory-based task. However, concerns remain as neither study showed marked malevolent detection improvements when time pressure was not a factor.

Given these results, it is important further develop and test interventions to reduce susceptibility to computer-based communications such as malevolent pop-up messages. Education-based training interventions are not always effective [20] with some finding that people are more suspicious of scams that they are familiar with versus those that are less familiar [21]. [22] tested the effectiveness of emails containing cues to malevolence although found that not all people read and processed the content to a deep enough level to identify them effectively. These findings fit SCAM parameters regarding the use of automatic heuristic processing strategies, especially when suspicion is low. It could be that training effectiveness is dependent on the extent of encouragement to engage in and cognitively process training materials. Another factor that could potentially increase engagement in training is competition, which has been shown to facilitate motivation and performance in some instances [23]. Short term improvements were found when testing a competitive e-learning interface that displayed a rank order for the best performing students [24]. Competitive ranking may encourage individuals to engage more in the task to gain an accurate appraisal of their own performance compared to others, and thus improve upon such performances. The social process of evaluating one's accuracy in relation to their ability may encourage a desire to improve performance to increase own sense of self-worth [25]. Thinking more about the content of the training, as well as gaining satisfaction from it, may increase the likelihood of the information being remembered and utilized subsequently.

The current experiment has four main aims. One is to attempt to replicate the findings of [1] and [2] on susceptibility to pop-ups with cues to malevolence when they occur as interruptions to a memory-based task. Second, to examine whether, and if so to what extent, susceptibility can be alleviated through an intervention involving malevolent cue identification training (abbreviated to MCIT hereafter). The training was designed not to only increase suspicion and awareness of cues to malevolence but also to encourage more effortful cognitive processing of message content. Third, we examined whether a form of incentivized MCIT that encourages competitiveness through social comparison might further increase the intervention effectiveness. A fourth aim was to establish whether beneficial effects of MCIT transfer to conditions involving novel cues to malevolence that have not been experienced as part of the training intervention.

2 Method

2.1 Participants

Fifty Cardiff University Psychology undergraduate students (age: 19.32; *SD* 1.06) were recruited, via opportunity sampling, in return for course credits with adequate a priori power (.8 detect medium to large effect sizes (Cohen's *f* .25 -.4). Participants were first-language English or highly proficient in English as a second language, and had normal/correct vision. They were assigned to one of three cue identification training groups. There were 16 in the Non-Malevolent Cue Identification (N-MCIT)/Control group (*M* age: 19.63-years, four male), 17 in the Non-Incentivized Malevolent Cue Identification (N-IMCIT) group (*M* age: 19.06-years, six male), and 17 in the Incentivized Malevolent Cue Identification (IMCIT) group (*M* age: 19.29-years, two male).

2.2 Design

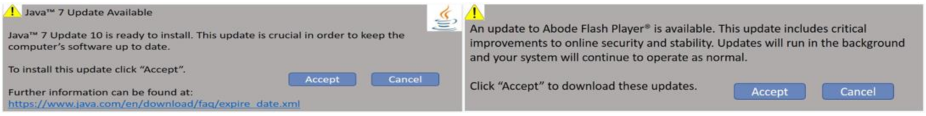
A mixed factorial design was employed. The between-participants' independent variable (IV) was CIT Group with three levels: Control, N-IMCIT, and IMCIT. There were three repeated measures IVs. One was serial recall phase with two levels: Phase 1/Pre-Intervention 1, and, Phase 3/Post-Intervention. Another was the malevolency (Message Type) of the pop-up with two levels: Non-Malevolent/Genuine, and, Non-Genuine/Malevolent. The third (Phase 3 only) was whether malevolent pop-ups contained the same (Malevolent-Old) or different malevolence cues than in Phase 1. There were two main dependent variables (DVs). The first was decision response to the pop-up request where two responses were possible: Accept, or, Decline. The second was the time to make a response. During the intervention phase, participant performance was recorded in two stages. The first stage required participants to respond to whether they identified at least one cue to indicate a category exemplar (Control group) or cue to malevolence (other groups), by choosing Yes or No. If choosing Yes, participants then had to record the number of cues identified (maximum 3 per passage of text with five passages in total).

2.3 Materials

Phase 1 and 3 Serial Recall and Interruption Pop-Up Tasks

Tasks were programmed on run on *Intel® Core™* i5 PCs connected to 1920x1080 *iiyama* 24" flat-panel monitors. The serial recall task was created using PsychoPy2 software [26]. There were 18 trials in Phase 1 and 30 in Phase 3. During each trial, a different string of nine letters and numbers, e.g., 96KJ3785H were presented in the center of the screen for 9-seconds before disappearing. An instruction ('enter code') appeared after a 2-second retention interval to inform participants that they should try and recall and write down letters and numbers in the order in which they were presented.

Twelve trials were interrupted in Phase 1: six with non-malevolent and six with malevolent pop-ups. Six trials were not interrupted. Twenty-four trials were interrupted in Phase 3: twelve with non-malevolent and twelve with malevolent pop-ups, with six of these containing the same (Old) malevolency cues as in Phase 1 and six containing New cues. Pop-up messages appeared in the center of the screen after the letter/number string had disappeared and before the recall instruction appeared, and remained on the screen until either an accept ('A' key) or cancel ('C') response was registered. Immediately after this response, the serial recall task was reinstated from the point in which it had been suspended (i.e., 'enter code' would appear next). Each new trial was initiated after the spacebar was pressed. Each pop-up contained text describing the scenario, plus an extra line of text with an instruction (e.g., 'Click 'accept' to download the [XXXX: name] pop -p') with boxes for Accept and Cancel. All non-malevolent and some malevolent pop-ups also contained a company logo in the top right corner and a hyperlink to a related website underneath text that read 'Further information can be found here:'.



Java™ 7 Update Available

Java™ 7 Update 10 is ready to install. This update is crucial in order to keep the computer's software up to date.

To install this update click "Accept".

Further information can be found at:
https://www.java.com/en/download/faq/expire_date.xml

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

Click "Accept" to download these updates.

Fig 1. Examples of a non-malevolent pop-up (left) and malevolent pop-up (right)

Non-malevolent pop-ups contained cues (or indeed not lack of) to suggest that they were genuine (Figure 1, left). These included a company logo, name (corresponding to logo), and website link, and accurate grammar and accurate spelling. Malevolent pop-ups (Figure 1, right) contained three of six cues to malevolence: lack of company logo, name, website link, and an instance of inaccurate grammar or a misspelt word(s). During Phase 3, malevolent pop-ups contained either three Old or three New cues. New cues included: misspelling within website link, non-capitalization of company names, missing a key detail, having a fake logo, or capitalization of a word that should not be.

Prior to the start of Phase 1 and 3 trials, the following message was displayed in the middle of the computer screen for 15-seconds:

> *'This system is protected by virus protection software and pop-ups 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 of data are legitimate.'*

Phase 2 Intervention and Control Non-Intervention Tasks

Participants in the intervention conditions were given explicit information on the errors/cues to malevolence contained in Phase 1 malevolent pop-ups. They were also given a small whiteboard and marker pen to make notes on these if they wished to do so. All participants were required to read a set of five passages of text, with 5-minutes (~60-seconds per passage) from fictitious companies. The passages each contained textual information relating to five nominal categories (drinks, transport, sport, clothes, color). Passages were adapted for the N-IMCIT and IMCIT conditions to contain the same errors/cues to malevolence as in Phase 1. Participants in the Control group were required to first indicate whether category (e.g., color) words were present within the passage by clicking ‘yes’ or ‘no’ within an online answer sheet, and if choosing Yes, they then had to type the number of instances they could find (max = three per passage) before moving to the next passage. Participants in the intervention groups had to do this for cues indicating malevolence (max = 3 per passage) rather than category instances. Answer sheets were set out as a separate tab containing a table to be completed in relation to each passage. For the IMCIT group, each tab was followed by a leader board with performance *appearing* to be ranked against all previous participants with their position increasing after completion of each passage. Leaderboard positions were preset with the intention of encouraging (through social comparison) participants to try harder and apply more cognitive effort for each new passage.

2.4 Procedure

Before providing consent, participants read through an information sheet and experimental instructions (which were also verbally read by the experimenter) before completing two practice trials: one with a non-interrupted serial recall task, and another with a serial recall task interrupted by a non-malevolent pop-up. They were not informed about the cyber security element of the experiment during this process. At the beginning of Phase 1, participants were presented with the computer security message (see Materials). After this disappeared, they were able to press the spacebar to start trial one of 18, with 12 of the trials interrupted (see Materials). Phase 2 was the intervention phase. Participants read an instruction sheet appropriate for their group. All were instructed they had 5-minutes to read 5-passages (one-at-a-time) and complete the cue identification task relevant to their group. The Control group had to indicate (Yes or No) whether the passage of text contained at least one cue relating to its category description (e.g., color: look for color words). If answering yes, they then had to indicate how many category words they could identify within the passage (i.e., 1-3). N-IMCIT and IMCIT groups were first given written information pertaining to the malevolency cues contained within pop-ups experienced in Phase 1. These were explained verbally by the experimenter who checked participants’ understanding. As with the Control group, participants in the MCIT groups were then presented with 5-passages of text, one-at-a-time, and had to indicate (Yes or No) whether the passage it contained at least one trained cue indicating potential malevolence. Participants were also provided with a small whiteboard and marker to make notes, if desired. Phase 3 (post-intervention) involved 30 serial recall trials with 24 interrupted. After Phase 3, participants completed demographics and pop-up awareness questionnaires. Participants were debriefed, with information about cyber-security and awareness aims.

3 Results and Discussion

All analyses are two-tailed with α = .05. One dataset was excluded, as it was found to be a statistical outlier (z-scores > 3.29, *ps* < .001) on more than one measure.

Percentage of Pop-Up Messages Accepted/Declined

First, we consider mean percentages of '*malevolent'* pop-ups accepted across Phases 1 (pre-intervention) and 3 (post-intervention), collapsing across New and Old cue malevolent pop-ups in Phase 3 (Table 1). The percentage of malevolent pop-ups accepted looks to have decreased in Phase 3 for both MCIT groups, although increased for the Control group. Somewhat surprisingly, the mean percentage is markedly lower in the Control versus the N-IMCIT and IMCIT groups.

A mixed 3 x 2 analysis of variance (ANOVA) with Training Group as the between-subjects variable (Control, N-IMCIT, IMCIT) and Phase (pre-intervention, post-intervention) revealed non-significant main effects of Training Group, *F*(2, 47) = 1.07, *MSE* = .08, *p* = .35, and, Phase, *F*(1, 47) = 1.03, *MSE* = .04, *p* = .32. There was however a significant interaction, *F*(2, 47) = 3.44, *MSE* = .04, *p* = .04. Bonferroni pot-hoc tests revealed a non-significant (although trend) reduction in the percentage of malevolent pop-ups accepted in Phase 3 compared with Phase 1 for the IMCIT group (*p* = .07). However, the significant interaction might be better explained by the percentage of malevolent pop-ups accepted by the Control group in Phase 1 being significantly lower than in the N-IMCIT and IMCIT groups within Phase 1 (*ps* < .025). Given this unexpected difference (discussed later), another mixed ANOVA, this time 2 (Training Group: MCIT, IMCIT) x 2 (Phase: 1, 3), was conducted. This revealed a significant main effect of Phase, *F*(1, 32) = 5.63, *MSE* = .04, *p* = .02 with a lower percentage of malevolent pop-ups accepted in Phase 3 than in Phase 1. There was a non-significant main effect of Training Group, *F*(1, 32) = .96, *MSE* = .08, *p* = .33, and a non-significant interaction, *F*(1, 32) = .03, *MSE* = .04, *p* = .86.

Taken together, these findings suggest that: (1) MCIT worked in terms of reducing the percentage of malevolent pop-up messages accepted post-intervention, (2) IMCIT did not lead to better performance than N-IMCIT, and, (3) participants in the Control group, in Phase 1 at least, performed differently (i.e., chose to accept far less malevolent pop-ups) to those in MCIT conditions. In relation to (1), findings are in line with SCAM predictions that heightening suspicion will lead to increased cognitive and less automatic processing of stimuli [3], thus improving the likelihood of identifying malevolence cues. However, the percentage of malevolent pop-ups accepted was still very high, even after the intervention. In relation to (2), incentivized MCIT through social comparison (using an onscreen leaderboard technique), was not effective enough to cause even more suspicion and increased cognitive processing of potential cues to suggest malevolence within pop-up messages compared to non-incentivized MCIT. This finding (despite there being a trend) is not in line with [22]and possible reasons are considered in the Limitations section. Considering (3), the only difference was when the groups were tested: The Control group were tested after the MCIT groups.

Table 1. Percentage of Malevolent and Genuine pop-ups accepted during Phases 1 and 2 and across each Training Group. *Note.* SD = Standard Deviation.

		Malevolent Pop-Ups	Genuine Pop-Ups

Phase	Condition	*Mean*	*SD*	*Mean*	*SD*
1	Control	56.30	.34	63.54	.39
	N-IMCIT	73.41	.31	84.31	.30
	IMCIT	81.29	.29	87.25	.29
3	Control	67.69	.33	70.83	.35
	N-IMCIT	60.35	.31	85.29	.24
	IMCIT	70.12	.31	92.65	.11

Next, we consider mean percentages of *'genuine'* pop-ups accepted in Phases 1 and 3, noting again that both New and Old cue malevolent pop-up data are collapsed across (Table 1). The percentage of genuine pop-ups accepted increased marginally in Phase 3 across all groups. However, and as with malevolent pop-ups, the mean percentage of genuine pop-ups accepted in Phase 1 was markedly lower in the Control versus MCIT groups. A mixed 3 x 2 analysis of variance (ANOVA) with Training Group as the between-subjects variable and Phase revealed a marginally non-significant main effect of Training Group, $F(2, 47) = 3.12$, $MSE = .07$, $p = .054$, and a non-significant main effect of Phase, $F(1, 47) = 2.57$, $MSE = .02$, $p = .12$. There was a non-significant interaction. However, these findings might again be affected by the unusual pattern of data in the Control condition during Phase 1 compared to the MCIT condition. Therefore, a 2 (Training Group: MCIT, IMCIT) x 2 (Phase: 1, 3) mixed ANOVA was conducted. There were non-significant main effects of Training Group, $F(1, 32) < 1$, $p = .50$, and Phase, $F(1, 32) < 1$, $p = .39$, and a non-significant interaction, $F(1, 32) < 1$, $p = .55$.

Taken together, these findings suggest that (1) the ability to identify genuine pop-up messages was high, (2) MCIT did not have any effect on this, and (3) participants in the Control group, in Phase 1 at least, performed quite differently (i.e., accepted fewer genuine pop-ups) to those in the MCIT conditions. It is difficult to determine why participants in MCIT groups seemed to be very good at classifying most genuine pop-ups as genuine and then chose to accept rather than decline. It might have been relatively easier to check whether pop-ups contained no cues to malevolence than to check and register a cue(s) to malevolence. Although, and given the very high (and somewhat worrying) percentages of malevolent pop-ups accepted, it could be that participants, particularly in Phase 1, were more inclined to adopt a trusting stance [3] and accept most pop-ups as being genuine unless they noted at least one cue that was enough to raise suspension and cause them to respond in a different way (i.e., decline the pop-up. In order to speak to these possibilities, we will later examine the amount of time participants took before making a decision to accept/decline messages.

Next, we examine for possible differences between the percentage of Old (i.e., contained same cue types as in Phase 1, trained on these cues in MCIT conditions in Phase 2) versus New (i.e., contained different cue types as in Phase 1, not trained on these cues in MCIT conditions in Phase 2) malevolent pop-ups in Phase 3 only (Table 2). Whilst there is no difference within the Control Group, participants in the MCIT groups appear to have accepted marginally more New than Old malevolent messages, particularly in the IMCIT condition. However, a 3 (Training Group) x 2 (Cue Familiarity: Old, New) mixed ANOVA revealed non-significant main effects of Training Group, $F(2, 47) < 1$, $p = .64$, Cue Familiarity, $F(1, 47) < 1$, $p = .92$, and a non-significant interaction, $F(2, 47) < 1$, $p = .33$. Given the unusual accept/decline behavior of the Control Group in Phase 1 (see above), an additional analysis (2 x 2 mixed ANOVA) was conducted with the Control group excluded. There were still non-significant main effects of Training Group, *F*(1, 32) < 1, *p* = .36, Cue Familiarity, *F*(1, 32) = 1.61, *p* = .21, and a non-significant interaction, *F*(1, 32) < 1, *p* = .62.

Table 2. Percentage of Old and New pop-ups accepted during Phase 3 across each Training Group. *Note.* SD = Standard Deviation.

		Malevolent Pop-Ups	
Phase	Condition	*Mean*	*SD*
Old	Control	67.75%	.34
	N-IMCIT	58.82%	.35
	IMCIT	66.65%	.34
New	Control	67.75%	.36
	N-IMCIT	61.79%	.31
	IMCIT	73.35%	.30

We anticipated that participants in both MCIT groups, and in particular the I-MCIT group would be less likely to spot new cues. However, there is no statistical evidence to suggest that any form of MCIT led to participants accepting more New messages, despite an ~11.5% higher acceptance of these in the IMCIT versus the N-IMCIT condition in Phase 3. Of course, this could be a power issue, and future studies should consider this before ruling out the possibility that MCIT will not put people at a disadvantage in terms of spotting malevolent cues that they have not be trained to identify,

Time to Accept/Decline Pop-Up Messages

Next, we consider the time taken at make an accept/decline response. Noting that the time to accept/decline malevolent pop-ups was 5.37-s for younger adults in the [2] study, and 10-92-s for older adults in the [1] study. In the same studies, the times to accept genuine pop-ups were 5.47-s and 10.45-s respectively. Mean pop-up accept/decline times for the current study are displayed in Table 3 (with one outlier removed: z-scores >3.29, *p* < .001). Malevolent and genuine pop-ups, accept/decline times are noticeably lower (~1-2-seconds) than in e.g., [2]. Also, response times appear to reduce for each Group in Phase 3 versus Phase 1. The third, and somewhat counterintuitive observation, is that response times are noticeably lowest (and very short) for the Control Group (*M* 3.39 Phase 1, *M* 2.97 Phase 3).

A 3 (Training Group) x 2 (Phase) x 2 (Message Type) mixed factorial ANOVA revealed a significant main effect of Phase, *F*(1, 46) = 4.55, *MSE* = 2.87, *p* = .038, with less time taken in Phase 3 (*M* = 3.62-s) than Phase 1 (*M* 4.13). There was a significant main effect of Message Type, *F*(1, 46) = 5.46, *MSE* = .45, *p* = .024, with more time spent before making an accept/decline response for malevolent (*M* 3.99) than genuine (*M* 3.76-s) messages. There was a non-significant main effect of Training Group, *F*(2, 46) = 1.55, *MSE* = 4.47, *p* = .22, and none of the interactions were significant (*ps* > .08).

Table 3. Time (seconds) before making an accept/decline response to Malevolent and Genuine pop-ups during Phases 1 and 2 and across each Training Group. *Note.* SD = Standard Deviation.

		Malevolent Pop-Ups		Genuine Pop-Ups	
Phase	Condition	*Mean*	*SD*	*Mean*	*SD*
1	Control	3.45	2.65	3.33	2.23
	N-IMCIT	4.41	2.49	4.26	2.15

	IMCIT	4.77	2.99	4.57	3.05
3	Control	3.08	1.82	2.85	1.87
	N-IMCIT	3.66	1.57	3.51	1.55
	IMCIT	4.54	3.07	4.05	2.31

Contrary to our prediction, participants were faster to respond to pop-up messages in Phase 3 than Phase 1, and despite a non-significant Phase x Training Group interaction, this was the case for the IMCIT (*M Diff* -0.23-s) and N-IMCIT (*M Diff* -0.75-s) groups. Given that participants in the MCIT groups did not take additional time to try and identify cues to malevolence in malevolent pop-up messages, the improved detection performance must have been due to increased suspicion [3] and making better use of the very short inspection times to identify at least one cue to rouse suspicion.

Given much lower acceptance rates of malevolent pop-ups amongst the Control group in Phase 1 (Table 1), it was expected that those participants took more time to try and identify cues than in the MCIT groups. This was not the case. Also, their acceptance rate for malevolent pop-ups in Phase 3 *increased* by over 10% and the time taken to accept/decline messages reduced by almost half a second. Upon closer inspection of the data, three Control group participants almost always declined malevolent messages compared with the others whose performance was largely in line with those in the MCIT groups. However, they were not statistical outliers the $p < .001$ (z-scores > 3.29) level.

4 Limitations

There are limitations. First, there was no statistical evidence to suggest that those in the IMCIT group were better at identifying malevolent pop-ups than those in the N-IMCIT group, despite a trend. Perhaps using a leaderboard with individual position increasing after each task (e.g., 19th/20 after the first task, 1st after the last task) was not effective enough. This may be influenced by some participants potentially being aware that they were performing optimally and met with incongruent feedback to suggest otherwise. Competing with other people *in situ* may have promoted stronger social comparison and led to more intense cognitive processing strategies [3]. Second, within both MCIT conditions, participants had to identify whether they detected malevolent cues and then type a number corresponding to how many. This method meant that accuracy of malevolent cue identical could not be measured. Third, participants had one-minute per training task, only five tasks to complete, with each passage containing only three malevolent cues. They were also aware that there would be a maximum of three malevolent cues. This may not have been cognitively engaging enough. Finally, Control group participants were treating pop-ups with higher levels of suspicion in Phase 1. Ideally, this condition would be re-run to check for a possibly anomalous effect.

5 Implications

We successfully demonstrated that MCIT can be used as an intervention to reduce susceptibility to potentially fraudulent computer pop-ups. More cognitively engaging and demanding versions of the intervention might be even more effective. Whilst an incentivized version of this intervention did not quite result in further improvements in identifying fraudulent pop-ups, an improved version that better encourages social comparison might work better. Whilst there was no statistical evidence to suggest that MCIT can impair the ability to detect malevolence cues that participants had not been trained on, trends indicated a performance deficit, and methods to mitigate this need to be considered in the future development of MCIT interventions. Finally, it is important to note that time spent viewing malevolent pop-up messages was incredibly low and the propensity to accept (rather than decline) them was alarmingly high, both pre- and post-intervention, and even higher than in the studies by [1] and [2]. This further emphasises the vital need to develop interventions to help alleviate such susceptibility.

References

1. Morgan, P. L., Williams, E. J., Zook, N. A., & Christopher, G.: Exploring Older Adult Susceptibility to Fraudulent Computer Pop-Up Interruptions. In: International Conference on Applied Human Factors and Ergonomics, pp. 56--68. Springer, Cham (2018)
2. Williams, E. J., Morgan, P. L., & Joinson, A. N.: Press accept to update now: Individual differences in susceptibility to malevolent interruptions. Decision Support Systems. 96, 119--129 (2017)
3. Vishwanath, A., Harrison, B., & Ng, Y. J.: Suspicion, cognition, and automaticity model of phishing susceptibility. Communication Research, 1--21 (2016)
4. Anti-Phishing Working Group (APWG), https://www.antiphishing.org/resources/apwg-reports/
5. Department for Culture, Media & Sport.: Cyber security breaches survey 2017, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60918 6/Cyber_Security_Breaches_Survey_2017_main_report_PUBLIC.pdf
6. National Cyber Security Centre.: Weekly threat report 30th June 2017, https://www.ncsc.gov.uk/report/weekly-threat-report-30th-june2017
7. Perera, D.: Researcher: Sony hackers used fake emails. Politico, https://www.politico.com/story/2015/04/sony-hackers-fake-emails-117200
8. Forbes Cyber Security report, https://www.forbes.com/sites/ellistalton/2018/04/23/the-u-s-governments-lack-of-cybersecurity-expertise-threatens-our-infrastructure/#20d248be49e0
9. HM Government. National cyber security strategy 2016-2021, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/56724 2/national_cyber_security_strategy_2016.pdf
10. Conteh, N. Y., & Schmick, P. J.: Cybersecurity: risks, vulnerabilities and countermeasures to prevent social engineering attacks. International Journal of Advanced Computer Research. 6(23), 31 (2016)
11. Downing, D., Covington, M., Covington, M., Barrett, C. A., & Covington, S.: Dictionary of computer and internet terms. Barron's Educational Series, New York (2000)
12. Daintith, J., & Wright, E.: A dictionary of computing. Oxford University Press (2008)

13. Norton How To 2018, https://us.norton.com/internetsecurity-how-to-the-importance-of-general-software-updates-and-patches.html
14. Altmann, E. M., Trafton, J. G., & Hambrick, D. Z.: Momentary interruptions can derail the train of thought. Journal of Experimental Psychology: General. 143(1), 215--226 (2014)
15. Hodgetts, H. M., & Jones, D. M.: Interruption of the Tower of London task: support for a goal-activation approach. Journal of Experimental Psychology: General. 135(1), 103--115 (2006)
16. Monk, C. A., Trafton, J. G., & Boehm-Davis, D. A.: The effect of interruption duration and demand on resuming suspended goals. Journal of Experimental Psychology: Applied. 14(4), 299--313 (2008)
17. Altmann E. M., Trafton J. G.: Memory for goals: An activation-based model. Cognitive Science. 26, 39--83 (2002)
18. Altmann E. M., Trafton J. G. Timecourse of recovery from task interruption: Data and a model. Psychonomic Bullletin & Review. 14(6), 1079—1084 (2017)
19. Cacioppo, J. T., Petty, R. E., & Feng Kao, C.: The efficient assessment of need for cognition. Journal of personality assessment. 48(3), 306--307 (1984)
20. Anandpara, V., Dingman, A., Jakobsson, M., Liu, D., & Roinestad, H.: Phishing IQ tests measure fear, not ability. In: International Conference on Financial Cryptography and Data Security, pp. 362--366. Springer, Berlin (2007)
21. Downs, J. S., Holbrook, M. B., & Cranor, L. F.: Decision strategies and susceptibility to phishing. In: Proceedings of the second symposium on Usable privacy and security, pp. 79--90. ACM (2006)
22. Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J.: Teaching Johnny not to fall for phish. ACM Transactions on Internet Technology (TOIT). 10(2), 1--30 (2010)
23. Clifford, M. M.: Effects of competition as a motivational technique in the classroom. American Educational Research Journal. 9(1), 123--137 (1972)
24. Aleman, J. L. F., de Gea, J. M. C., & Mondéjar, J. J. R.: Effects of competitive computer-assisted learning versus conventional teaching methods on the acquisition and retention of knowledge in medical surgical nursing students. Nurse education today. 31(8), 866--871 (2011)
25. Festinger, L.: A theory of social comparison processes. Human relations. 7(2), 117--140 (1954)
26. Peirce, J. W.: PsychoPy—Psychophysics software in Python. Journal of Neuroscience Methods. 162(2), 8--13 (2007)