Received: 9 January 2020 | Revised: 9 November 2020 | Accepted: 24 November 2020

DOI: 10.1111/isj.12318

RESEARCH ARTICLE

WILEY

A comparison of features in a crowdsourced phishing warning system

Christopher Nguyen[1] | Matthew L. Jensen[2] | Alexandra Durcikova[2] | Ryan T. Wright[3]

[1]Logistics Management Institute, Tysons, Virginia

[2]Price College of Business, University of Oklahoma, Norman, Oklahoma

[3]McIntire School of Commerce, University of Virginia, Charlottesville, Virginia

Correspondence
Christopher Nguyen, Logistics Management Institute, 7940 Jones Branch Dr, Tysons, VA 22102, USA.
Email: cnguyen@lmi.org

Funding information
Directorate for Social, Behavioral and Economic Sciences, Grant/Award Number: 1421580

Abstract

Initial research on using crowdsourcing as a collaborative method for helping individuals identify phishing messages has shown promising results. However, the vast majority of crowdsourcing research has focussed on crowdsourced system components broadly and understanding individuals' motivation in contributing to crowdsourced systems. Little research has examined the features of crowdsourced systems that influence whether individuals utilise this information, particularly in the context of warnings for phishing emails. Thus, the present study examined four features related to warnings derived from a mock crowdsourced anti-phishing warning system that 438 participants were provided to aid in their evaluation of a series of email messages: the number of times an email message was reported as being potentially suspicious, the source of the reports, the accuracy rate of the warnings (based on reports) and the disclosure of the accuracy rate. The results showed that crowdsourcing features work together to encourage warning acceptance and reduce anxiety. Accuracy rate demonstrated the most prominent effects on outcomes related to judgement accuracy, adherence to warning recommendations and anxiety with system use. The results are discussed regarding implications for organisations considering the design and implementation of crowdsourced phishing warning systems that facilitate accurate recommendations.

KEYWORDS
accuracy, crowdsourcing, human-automation interaction, phishing, report source, warning

1 | INTRODUCTION

Most users are familiar with tempting or frightening phishing messages (eg, emails, texts, social media posts) that prompt us to click on a link or download an attachment. Responding to these fraudulent messages allows criminals to access sensitive data and/or to instal malicious software. Phishing is more prevalent now more than ever. The Anti-Phishing Working Group reported that phishing activity is now at a 3-year high, with cybercriminals using COVID-19-themed attacks against healthcare facilities and the recently unemployed (Anti-Phishing Working Group, 2020). Verizon's Data Breach Report, which provides details on over 2013 confirmed data breaches and 41 686 security incidents, indicated that malware used in the breaches was delivered using phishing emails in 93% of the cases (Verizon RISK, 2019). Clearly, phishing remains a scourge on modern organisations that often results in data breaches, financial losses and loss of intellectual property. Additionally, organisations must also handle losses and risks associated with their customers, as well as threats to their reputation (Syed, 2019). Phishing messages entice individuals to respond by mimicking legitimate messages from genuine sources. For example, a successful phishing attack may target individuals using messages that appear to come from a familiar (eg, coworker, service provider) or authoritative source (eg, manager, government; Bose & Leung, 2008). Examples of recent successful phishing attacks include prominent healthcare providers (ie, Los Angeles County hospitals; Karlamangla, 2019) and a message from the UK's tax office informing of COVID-19-related grants (Natvig et al., 2020).

Within recent years, researchers have focussed a large amount of attention on information technology security to reduce these data breaches and compromises, acknowledging that individual employees are often the greatest liability to organisations (eg, Lowry & Moody, 2015; Willison et al., 2018). In fact, phishing susceptibility has become an increasingly popular area of research over the past 20 years (see Appendix A). One common theme in this literature is that individuals often have a difficult time identifying phishing emails and websites (c.f. Jensen et al., 2013; Moody et al., 2017; Wang et al., 2012; Wright et al., 2014). Additionally, behavioural researchers have also focussed on three fundamental types of phishing susceptibility research: (a) the examination of properties of the email messages and websites that make individuals more susceptible to act on a phishing message such as clicking on the link or downloading an attachment (eg, Wright et al., 2014); (b) the examination of demographic and psychographic properties of individuals who are susceptible to phishing attacks, including dispositional factors (eg, trust, suspicion) and experience factors (eg, security knowledge and computer self-efficacy; Wright & Marett, 2010), gender (eg, Sjouwerman, 2019) and age (eg, Kumaraguru et al., 2009; Sheng et al., 2010) and (c) the examination of methods such as training (eg, Dodge et al., 2007) and warnings (eg, Volkamer et al., 2017) to mitigate phishing susceptibility.

To prepare for these types of attacks, organisations invest in hardware and software that automatically detect and filter phishing messages and associated websites (Hong, 2012; Ramesh et al., 2014), as well as invest in warning systems that alert individuals when messages appear suspicious (Petelka et al., 2019). Past research has demonstrated that automated filtering and warning systems and tools can be extremely effective in preventing individuals from responding to phishing attacks (Gupta et al., 2017; Tan et al., 2016). Researchers have also noted the importance of viewing the design of warnings as an integral part of systems design and following certain guidelines and principles to employ effective warning systems (Wogalter et al., 2021). For example, Carpenter et al. (2014) used an experimental setting to show that the term 'hazard' was most effective for reducing disclosure of information online. However, automated systems can be limited in the protection they offer against new permutations of phishing attacks. As phishing attacks become more customised for their intended targets, they may not include the typical cues used by automated systems and go undetected (Symantec, 2014). Additionally, the effectiveness of automated

systems can be further limited if individuals do not pay attention to warnings or recommendations they receive from the systems due to habituation or generalisation (eg, Kirwan et al., 2020; Vance et al., 2018; Vishwanath et al., 2011), difficulty in understanding security warnings (eg, Anderson et al., 2018; Felt et al., 2015) and other contextual factors and user concerns (eg, Reeder et al., 2018).

As an additional line of defense in their layered security approach, organisations also train their members to identify and report new permutations of phishing attacks (Kumaraguru et al., 2010). Members' reports can serve as inputs for improving automated filtering and warning tools by making quicker updates to automated systems. Furthermore, rather than having members work in isolation when detecting phishing messages, utilising crowds to report potential phishing websites can be a more effective strategy for increasing accuracy and reducing time in verifying the legitimacy of suspicious websites (Liu et al., 2011). In this manner, crowdsourcing can help individuals work together to collectively identify phishing messages by utilising reports from others who have previously flagged similar suspicious messages. Researchers (eg, Liu et al., 2011) and organisations (eg, Google, Microsoft, PhishMe) have begun investigating coordination and collaboration techniques for harnessing the abilities of large numbers of people through crowdsourcing to combat phishing attacks. Crowdsourcing the detection of phishing messages is a promising area of research, with efforts focussing on incentivising participation (Jensen, Durcikova, & Wright, 2017) or providing publicly available databases for tracking phishing websites (Dobolyi & Abbasi, 2016).

There is, however, little empirical research to provide insight into effective approaches for implementing crowdsourced phishing warning systems. Most research has instead focussed on detecting the website and providing warnings after the user has clicked on a link (eg, Abbasi et al., 2015; Tan et al., 2016) or implementing crowdsourced systems more broadly and outside of phishing contexts (eg, Durward et al., 2020; Love & Hirschheim, 2017). Although utilising the power of crowdsourcing seems to be a promising strategy for combatting phishing, what these crowdsourced systems should look like and how they should operate within this context have not yet been determined. Furthermore, as organisations incorporate crowdsourcing to combat phishing, additional research is needed to understand what features influence the efficacy and acceptance of warnings from crowdsourced systems.

Thus, the goal of this study was to examine the features of crowdsourced phishing warning systems that may influence how suspicious email messages are identified and evaluated. We build upon existing knowledge of automated phishing warning systems by incorporating crowdsourcing elements. Therefore, we drew upon and integrated two theoretical perspectives: human-automation interaction and crowdsourcing. Specifically, we examined four features related to the recommendations derived from a prototypical crowdsourced phishing warning system: the number of times an email message was reported as being potentially suspicious, the source of the reports (human vs automated system), the accuracy rate of the warnings based on reports (40% vs 80%) and disclosure of the accuracy rate (disclosed vs not disclosed). Some of these features (ie, number of reports and report source) are pertinent to crowdsourced systems and have only been studied in other contexts (eg, e-commerce; Benlian et al., 2012; Kim & Gambino, 2016; Lin, 2014), whereas other features (ie, accuracy rate and accuracy disclosure) have been previously examined in research investigating trust in human-automation interactions (eg, Parasuraman et al., 2000; Wang et al., 2009) and are also applicable to crowdsourced systems. Integrating the research from these two literatures can help disentangle the features that prompt individuals to use recommendations from crowdsourced systems and aid in the development of better anti-phishing tools. Additionally, because these various features will likely have differing effects on whether individuals use these systems and benefit in terms of performance, determining the appropriate combination of features will be essential to building more effective crowdsourced systems for phishing email detection.

2 | BACKGROUND

To develop an expanded understanding of how individuals may respond to and utilise crowdsourced phishing warning systems, we first probe important nuances relevant to features influencing individuals' trust in automated

systems (accuracy rate and disclosure of the accuracy rate) and then discuss how combining these features with unique features of crowdsourcing (human source for reports and number of times messages were reported as being suspicious) can lead to more effective phishing detection. We then evaluate the effect of these features on outcomes related to the influence of crowdsourced system warnings and users' judgement accuracy when aided by the system.

2.1 | Human-automation interaction

Automation refers to the execution of a function by a machine that was previously (or could have been) carried out by another person (Parasuraman & Riley, 1997). Automation varies in terms of functions a system performs, as well as the amount of autonomy an individual has in terms of control over the automated system (Parasuraman et al., 2000). For example, lower levels of automation may only have a system organise and filter relevant information to provide recommendations to users and leave them with the responsibility to make and enact decisions, whereas higher levels of automation may have a system gather information, make a decision and then enact the decision with limited or no manual intervention (Miller & Parasuraman, 2007). Various combinations of these types and levels of automation have been shown to affect users' task performance, mental workload, situational awareness and complacency (Parasuraman et al., 2000). Although introducing automated systems is typically intended to help individuals perform more optimally, individuals may also rely too much and inappropriately use the system (misuse) or underutilise the system and ignore recommendations (disuse; Lee, 2008; Parasuraman & Riley, 1997). In some instances, this decision to trust, and ultimately whether to misuse or disuse automation, could lead to detrimental consequences (Hussein et al., 2020). In particular, Endsley (2017) highlights that as automated systems become increasingly reliable and robust, individuals will need to remain just as increasingly vigilant when monitoring system performance and intervene when necessary.

Although the topic of human-automation interaction has been a long-studied topic within information systems research, this area of research remains a promising area of study as technological advances require greater exploration as to how humans can better collaborate, understand and work alongside automated systems. For example, researchers have begun examining emerging themes within this stream of human-automation interaction literature, including how to better collaborate in teams with automation (eg, Lakhmani et al., 2020) and maximising performance through knowledge sharing from humans and automation (eg, Jiao et al., 2020). Attitudes towards automated systems often influence whether individuals decide to use them when working on tasks or when tasks are interrupted with warnings. In particular, trust in automation plays a major role in whether individuals are willing to follow warning recommendations or utilise automation when necessary (Hoff & Bashir, 2015). Marsh and Dibben (2003) distinguish among three factors that influence human-automation trust: (a) dispositional trust (eg, individual differences such as culture, age, gender and personality); (b) situational trust (eg, external environmental factors and internal, context-dependent characteristics of the individual) and (c) learned trust (eg, evaluations of an automated system based upon its pre-existing reputation, performance and design features). Additionally, a system's perceived usefulness and ease of use, as well as individuals' attitudes and behavioural intentions to use the system, can influence trust and ultimately affect system utilisation or recommendation acceptance (Davis, 1989).

Although scholars have debated whether systems can be the recipients of trust (eg, Lankton et al., 2015, 2016), there is evidence that trust in systems and trust in other people are not fundamentally different (Wang & Benbasat, 2005). Because all technological systems have human designers or operators, individuals can essentially form their bases of trust in systems in a similar manner as they would with another person. Similar to the notion of trust with other humans, trust in automation is likely to vary based on how well an automated system executes a task (Muir, 1988). In this vein, research has consistently found that performance-based trust, or the attitudes shaped by the accuracy (or reliability) of the automated system, often governs the decision to utilise the system, as well as the extent to which it is used (Merritt & Ilgen, 2008; Parasuraman et al., 2000; Wang et al., 2009). Unreliable

automation, particularly in the case of systems that frequently give false alarms, lowers trust and may result in disuse, thus undermining the potential performance benefits from utilising automation (Parasuraman et al., 2000). Similarly, Hancock et al.'s (2011) meta-analytic results reveal that factors related to the system itself (eg, reliability and performance) had the greatest impact on individuals' trust towards the system. Within the context of identifying phishing emails, individuals have been shown to be less likely to follow recommendations of automated phishing warning systems that have longer response times and lower detection rates (Abbasi & Chen, 2009; Wu et al., 2006).

Similarly, transparency of automation accuracy has also been found to influence trust in automation. Studies have found that disclosing information about system accuracy can better facilitate trust and improve human-automation task performance (Hoff & Bashir, 2015; Wang et al., 2009). In this vein, Lee and See (2004) recommended that when designing automated systems, it is important to inform individuals of the automation's past performance to elicit more trust in automation. Although designing automated systems with features such as improved accuracy and transparency are important for eliciting trust and ultimately better system acceptance and utilisation, trained individuals should also be working alongside technology to combat phishing within organisations. However, when individuals make decisions regarding whether an email is phishing, they are often working individually or in isolation. Research has shown that individuals are often overconfident in their ability to protect themselves from security-related attacks and may choose to ignore warnings, even though the warnings come from highly accurate systems (Kumaraguru et al., 2007). Additionally, challenges related to habituation and comprehension of warnings may compound these issues (eg, Anderson et al., 2018; Kirwan et al., 2020). For example, with email management becoming a daily task for most individuals nowadays, it is likely that the same behaviours are being repeated in recurring contexts without much conscious thought put behind the process (Wood & Rünger, 2016). Thus, finding ways to promote increased trust in automated system warnings, while also encouraging individuals to work collectively to identify suspicious messages, is needed. To this aim, accessing the potential of many users through crowdsourcing may be a particularly effective method for combating phishing.

2.2 | Crowdsourcing

Crowdsourcing refers to a technique in which a task or job is outsourced to a large, undefined group of individuals (Howe, 2006). Crowdsourced systems are unique from automated systems in that they contain a manual element in their design where individuals provide inputs or contributions to the system. Although the applications of crowdsourcing can vary (eg, idea generation, microtasking, open source software), the overall intent of utilising crowdsourcing is to accomplish tasks that go beyond what would be possible when individuals or automated systems operate on their own (Hossain & Kauranen, 2015). Organisations such as Wikipedia and Amazon (via their platform Mechanical Turk) have greatly benefitted from incorporating crowdsourcing in their services (Howe, 2006; Love & Hirschheim, 2017). Although the premise of crowdsourcing is that it relies upon a large number of contributions to be effective, the quality of the contributions can often vary greatly in crowdsourced systems (Thuan et al., 2013). As many individuals contribute to crowdsourced systems, there exists the potential for erroneous or low-quality contributions, which may undermine the successful completion of the crowdsourced task and the utilisation of outputs (eg, recommendations) made by the system. Even so, Warby et al. (2014) found that crowdsourcing contributions from non-experts on a subject matter can be an effective way for accomplishing difficult tasks and can result in even greater performance than from automated methods.

The decision to use crowdsourcing within an organisation, however, is a complicated issue. Due to the voluntary, and often anonymous, nature of contributions made to crowdsourced systems, developing these systems involves navigating various components related to user management (eg, evaluating and coordinating users), task management (eg, designing and assigning tasks), contribution management (eg, evaluating and selecting contributions) and workflow management (eg, defining and managing workflow; Hetmank, 2013). Additionally, factors related to an organisation's environment, management, people and tasks performed play a large role in determining whether

integrating crowdsourcing into organisational practices would be appropriate and effective (Thuan et al., 2013). Within an information security context, these factors are also likely to influence how phishing and other cybersecurity threats manifest in organisations, such as the types of phishing emails received, organisational strategies for combatting phishing and employees' ability to detect these emails.

Within recent years, crowdsourcing has been increasingly utilised in security services such as detecting phishing attacks, detecting cybersecurity threats and browser security (Liu et al., 2011; Moore & Clayton, 2008). For example, organisations such as Google, Microsoft and PhishTank have utilised crowdsourcing to develop and update blacklists which contain repositories of manually verified phishing websites. Utilising this 'wisdom of crowds' allows these organisations to contribute to a central knowledge system where content is constantly added, reviewed and modified (Moore & Clayton, 2008; p. 1). As phishing attacks become more dynamic, complex and prevalent for organisations, the need for integrating crowdsourced systems into organisational security strategies is becoming increasingly important.

As shown in Appendix A, the research on and understanding of phishing susceptibility has become much more nuanced, highlighting the importance of integrating these findings into crowdsourced warning system design. Crowdsourced warning systems have the potential to provide timely information about suspicious incoming email messages, and individuals can use these systems in detecting and informing others of phishing attacks. For example, individuals who encounter phishing messages in their inboxes can report the message, where reports would be sent to a central organisational repository. The information sent to this repository can then be distributed both internally (eg, to individuals within the respective organisation) or externally (eg, to individuals at other organisations within similar sectors) to aid in decision-making. These crowdsourcing methods, however, have not been rigorously examined, as most of the research on phishing warning systems has primarily focussed on fully automated systems that flag suspicious websites (eg, Abbasi et al., 2010; Ramesh et al., 2014) and training users to individually identify suspicious messages (Jensen, Dinger, et al., 2017; Kumaraguru et al., 2007, 2009).

Furthermore, the majority of research on crowdsourcing from an information systems perspective has focussed on crowdsourced systems broadly (rather than in phishing detection contexts) and has primarily examined issues related to understanding system components and elements (eg, Pedersen et al., 2013), designating types of crowdsourced systems (eg, Geiger et al., 2012) or identifying characteristics of people who comprise the crowds and what features and incentives influence individuals' motivation to continue to contribute to these systems (eg, Durward et al., 2020; Estellés-Arolas & González-Ladrón-de-Guevara, 2012; Finnerty et al., 2013). There is scant empirical evidence available regarding which features embedded in crowdsourced phishing warning systems provide protection for individuals against phishing attacks. Although researchers have begun developing frameworks and models to better conceptualise crowdsourced systems (eg, Knop et al., 2017; Love & Hirschheim, 2017; Zuchowski et al., 2016), there is little research that examines the features to incorporate in the design of crowdsourced systems that will aid in decision-making specific to a phishing detection context, maximise the acceptance of the system's recommendations and improve detection accuracy.

3 | HYPOTHESES

The purpose of the present study was to examine the features of reports made to crowdsourced phishing warning systems that influence individuals' utilisation of warning recommendations when evaluating email messages. When determining the list of features to examine, we drew upon previous research on system features that have shown to be impactful in whether or not individuals trust or utilise automation, as well as features that are unique to crowdsourced systems that may be influential but have not been previously examined within a phishing context. Thus, we manipulated four different features in a mock crowdsourced phishing warning system: (a) the number of times an email message was reported as being potentially suspicious; (b) the source of the reports; (c) the accuracy rate of the warning based on the reports and (d) disclosure of the accuracy rate. We examined the efficacy of these

four features on two types of outcomes: (a) accuracy of message judgements and (b) influence of crowdsourced system warnings. These outcomes were measured to capture how these features influence individuals' task performance but also their perceptions of the crowdsourced phishing warning system. Judgement accuracy was measured by hits (ie, correctly identifying a phishing message as phishing) and false-positives (ie, incorrectly identifying a legitimate message as phishing). Influence of the crowdsourced system warnings was measured by how many times participants followed the systems' warnings (ie, adherence to warning recommendations), as well as participants' anxiety with using the crowdsourced system. Judgement accuracy and influence of warnings are related but are important to examine separately. For example, adherence to warnings may lead to incorrect judgements, and rejection of warnings may produce correct judgements. Additionally, the features that drive warning efficacy may be different than those inducing judgement accuracy. For example, anxiety with security-related tools can be a deterrent in individuals' decision to utilise recommendations in an appropriate manner (Venkatesh & Davis, 2000). Because the level of control an individual has over an automation can vary also (Parasuraman et al., 2000), the focus of the present study was on the design of a mock crowdsourced phishing warning system with a low degree of automation where individuals are provided with warning recommendations for potentially suspicious emails, but the decision to follow those recommendations is ultimately left to the individual.

As mentioned previously, one of the features of an effective crowdsourced system is a large number of contributions. In terms of the number of reports, Liu et al. (2011) found that clustering similar phishing websites together led to greater accuracy in identifying phishing messages compared to providing individuals with single instances of phishing websites. Similarly, when individuals are provided with warning recommendations where the same email messages have been reported multiple times as being potentially suspicious, individuals may be more likely to follow those recommendations. Numerous reports provide tangible evidence of suspicion and are likely to increase confidence in the system's recommendation. Additionally, having a higher number of reports for a suspicious message may also reduce individuals' anxiety when using the warning system by reducing ambiguity surrounding the true nature of the message. To this end, Lin (2014) found that having a larger volume of user recommendations (ie, number of reviews) made recommendations appear more trustworthy and reduced uncertainty, leading to greater utilisation of recommendations towards product sales. Additionally, Kim and Gambino (2016) found that higher bandwagon cues (ie, star ratings and number of reviews) were associated with more positive user attitudes and user intentions towards both websites and recommendations. Thus,

Hypothesis 1. A higher number of reports (either human or automated) in phishing warning systems will be associated with (a) adherence to more warning recommendations and (b) less anxiety using the system.

In contrast to fully automated decision aids that gather and analyse data without intervention or supervision, the distinguishing feature of crowdsourced systems is that other people are the sources of information upon which system recommendations are based. Therefore, the source of the information that forms the recommendation may alter whether or not the recommendation is followed. Although human trust in automation is complex and involves various factors, de Vries et al. (2003) state that there is a 'fundamental bias' for individuals to trust one's ability over those of an automated system (p. 734). Further, this bias is exacerbated when one has high self-confidence in one's own technological abilities or when the task can be achieved without assistance from an automated decision aid (Madhavan & Phillips, 2010). In general, individuals are often overly confident in their ability to protect themselves against security-related attacks and may overestimate their abilities in recognising phishing emails (Kumaraguru et al., 2007). Following this rationale, individuals may project this same sense of confidence into others and be more likely to trust the judgement of other individuals. Research has also found that consumer recommendations/reviews (ie, recommendations from individuals) were more effective in generating greater trusting beliefs compared to system recommendations/reviews (Ashraf et al., 2019; Benlian et al., 2012). Similarly, Lin (2014) found that user recommendations were more effective at driving product sales compared to system recommendations. Although these studies were conducted within the context of e-commerce transactions, the rationale behind individuals' increased trust towards recommendations from other individuals vs automated systems may also be applicable to a phishing

context such that individuals are more willing to accept more recommendations and feel less anxiety from using such a crowdsourced warning system when there are people involved in the evaluation. Thus,

Hypothesis 2. Reports from human sources in phishing warning systems will be associated with (a) adherence to more warning recommendations and (b) less anxiety using the system compared to reports which come from automated sources.

Previous research has consistently shown that working with information or alongside systems that are more accurate will likely result in better performance (Hiltz & Turoff, 1985). When individuals have access to more accurate information, they should make better and more appropriate decisions. Although increased accuracy may result in misuse or overreliance on systems (Parasuraman & Riley, 1997), within phishing contexts, greater accuracy of automated tools has been highly effective in reducing successful phishing attacks (Abbasi & Chen, 2009). As individuals have direct experience with accurate phishing warning systems, individuals may also be more likely to trust and, therefore, accept more recommendations. Additionally, when the system more accurately evaluates suspicious email messages, individuals should have less anxiety using the system because they have greater trust in the system and sense less ambiguity during message evaluation. Therefore,

Hypothesis 3. Higher accuracy rates of reports (either human or automated) in phishing warning systems will be associated with (a) adherence to more warning recommendations and (b) less anxiety using the system.

In addition to their independent effects on the outcomes, we expected to see combined effects for the two features unique to crowdsourced systems (ie, report source and number of reports). In other research regarding acceptance of crowdsourced information (eg, contributions to online knowledge bases), the combination of multiple information quality indicators increased acceptance more than the sum of indicators' individual effects (Meservy et al., 2014). In other words, the effects from features of crowdsourced systems may exceed their additive effects. Due to the detrimental consequences associated with falling for a phishing attack, it is possible that individuals may trust and utilise the recommendations from other people because they see others as extensions of themselves who are also trying to protect themselves and their assets. We anticipate that these effects will be even stronger when there are a large number of reports, as more contributions may reduce the uncertainty surrounding the status of the message and even promote credibility. Although previous researchers did not incorporate both comparisons of human vs automated recommendations and lower vs higher number of recommendations in a single study to examine individuals' trusting beliefs and decisions to utilise recommendations, we use their findings to predict that a combination of a higher number of reports coming from human sources will yield more positive outcomes as it relates to adherence to more warning recommendations and less anxiety using systems (Ashraf et al., 2019; Benlian et al., 2012; Lin, 2014). Thus,

Hypothesis 4. The number of reports and the source of reports will interact such that when the number of reports is higher and the reports come from human sources, individuals will have (a) adherence to more warning recommendations and (b) less anxiety using the system.

Additionally, we also expected to see combined effects for the two features influencing human trust in automation (ie, accuracy rate of reports and disclosure of accuracy rates). In other words, the effect of disclosure of accuracy rates may not have an independent effect because it is contingent upon the actual accuracy rate of the reports. Although providing more accurate information may be a critical characteristic of these systems, it may be necessary to disclose to individuals the system's accuracy or performance to elicit greater trust (Hoff & Bashir, 2015; Lee & See, 2004; Wang et al., 2009). By disclosing the accuracy rate, individuals can limit the process of sensemaking to determine whether a system is trustworthy or not because they know exactly how effective the system is and ultimately how much reliance to place upon it. For example, if individuals are informed about how a system is not very accurate in detecting phishing

messages, they may have reservations about its recommendations and decide to rely upon their own detection abilities instead, which could create instances of disuse and mistrust of the system. In contrast, if individuals are informed about how a system has high accuracy, they may be in a better position to protect themselves, not only because following more accurate recommendations will lead to greater accuracy but also because there are less doubts regarding the system's trustworthiness or negative feelings towards the system. Because phishing detection has negative consequences associated with incorrect decision-making, we anticipate that there will be more advantageous and positive outcomes associated when disclosure is combined with high (vs low) accuracy rates as it relates to greater judgement accuracy, adherence to more warning recommendations and less anxiety with system use. Therefore,

Hypothesis 5. The accuracy rate of reports and the disclosure of accuracy rates will interact such that when the accuracy rate of reports is higher and the accuracy rate is disclosed, individuals will have (a) higher hit rates; (b) lower false-positive rates; (c) adherence to more warning recommendations; and (d) less anxiety using the system.

In addition to the combined effects of features unique to crowdsourced systems and combined effects of features influencing human trust in automation individually, integrating these two research literatures and examining the unique combinations of these features can provide insight in developing successful crowdsourced phishing warning systems. Because the quality of crowdsourced reports or contributions can vary greatly, it is important to not only consider crowdsourced report features regarding the sheer number of reports and the report sources but also the accuracy of those reports when implementing the most effective design features. This consideration may be particularly important in a phishing detection context due to the potentially detrimental consequences associated with phishing decisions vs e-commerce decisions that do not have such negative repercussions. In other words, accuracy of recommendations plays a much greater role in a context in which there is a definite correct and incorrect answer in a given scenario. For example, as mentioned previously, research findings have indicated greater bias towards trusting recommendations from other individuals vs automation; however, Pearson et al. (2019) describe how this relationship may be dependent on the perceived expertise of those individuals. Similarly, although previous research has shown the benefits of having a large number of reports from a crowdsourced system for influencing individuals' positive perceptions and utilisation of the system, the accuracy of these reports may be particularly important in contexts such as phishing detection to enable individuals to effectively utilise them. Thus, more effective phishing detection, utilisation of recommendations and positive feelings towards the system may be associated not only with highly accurate reports that come from human sources but also highly accurate reports that are higher in quantity. Therefore,

Hypothesis 6. The source of reports and the accuracy rate of reports will interact such that when the reports come from human sources and the accuracy rate of reports is higher, individuals will have (a) higher hit rates; (b) have lower false-positive rates; (c) adherence to more warning recommendations; and (d) less anxiety using the system.

Hypothesis 7. The number of reports and the accuracy rate of reports will interact such that when the number of reports is higher and the accuracy of reports is higher, individuals will have (a) higher hit rates; (b) lower false-positive rates; (c) adherence to more warning recommendations; and (d) less anxiety using the system.

4 | METHOD

4.1 | Participants

Students enrolled in introductory Management Information Systems courses at a Midwestern University were recruited to participate in the study in exchange for extra credit for a course. A total of 488 students began the study.

However, data from 50 participants were removed from analyses due to incomplete responses or failure to correctly answer attention check items. This yielded a final sample of 438 participants (63% were male). Participants ranged in age from 18 years to 54 years, with a mean age of 21.11 (SD = 3.32). In terms of classification, participants consisted of 5% freshmen, 36% sophomores, 43% juniors and 16% seniors, with 87% identifying a business major as their field of study.

4.2 | General procedures

A 2 (number of reports: one or five times) × 2 (report source: human or automated) × 2 (accuracy rate: 40% or 80%) × 2 (disclosure of accuracy rate: disclosed or not disclosed) factorial design was used for the study. Participants were asked to complete a survey that was constructed through the online survey platform qualtrics. Once participants began the study, the opening instructions explained that the purpose of the study was to pilot test a new email system and examine how individuals interpret different email messages. Participants first completed a series of Likert scale measures of the covariates related to their email and Web usage, as well as their past phishing experience. Participants then completed an email identification task in which they were asked to read through 10 email messages (five legitimate and five phishing) and determine whether each message was a phishing message. The email messages were developed based on real-world examples of legitimate and phishing emails. Each phishing email message contained a minimum of two cues that are commonly associated with phishing email messages (eg, suspicious-looking link or use of influence tactics within the message content; Cialdini, 2009). The 10 emails were displayed in a randomised order. Appendix B displays all phishing and legitimate email messages included in the study.

During the email identification task, participants received warning messages for some email messages to aid in their decision-making. Five of the 10 email messages (both phishing and legitimate emails) were randomly chosen to contain warning messages. Participants were randomly assigned to conditions in which the warning messages were manipulated along the four features: (a) number of the times email messages were reported; (b) report source; (c) accuracy rate and (d) disclosure of the accuracy rate. The combination of the four warning message features resulted in participants being assigned to one of 16 different study conditions. The number of times email messages were previously reported was either one time or five times. The report source was either a human source (eg, different university students reporting emails) or an automated source (eg, the same automated anti-phishing software program reporting emails). The accuracy rate of the reports made by the source was either 40% (eg, two phishing messages correctly labelled with warning messages and three legitimate messages incorrectly labelled with warning messages) or 80% (eg, four phishing messages correctly labelled with warning messages and one legitimate message incorrectly labelled with a warning message). Participants were either disclosed or not disclosed the accuracy rate of the reports before beginning the email identification task. Appendix C displays the instructions that participants received before the task explaining the warning messages and an example of the warning conditions.

After completing the email identification task, participants completed additional Likert-scaled measures related to their experiences during the task and their perceptions of the phishing warning system. Finally, participants completed demographic measures and were debriefed about the study.

4.3 | Covariates

Several covariates that have been examined in previous phishing research (eg, Wright & Marett, 2010) were included to isolate the effects of the experimental manipulations. These covariates included disposition to trust (Hurley, 2006; McKnight et al., 2002), mindfulness in technology (Thatcher et al., 2018), perceived internet risk (Malhotra et al., 2004; Pelaez et al., 2019), computer self-efficacy (Compeau & Higgins, 1995) and self-reported expertise in

identifying phishing messages, email experience and computer game interest. Appendix D (Table D1) lists the text and measurement properties of the items used in the study.

5 | RESULTS

In general, four 2 (number of reports: one or five times) × 2 (report source: human or automated) × 2 (accuracy rate: 40% or 80%) × 2 (disclosure of accuracy rate: disclosed or not disclosed) between-subjects design analyses of covariance (ANCOVAs) were conducted for each of the four dependent variables (ie, hit rate, false-positive rate, adherence to warning recommendations and anxiety with system use). Main effects were examined for Hypotheses 1 through 3. Additionally, interaction effects were examined for Hypotheses 4 through 7. Appendix D (Table D2) lists the means, SDs, reliabilities, and correlations for the study variables.

Table 1 displays the adjusted means and standard errors for the dependent variables by study conditions. Table 2 reports the results of the ANCOVAs predicting each of the four dependent variables.

5.1 | Main effects

As seen in Table 2, after controlling for covariates, there was not a significant main effect of number of reports on the adherence to warning recommendations, $F(1,414) = 0.298$, $P = .59$, or anxiety with system use, $F(1,414) = 0.072$, $P = .79$. Thus, Hypothesis 1 was not supported.

In addition, there was not a significant main effect of report source on the adherence to warning recommendations, $F(1,414) = 0.457$, $P = .50$, or anxiety with system use, $F(1,414) = 0.001$, $P = .98$. Thus, Hypothesis 2 was not supported.

As seen in Table 2, after controlling for the covariates, there was a significant main effect of accuracy rate on the adherence to warning recommendations, $F(1,414) = 10.65$, $P = .001$. Individuals accepted significantly more recommendations when the accuracy rate of the reports was 80% (M = 2.89, SE = 0.09) than when the accuracy rate of the reports was 40% (M = 2.49, SE = 0.09). Additionally, there was a significant main effect of accuracy rate on anxiety with system use, $F(1,414) = 7.07$, $P = .01$. Anxiety with system use was significantly lower when the accuracy rate of reports was 80% (M = 4.05, SE = 0.07) than when the accuracy rate of reports was 40% (M = 4.31, SE = 0.07). Thus, Hypothesis 3 was supported.

5.2 | Interaction effects

As seen in Table 2, there was a significant interaction between the source of reports and number of reports on the adherence to warning recommendations, $F(1,414) = 6.10$, $P = .01$. As shown in Figure 1, there was a significant difference in the adherence to warning recommendations when the number of reports was higher (ie, reported five times) but not when the number of reports was lower (ie, reported one time). Specifically, when the number of reports was higher, individuals accepted significantly more recommendations from human sources than from automated sources.

Additionally, there was a significant three-way interaction between the source of reports, number of reports and disclosure of the accuracy rate on anxiety with system use, $F(1,414) = 4.53$, $P = .03$. As shown in Figure 2, there was a significant interaction between disclosure of accuracy rates and number of reports on anxiety with system use when the source of reports was a human but not when the source of reports was automated. For human report sources, the significant difference in anxiety with system use was found when the number of reports was higher (ie, reported five times) vs lower (ie, reported one time). Specifically, when the number of reports was higher, anxiety

TABLE 1 Results of analyses of covariance: Adjusted means and SEs by study conditions and outcomes

Condition				*n*	Hit rate			False-positive rate			Adherence to warning recommendations			Anxiety with system use		
					M	SE	95% CI	M	SE	95% CI	M	SE	95% CI	M	SE	95% CI
Human source	One report	40% accuracy	No disclosure	23	2.40	0.28	1.85, 2.94	1.86	0.25	1.37, 2.35	2.43	0.27	1.90, 2.96	4.22	0.21	3.80, 4.63
			Disclosure	25	2.18	0.26	1.66, 2.69	1.31	0.24	0.86, 1.65	2.00	0.26	1.49, 2.50	4.62	0.20	4.23, 5.01
		80% accuracy	No disclosure	33	3.35	0.23	2.90, 3.80	1.24	0.21	.084, 1.65	2.94	0.22	2.50, 3.38	3.89	0.17	3.55, 4.23
			Disclosure	23	3.20	0.27	2.67, 3.74	0.92	0.25	0.44, 1.40	2.81	0.27	2.28, 3.33	3.96	0.21	3.56, 4.37
	Five reports	40% accuracy	No disclosure	28	2.90	0.25	2.41, 3.39	1.52	0.23	1.07, 1.96	2.70	0.24	2.22, 3.18	4.61	0.19	4.24, 4.99
			Disclosure	26	2.63	0.26	2.12, 3.13	1.54	0.23	1.08, 2.00	2.67	0.25	2.17, 3.17	3.85	0.20	3.46, 4.23
		80% accuracy	No disclosure	30	3.35	0.24	2.88, 3.83	1.66	0.22	1.23, 2.08	3.25	0.24	2.79, 3.72	4.18	0.18	3.82, 4.54
			Disclosure	25	3.39	0.26	2.87, 3.91	1.41	0.24	.94, 1.88	3.04	0.26	2.53, 3.55	4.13	0.20	3.73, 4.52
Automated source	One report	40% accuracy	No disclosure	25	2.81	0.26	2.30, 3.33	1.73	0.24	1.26, 2.19	2.75	0.26	2.25, 3.25	4.31	0.20	3.92, 4.70
			Disclosure	31	2.65	0.24	2.19, 3.11	1.55	0.21	1.13, 1.96	2.66	0.23	2.21, 3.11	4.43	0.18	4.08, 4.78
		80% accuracy	No disclosure	32	2.40	0.23	1.94, 2.85	1.08	0.21	0.67, 1.49	2.47	0.23	2.03, 2.92	4.16	0.18	3.71, 4.50
			Disclosure	22	3.19	0.28	2.64, 3.74	1.31	0.25	0.81, 1.80	3.17	0.27	2.64, 3.71	3.76	0.21	3.34, 4.17
	Five reports	40% accuracy	No disclosure	24	2.51	0.27	1.98, 3.03	1.62	0.24	1.15, 2.10	2.60	0.26	2.09, 3.12	4.20	0.20	3.80, 4.60
			Disclosure	24	2.76	0.27	2.24, 3.29	1.26	0.24	0.79, 1.73	2.08	0.26	1.57, 2.60	4.23	0.20	3.83, 4.62
		80% accuracy	No disclosure	33	2.72	0.23	2.27, 3.17	1.21	0.21	0.80, 1.61	2.53	0.25	2.09, 2.97	4.16	0.17	3.82, 4.50
			Disclosure	34	3.08	0.23	2.64, 3.53	1.38	0.20	0.98, 1.78	2.89	0.22	2.45, 3.32	4.20	0.17	3.86, 4.53
Total				438	2.85	0.06	2.72, 2.97	1.41	0.06	1.30, 1.52	2.69	0.06	2.57, 2.81	4.18	0.05	4.09, 4.27

TABLE 2 Results of analyses of covariance predicting study outcomes

	Hit rate		False-positive rate		Adherence to warning recommendations		Anxiety with system use	
Variable	*F*	η_p^2	*F*	η_p^2	*F*	η_p^2	*F*	η_p^2
Disposition to trust	1.21	0.00	0.77	0.00	1.10	0.00	2.20	0.01
Mindfulness in technology	3.58†	0.01	0.24	0.00	0.21	0.00	0.10	0.00
Perceived internet risk	1.02	0.00	2.50	0.01	0.28	0.00	11.30**	0.00
Computer self-efficacy—Internal	0.10	0.01	0.42	0.00	1.11	0.00	0.00	0.00
Computer self-efficacy—External	4.89*	0.00	0.00	0.00	0.95	0.00	0.04	0.00
Phishing identification expertise	8.10**	0.02	1.29	0.00	0.00	0.00	5.82*	0.01
Email experience	5.84*	0.01	8.69**	0.02	5.01*	0.01	3.45	0.01
Computer game interest	1.96	0.01	0.03	0.00	1.21	0.00	5.32	0.01
Report source	1.61	0.00	0.13	0.00	0.46	0.00	0.00	0.00
Number of reports	1.32	0.00	0.42	0.00	0.30	0.00	0.07	0.00
Accuracy rate	14.74***	0.03	5.72*	0.01	10.65**	0.03	7.07**	0.02
Disclosure of accuracy rate	0.41	0.00	1.80	0.00	0.14	0.00	0.52	0.00
Source × number	1.23	0.00	1.18	0.00	6.10*	0.02	0.00	0.00
Source × accuracy	6.34*	0.02	0.05	0.00	1.66	0.00	0.10	0.00
Source × disclosure	3.37†	0.01	1.05	0.00	1.60	0.00	0.03	0.00
Number × accuracy	0.12	0.00	3.01†	0.01	0.01	0.00	4.17*	0.0
Number × disclosure	0.02	0.00	0.20	0.00	0.21	0.00	1.54	0.00
Accuracy × disclosure	1.98	0.01	0.94	0.00	3.09†	0.01	0.02	0.00
Source × number × accuracy	1.33	0.00	0.22	0.00	0.81	0.00	0.01	0.00
Source × number × disclosure	0.02	0.00	0.94	0.00	1.22	0.00	4.53*	0.01
Source × accuracy × disclosure	0.45	0.00	1.15	0.00	2.38	0.01	1.32	0.00
Number × accuracy × disclosure	0.36	0.00	0.16	0.00	0.17	0.00	4.21*	0.01
Source × number × accuracy × disclosure	1.17	0.00	0.45	0.00	0.33	0.00	0.46	0.00
R^2	0.12		0.09		0.08		0.12	

$^{\dagger}P < .10$;
$^{*}P < .05$;
$^{**}P < .01$;
$^{***}P < .001$.

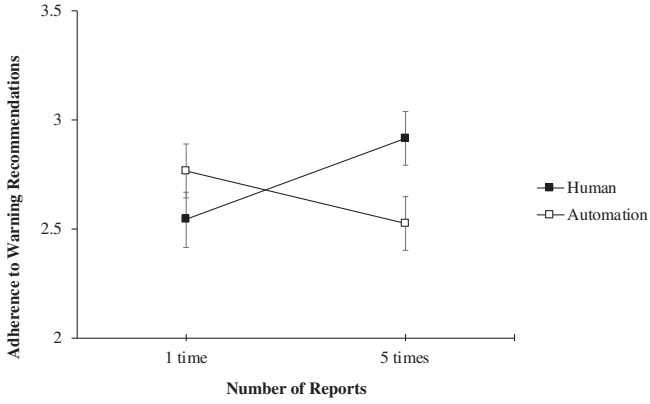


FIGURE 1 Interaction effect between number of reports and source of reports on adherence to warning recommendations

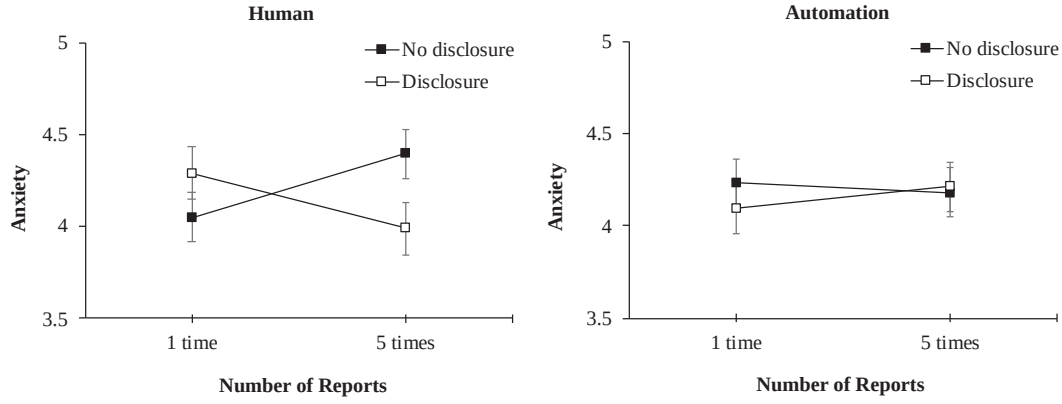


FIGURE 2 Comparison of interaction effects between disclosure of accuracy rates and number of reports by source of reports on anxiety with system use

with system use was significantly lower when the accuracy rate was disclosed than when the accuracy rate was not disclosed. Thus, Hypothesis 4 was supported.

As seen in Table 2, although were no significant interactions between accuracy rate of reports and disclosure of the accuracy rate on hit rates, $F(1,414) = 1.98$, $P = .16$, or false-positive rates, $F(1,414) = 0.94$, $P = .33$, there was a marginally significant interaction between accuracy rate of reports and disclosure of the accuracy rate on adherence to warning recommendations, $F(1,414) = 3.09$, $P = .08$. Additionally, there was a significant three-way interaction between accuracy rate of reports, number of reports and disclosure of the accuracy rate on anxiety with system use, $F(1,414) = 4.21$, $P = .04$. As shown in Figure 3, there was a significant interaction between accuracy rate of reports and number of reports when the accuracy rate was disclosed but not when the accuracy rate was not disclosed. For disclosed accuracy rates, there was a significant difference in anxiety with system use when the number of reports was lower (ie, one time) compared to higher (ie, five times). Specifically, when the number of reports was lower, anxiety with system use was significantly lower when the accuracy rate of reports was 80% instead of 40%. Thus, Hypothesis 5 was partially supported.

As seen in Table 2, although there were no significant interactions between source of reports and accuracy of reports on false-positive rates, $F(1,414) = 0.05$, $P = .83$, acceptance of recommendations, $F(1,414) = 1.66$, $P = .20$ or anxiety with system use, $F(1,414) = 0.10$, $P = .75$, there was a significant interaction between the source of reports and accuracy rate of reports on hit rates, $F(1,414) = 6.34$, $P = .01$. As shown in Figure 4, there was a significant

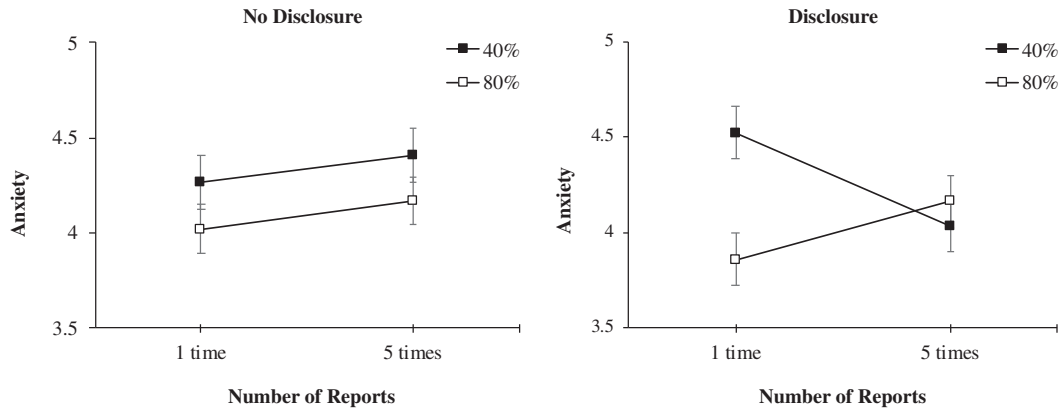
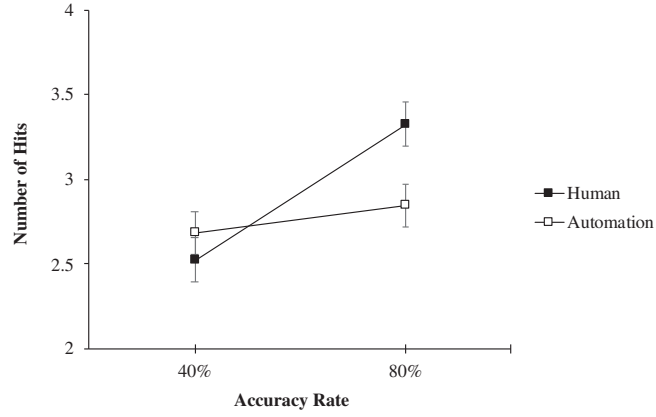


FIGURE 3 Comparison of interaction effects between accuracy rate of reports and number of reports by disclosure of accuracy rates on anxiety of with system use

FIGURE 4 Interaction effect between accuracy rate of reports and source of reports on number of hits



difference when the accuracy of reports was higher (ie, 80%) but not when the accuracy of reports was lower (ie, 40%). Specifically, when the accuracy of reports was higher, hit rates were significantly higher when the source of reports was a human instead of automated. Thus, Hypothesis 6 was partially supported.

As seen in Table 2, although were no significant interactions between number of reports and accuracy rate of reports on hit rates, $F(1,414) = 0.12$, $P = .73$, or adherence to warning recommendations, $F(1,414) = 0.01$, $P = .93$, there was a marginally significant interaction between number of reports and accuracy rate of reports on false-positive rates, $F(1,414) = 3.01$, $P = .08$. Additionally, there was a significant interaction between the number of reports and accuracy rate of reports on anxiety with system use, $F(1,414) = 4.17$, $P = .04$. As shown in Figure 5, there was a significant difference when the number of reports was lower (ie, one time) but not when the number of reports was higher (ie, five times). Specifically, when the number of reports was lower, anxiety with system use was significantly lower when the accuracy rate of reports was 80% instead of 40%. Thus, Hypothesis 7 was partially supported.

6 | DISCUSSION

Table 3 provides a summary of the results found for the study hypotheses, and Table 4 outlines the study implications as it relates to theoretical and practical contributions. The purpose of our study was to examine the features of

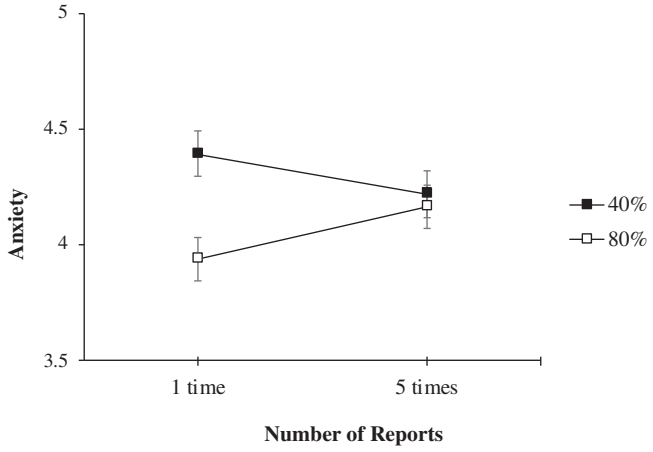


FIGURE 5 Interaction effect between accuracy rate of reports and number of reports on anxiety with system use

reports from crowdsourced phishing warning systems that influence individuals' utilisation of system recommendations and evaluation of email messages. Due to the lack of current empirical research on the design of these crowdsourced phishing warning systems, we built upon two literatures to explore this issue: human-automation interaction and crowdsourcing. The results showed that features related to human trust in automation and features in crowdsourced systems features work together to encourage warning acceptance and reduce anxiety, with accuracy rate demonstrating the most prominent effects on outcomes related to judgement accuracy, adherence to warning recommendations and anxiety with system use. The results offer several contributions to theory and practice, which are discussed below.

6.1 | Theoretical implications

Our findings echo Madhavan and Phillips's (2010) notion that 'the more reliable the system, the better utilised it will be' (p. 204). Specifically, individuals had significantly higher message judgement accuracy (eg, more hits and fewer false-positives), adhered to significantly more warning recommendations and reported significantly less anxiety using the system when accuracy rates were higher. These findings are consistent with the human-automation interaction body of research that suggests that automated systems with improved accuracy are important for eliciting more trust and ultimately better system acceptance and utilisation (Lee & See, 2004; Parasuraman et al., 2000). When provided with more accurate warning recommendations, individuals can make more optimal decisions and properly utilise recommendations. In contrast, the report source, number of previous reports and disclosure of accuracy rates did not demonstrate main effects on any of the outcome variables. In other words, when adhering to warning recommendations, managing anxiety with using the system and making message judgements, whether reports came from human or automated sources, how high the number of contributed reports was or if individuals were aware of how accurate the reports were ultimately mattered less than system accuracy.

Our findings are also consistent with Thuan et al.'s (2013) notion that the quality of the contributions to crowdsourced systems can vary greatly and can ultimately impact decision-making. If the quality of contributions influences how information from the crowdsourced system is used, users need to be provided with this information, rather than just the sheer quantity of contributions or who the contributors are. In other words, the features unique to crowdsourced systems or human-automation interactions may need to be examined together. Our findings support this notion through the interaction effects found for our various outcome variables, indicating how these features do not exist independently but instead must work in conjunction with

TABLE 3 Summary of study hypotheses and associated support found

Hypothesis	Description	Type of test	Results
1	A higher number of reports (either human or automated) in phishing warning systems will be associated with (a) adherence to more warning recommendations and (b) less anxiety using the system	Main effect of number of reports on: (a) adherence to warning recommendations (b) anxiety with system use	Hypothesis 1a and 1b not supported
2	Reports from human sources in phishing warning systems will be associated with (a) adherence to more warning recommendations and (b) less anxiety using the system compared to reports which come from automated sources	Main effect of report source on: (a) adherence to warning recommendations (b) anxiety with system use	Hypothesis 2a and 2b not supported
3	Higher accuracy rates of reports (either human or automated) in phishing warning systems will be associated with (a) adherence to more warning recommendations and (b) less anxiety using the system	Main effect of accuracy rate on: (a) adherence to warning recommendations (b) anxiety with system use	Hypothesis 3a and 3b supported: Adherence to warning recommendations was significantly greater for 80% accuracy rate vs 40% Anxiety with system use was significantly lower for 80% vs 40%
4	The number of reports and the source of reports will interact such that when the number of reports is higher and the reports come from human sources, individuals will have (a) adherence to more warning recommendations and (b) less anxiety using the system	2 (number of reports: one or five times) × 2 (report source: human or automated) ANCOVA on: (a) adherence to warning recommendations (b) anxiety with system use	Hypothesis 4a and 4b supported: There was a significant difference in adherence to warning recommendations when the number of reports was higher (ie, five times) but not when the number of reports was lower (ie, one time). When the number of reports was higher (ie, five times), adherence to warning recommendations was significantly greater when the report source was human instead of automated There was a significant interaction between disclosure of accuracy rates and number of reports on anxiety with system use when the source of the reports was a human (vs not automated). There was a significant difference in anxiety only when the number of reports was higher (ie, reported five times). When the number of reports was higher, anxiety was significantly lower when the accuracy rate was disclosed than when the accuracy rate was not disclosed

(Continues)

TABLE 3 (Continued)

Hypothesis	Description	Type of test	Results
5	The accuracy rate of reports and the disclosure of accuracy rates will interact such that when the accuracy rate of reports is higher and the accuracy rate is disclosed, individuals will have (a) higher hit rates; (b) lower false-positive rates; (c) adherence to more warning recommendations and (d) less anxiety using the system	2 (accuracy rate: 40% or 80%) × 2 (disclosure of accuracy rate: disclosed or not disclosed) ANCOVA on: (a) hit rates (b) false-positive rates (c) adherence to warning recommendations (d) anxiety with system use	Hypothesis 5d supported: There was a significant interaction between accuracy rate and number of reports on anxiety with system use when the accuracy rate of the reports was disclosed (vs not disclosed). There was a significant difference in anxiety only when the number of reports was lower (ie, reported one time). When the number of reports was lower, anxiety was significantly lower when the accuracy rate was 80% than when the accuracy rate was 40% Hypothesis 5c marginally supported, and Hypotheses 5a and 5b not supported
6	The source of reports and the accuracy rate of reports will interact such that when the reports come from human sources and the accuracy rate of reports is higher, individuals will have (a) higher hit rates; (b) have lower false-positive rates; (c) adherence to more warning recommendations and (d) less anxiety using the system	2 (report source: human or automated) × (accuracy rate: 40% or 80%) ANCOVA on: (a) hit rates (b) false-positive rates (c) adherence to warning recommendations (d) anxiety with system use	Hypothesis 6a supported: When accuracy rate was lower (ie, 40%), there was no significant difference in hit rates when report source was human or automated. When accuracy rate was higher (ie, 80%), hit rates were significantly higher when report source was human than automated Hypotheses 6b, 6c and 6d not supported
7	The number of reports and the accuracy rate of reports will interact such that when the number of reports is higher and the accuracy of reports is higher, individuals will have (a) higher hit rates; (b) lower false-positive rates; (c) adherence to more warning recommendations and (d) less anxiety using the system	2 (number of reports: one or five times) × (accuracy rate: 40% or 80%) ANCOVA on: (a) hit rates (b) false-positive rates (c) adherence to warning recommendations (d) anxiety with system use	Hypothesis 7d supported: When the number of reports was higher (ie, five times), there was no significant difference in anxiety between the 40% and 80% accuracy rates. When the number of reports was lower (ie, one time), anxiety was significantly lower when the accuracy rate of reports was 80% instead of 40% Hypothesis 7b marginally supported, and Hypotheses 7a and 7c not supported

TABLE 4 Summary of study implications

Supported hypothesis	Description	Contributions to theory	Contributions to practice
3	Higher accuracy rates of reports (either human or automated) in phishing warning systems will be associated with adherence to more warning recommendations and less anxiety using the system	Corroborates human-automation findings: The more accurate the system, the better utilised it will be	Organisations will benefit from adding crowdsourced phishing systems to other anti-phishing efforts as part of a layered approach Building an effective crowdsourced system for phishing detection is a dynamic process, requiring a balance of various system design characteristics based on the information currently contributed to the system and available to users
4	The number of reports and the source of reports will interact such that when the number of reports is higher and the reports come from human sources, individuals will have adherence to more warning recommendations and less anxiety using the system	Corroborates crowdsourced systems findings: Features of crowdsourced systems must work in conjunction with each other to improve adoption and utilisation People will utilise warnings from other people but only if there are enough reports; however, when there are a large number of reports from others, the accuracy rate should be disclosed to people so they can decide how to utilise the recommendations	
5	The accuracy rate of reports and the disclosure of accuracy rates will interact such that when the accuracy rate of reports is higher and the accuracy rate is disclosed, individuals will have adherence to more warning recommendations and less anxiety using the system	When there is a small number of reports contributed to a crowdsourced system, anxiety with using the system can be reduced by informing people those reports are highly accurate	Recommendations to users about suspicious messages must be accurate, particularly if there are only a few recommendations contributed In crowdsourced phishing systems, accuracy rests in the hands of the contributors, reinforcing the importance of anti-phishing training
6 and 7	The source of reports and the accuracy rate of reports will interact such that when the reports come from human sources and the accuracy rate of reports is higher, individuals will have higher hit rates The number of reports and the accuracy rate of reports will interact such that when the number of reports is higher and the accuracy of reports is higher, individuals will have lower false-positive rates and less anxiety using the system	Accuracy of crowdsourced recommendations primarily drive phishing detection hit rates, particularly if recommendations come from other people When there is a small number of reports contributed to a crowdsourced system, anxiety with using the system can be reduced by ensuring those reports are highly accurate Characteristics of crowdsourced phishing systems behave differently than crowdsourced e-commerce systems, with a greater emphasis on accuracy vs the sheer number of contributions	

one another. Overall, our study attempted to bridge the lack of integration within both streams of literature in human-automation interaction and crowdsourcing that can help provide clarity on designing crowdsourced systems that best assist individuals in their decision-making and use of recommendations, particularly within the context of phishing detection.

In terms of adhering to warning recommendations, the sheer number of contributed reports may have differing effects depending on the report source. As seen in Figure 1, the results showed that individuals are more likely to accept recommendations from other individuals (ie, human sources) when there are a higher number of reports to the system, with the opposite being true for reports from automated sources. This finding supports previous findings that demonstrate the effectiveness of user recommendations compared to system recommendations (Ashraf et al., 2019; Benlian et al., 2012; Lin, 2014). Additionally, factors such as shared experiences and perceived similarities may also be attributed to greater trustworthiness and credibility when adhering to recommendations. In particular, participants may have identified with the human sources more so because they were also university students (Benlian et al., 2012; Hurley, 2006). Even so, these findings provide greater insight regarding how individuals may associate the credibility of recommendations from other people by the sheer number of people who have previously contributed to the system. In other words, establishing trust in crowdsourced recommendations may require input from multiple others before credibility is established, whereas establishing trust in automated sources may not.

Similarly, as seen in Figure 4, individuals had greater message judgement accuracy (ie, higher hit rates) when warning recommendations were reported from other individuals. These reports, however, must be highly accurate to help individuals' phishing detection, again highlighting the importance of having systems with accurate performance (Lee & See, 2004; Parasuraman et al., 2000). This finding is consistent with Pearson et al.'s (2019) notion that the relationship and degree of trust expended towards other individuals are dependent upon others' perceived expertise. Although it is surprising that the number of correctly identified email messages did not significantly improve when the accuracy rate increased for reports from automated sources, these results are consistent with individuals' preferences for recommendations from other individuals compared to recommendations from an automated system (Ashraf et al., 2019; Benlian et al., 2012; Lin, 2014). It is possible that individuals were overly confident in their ability to correctly judge emails and did not rely upon the reports from the automated source to make their judgements; rather, they relied more upon reports from other individuals by projecting their own confidence to human sources that thus improved their phishing detection.

The results also showed that anxiety with system use is impacted differently depending on various combinations of system features. For example, as seen in Figure 2, having a large number of reports from human sources created more anxiety when the accuracy of the reports was not disclosed. Although researchers (eg, Hoff & Bashir, 2015; Lee & See, 2004; Wang et al., 2009) discuss the importance of transparency of system performance in generating greater trust, the decision to disclose accuracy rates may ultimately depend on the how accurate the systems' recommendations are. As seen in Figures 3 and 5, disclosure actually increased anxiety when reports were low in accuracy and low in number, whereas disclosure reduced anxiety when accuracy rates were higher, even if the number of reports was low. This pattern of results further illustrates the importance of having quality contributions (ie, highly accurate), especially when there are a low number of contributions.

In this vein, although the present study draws upon e-commerce research to investigate similar crowdsourcing features, it is important to acknowledge how the present study's phishing context influences the interpretation of results. Whereas individuals purposefully browse and purchase new products in an e-commerce setting, phishing detection is a task inherent in daily email management that serves the sole purpose of avoiding negative outcomes. Because there can be disastrous consequences associated with incorrect decision-making in phishing detection (in contrast to affirmative search tasks in e-commerce), certain features for the crowdsourced system, although relevant in both contexts, may have a greater influence on individuals'

decision-making. For example, the effectiveness and utilisation of recommendations within an e-commerce context places precedence on a large number of contributions to influence consumer behaviours (eg, Kim & Gambino, 2016; Lin, 2014). However, our results also place precedence on accuracy in crowdsourced phishing systems (perhaps even to exceed number of contributions). E-commerce recommendations and product reviews are typically subjective evaluations and often do not have values of accuracy attached to them. There is, however, a definite correct/incorrect answer in email evaluation when determining whether an email message is phishing or not. This notion is in line with previous researchers' (eg, Abbasi & Chen, 2009; Wu et al., 2006) findings on the importance that detection rate accuracy plays on influencing individuals to utilise system recommendations.

Phishing literature (see Appendix A) has made significant strides in establishing greater understanding of the individual characteristics that increase phishing susceptibility, effective phishing training methods and influence tactics embedded within phishing emails. There has been, instead, less attention dedicated to collaborative methods and strategies for individuals to more effectively combat phishing. Liu et al.'s (2011) initial findings on the effectiveness of presenting or clustering large numbers of phishing messages together (vs individually) to help individuals more accurately and quickly identify messages demonstrate a promising area for using crowdsourcing for phishing detection. The decision to incorporate certain features over others in a crowdsourced system, however, must be done with particular care in contexts such as phishing detection in which negative consequences can be mitigated. As such, examining additional features that make an effective crowdsourcing system in a phishing context provides a novel perspective on utilising crowdsourcing for information security purposes such that it serves an organisational good and can have lasting impacts beyond the benefits for just one individual.

Overall, the findings provide greater insight regarding how individuals utilise information from human and automated sources within a recommendation system, as well as the nuances associated with the trust they place in these sources when they are paired with other system features (ie, number of reports, accuracy rate of reports, disclosure of accuracy rate). Although these findings are interesting and contribute to our understanding in developing crowdsourced systems that best integrate both features related to human trust in automation and features unique to crowdsourcing, these interaction effects are much weaker compared to the overall main effect of accuracy rate (see Table 2 for effect sizes). In other words, the results consistently show that crowdsourced systems can be an effective decision aid for phishing detection but must be designed to provide accurate information.

6.2 | Practical implications

Our findings indicate that organisations may benefit from integrating crowdsourcing to combat phishing attacks to keep their assets and employees secure. Although anti-phishing tools such as automated blacklists or email filters are effective, they are not foolproof and must constantly be updated to remain applicable. Instead, utilising the information harnessed from crowdsourcing can help employees more effectively identify phishing messages. Individual users are often considered the weakest link when it comes to security, so rather than treating phishing attacks as isolated events, organisations should emphasise the need to work collectively to overcome these challenges.

It is important to note, however, that even with access to information from crowdsourced phishing warning systems, the decision-making is still in the hands of individuals. Systems can organise and integrate information contributed from the crowd and offer recommendations, but the individual ultimately has the option to ignore or utilise recommendations when deciding how to act when processing emails. Thus, organisations need to consider features of the contributions made by their crowdsourced systems during the system design and implementation process that will lead to more frequent and appropriate recommendation usage. As the

nature of the contributions changes over time, whether it is who is making the contributions or how many contributions there are, user interaction with the crowdsourced systems will also shift, requiring a design process that is dynamic and adaptable to most effectively optimise system utilisation. Even so, our findings indicate that rather than focussing on crowdsourcing large numbers of reports that come from certain sources, efforts should instead be focussed on making sure that any warnings that are recommended to individuals are accurate. Therefore, actions that organisations can take to improve the accuracy of reports will likely yield the greatest gains in acceptance and overall protection. Past research has shown that anti-phishing training, public recognition of contribution and verification all play a role in increasing the accuracy of phishing message reporting (Jensen, Dinger, et al., 2017; Jensen, Durcikova, & Wright, 2017; Karumbaiah et al., 2016). By incorporating these various methods, individuals can use the information provided from crowdsourced systems to supplement their own decisions when detecting phishing messages, rather than solely relying upon these systems to make the decisions for them.

It is important to note that with any study examining security-related issues, there is the potential risk of dual use of research results for benevolent vs malevolent intentions (Rath et al., 2014). Specifically, within the context of phishing, research on reducing individuals' phishing susceptibility can aid organisations on how to better protect their employees and themselves as a whole from these attacks; however, these research findings can also potentially aid hackers in identifying key areas in which individuals are most susceptible to phishing and exploiting those weaknesses in their attack tactics. As we gain a greater understanding of what phishing messages typically look like, what common strategies phishers are using and what individual differences increase the likelihood of being phished, we must also be aware that phishers are also becoming increasingly advanced and less predictable in their methods. Even so, the risk of dual use of our study's findings is low due to the focus on understanding and developing a new method of combatting phishing through crowdsourcing. In other words, rather than directly examining what makes phishing attacks more effective, our study examines how to better leverage and overcome individuals' weaknesses of working in isolation when detecting phishing messages.

6.3 | Limitations and future research

There are several limitations of the present study that should be noted. First, the correct or incorrect identification of email messages in the study's task was not linked to consequences that are associated with good or bad decisions in real-world situations. In other words, users' decisions to accept or utilise warning recommendations could differ if there were actual consequences in falling for a phishing attack. Future research should include a field study to examine whether the influence of the features of reports from crowdsourced systems have similar effects on individuals' everyday email usage. Additionally, our study was conducted in an academic setting with college students. College students are regular targets of phishing attacks, but because the goal of the research is to ultimately offer recommendations to organisations, future research should examine the generalisability of the findings in different organisational settings and organisational samples. In this vein, the generalisability of these findings is also limited by the levels chosen for the study manipulations. Although previous research and rationale for mimicking real-world phishing scenarios guided these decisions, future researchers should examine additional levels of the study manipulations, particularly the number of reports and accuracy rate, which do not fall into distinct binary categories like report source and accuracy rate disclosure. Finally, crowdsourcing approaches may be limited in their effectiveness when it comes to combatting more targeted phishing attacks (eg, spear phishing). The intent of the present study was to examine how crowdsourcing first applies to typical phishing messages that contain common cues before examining these more nuanced, sophisticated phishing attacks. Future research should examine how crowdsourced systems can be effectively implemented to combat attacks that specifically target individuals and are much more challenging to detect and assess.

7 | CONCLUSION

In summary, the results of the present study highlight the importance of providing accurate recommendations by crowdsourced systems. Compared to effects from other features such as report sources, number of reports or the disclosure of accuracy rates, it is the accuracy rate of the warnings that matters most in terms of facilitating effective information usage from a crowdsourced system when evaluating emails. Specifically, when individuals are provided with more accurate information, they are more accurate in their identification of phishing emails, more likely to adhere to warning recommendations and less anxious using the system.

ACKNOWLEDGEMENTS

This research was supported by a grant from the NSF Social and Economic Division, Project# (Redacted for blind review). The views and conclusions contained herein are those of the authors and should not be interpreted as representing the official policies or endorsements, either expressed or implied, of NSF or the U.S. Government.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Matthew L. Jensen https://orcid.org/0000-0001-8711-1827

Ryan T. Wright https://orcid.org/0000-0002-9719-415X

REFERENCES

Abawajy, J. (2014). User preference of cyber security awareness delivery methods. *Behaviour & Information Technology*, *33* (3), 237–248.

Abbasi, A., & Chen, H. (2009). A comparison of tools for detecting fake websites. *IEEE Computer*, *42*(10), 78–86.

Abbasi, A., Dobolyi, D., Vance, A., & Zahedi, M. (2021). The phishing funnel model: A design artifact to predict user susceptibility to phishing websites. *Information Systems Research*, (forthcoming), 1–13.

Abbasi, A., Zahedi, F. M., Zeng, D., Chen, Y., Chen, H., & Nunamaker, J. F., Jr. (2015). Enhancing predictive analytics for anti-phishing by exploiting website genre information. *Journal of Management Information Systems*, *31*(4), 109–157.

Abbasi, A., Zhang, Z., Zimbra, D., Chen, H., & Nunamaker, J. F., Jr. (2010). Detecting fake websites: The contribution of statistical learning theory. *MIS Quarterly*, *34*(3), 435–461.

Akdemir, N., & Lawless, C. J. (2020). Exploring the human factor in cyber-enabled and cyber-dependent crime victimisation: A lifestyle routine activities approach. *Internet Research*, *30*(6), 1–23.

Alseadoon, I., Chan, T., Foo, E., & Gonzales, N. J. (2012). Who is more susceptible to phishing emails? A Saudi Arabian study. In *ACIS 2012: Location, location, location: Proceedings of the 23rd Australasian conference on information systems 2012* (pp. 1–11). ACIS.

Alsharnouby, M., Alaca, F., & Chiasson, S. (2016). Why phishing still works: User strategies for combating phishing attacks. *International Journal of Human–Computer Studies*, *89*, 69–82.

Anderson, B., Bjornn, D., Jenkins, J., Kirwan, B., & Vance, A. (2018). Improving security message adherence through improved comprehension: Neural and behavioral insights. In *Proceedings 2018 Americas conference on information systems (AMCIS)* (pp. 1–5). AMCIS.

Anti-Phishing Working Group (2020). Phishing activity trends report. Retrieved from https://apwg.org/trendsreports/

Arachchilage, N. A. G., & Love, S. (2013). A game design framework for avoiding phishing attacks. *Computers in Human Behavior*, *29*(3), 706–714.

Arachchilage, N. A. G., Love, S., & Beznosov, K. (2014). Security awareness of computer users: A phishing threat avoidance perspective. *Computers in Human Behavior*, *38*, 304–312.

Ashraf, M., Jaafar, N. I., & Sulaiman, A. (2019). System- vs. consumer-generated recommendations: Affective and social-psychological effects on purchase intention. *Behaviour & Information Technology*, *38*(12), 1259–1272.

Ayaburi, E., & Andoh-Baidoo, F. K. (2019). Understanding phishing susceptibility: An integrated model of cue-utilization and habits. In *Proceedings of the 2019 international conference on information systems*. ICIS.

Bansal, G. (2018). Got phished! Role of top management support in creating phishing safe organizations. In *Proceedings of the 2018 Midwest Association for information systems*. MWAIS.

Benenson, Z., Gassmann, F., & Landwirth, R. (2017). Unpacking spear phishing susceptibility. In *International conference on financial cryptography and data security* (pp. 610–627). Springer.

Benlian, A., Titah, R., & Hess, T. (2012). Differential effects of provider recommendations and consumer reviews in e-commerce transactions: An experimental study. *Journal of Management Information Systems*, *29*(1), 237–272.

Blythe, M., Petrie, H., & Clark, J. A. (2011). F for fake: Four studies on how we fall for phish. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 3469–3478). ACM.

Bose, I., & Leung, A. C. M. (2008). Assessing anti-phishing preparedness: A study of online banks in Hong Kong. *Decision Support Systems*, *45*(4), 897–912.

Canfield, C. I., Fischhoff, B., & Davis, A. (2016). Quantifying phishing susceptibility for detection and behavior decisions. *Human Factors*, *58*(8), 1158–1172.

Caputo, D. D., Pfleeger, S. L., Freeman, J. D., & Johnson, M. E. (2013). Going spear phishing: Exploring embedded training and awareness. *IEEE Security & Privacy*, *12*(1), 28–38.

Carpenter, S., Zhu, F., & Kolimi, S. (2014). Reducing online identity disclosure using warnings. *Applied Ergonomics*, *45*(5), 1337–1342.

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

Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, *19*(2), 189–211.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340.

de Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International Journal of Human-Computer Studies*, *58*(6), 719–735.

Dhamija, R., Tygar, J. D., & Hearst, M. (2006). Why phishing works. In *Computer human interaction conference* (pp. 581–590). ACM.

Diaz, A., Sherman, A. T., & Joshi, A. (2020). Phishing in an academic community: A study of user susceptibility and behavior. *Cryptologia*, *44*(1), 53–67.

Dobolyi, D. G., & Abbasi, A. (2016). PhishMonger: A free and open source public archive of real-world phishing websites. In *2016 IEEE conference on intelligence and security informatics (ISI)* (pp. 31–36). IEEE.

Dodge, R. C., Jr., Carver, C., & Ferguson, A. J. (2007). Phishing for user security awareness. *Computers and Security*, *26*(1), 73–80.

Dodge, R., Coronges, K., & Rovira, E. (2012). Empirical benefits of training to phishing susceptibility. In *International information security conference* (pp. 457–464). Springer.

Downs, J. S., Holbrook, M. B., & Cranor, L. F. (2006). Decision strategies and susceptibility to phishing. In *Proceedings of the second symposium on usable privacy and security* (pp. 79–90). ACM.

Durward, D., Blohm, I., & Leimeister, J. M. (2020). The nature of crowd work and its effects on individuals' work perception. *Journal of Management Information Systems*, *37*(1), 66–95.

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

Eminağaoğlu, M., Uçar, E., & Eren, Ş. (2009). The positive outcomes of information security awareness training in companies: A case Study. *Information Security Technical Report*, *14*(4), 223–229.

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors*, *59*(1), 5–27.

Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, *38*(2), 189–200.

Felt, A. P., Ainslie, A., Reeder, R. W., Consolvo, S., Thyagaraja, S., Bettes, A., Harris, H., & Grimes, J. (2015). Improving SSL warnings: Comprehension and adherence. In *Proceedings of the conference on human factors in computing systems* (pp. 2893–2902). ACM.

Finnerty, A., Kucherbaev, P., Tranquillini, S., & Convertino, G. (2013). Keep it simple: Reward and task design in crowdsourcing. In *Proceedings of the biannual conference of the Italian chapter of SIGCHI* (pp. 1–4). ACM.

Gavett, B. E., Zhao, R., John, S. E., Bussell, C. A., Roberts, J. R., & Yue, C. (2017). Phishing suspiciousness in older and younger adults: The role of executive functioning. *PloS One*, *12*(2), 1–16.

Geiger, D., Rosemann, M., Fielt, E., & Schader, M. (2012). Crowdsourcing information systems – Definition, typology, and design. In *Proceedings of the 33rd international conference on information systems* (pp. 1–5). ICIS.

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

Gupta, B. B., Tewari, A., Jain, A. K., & Agrawal, D. P. (2017). Fighting against phishing attacks: State of the art and future challenges. *Neural Computing and Applications*, *28*(12), 3629–3654.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human–robot interaction. *Human Factors*, *53*(5), 517–527.

Harrison, A., Samuel, B., Shan, Z., Cook, M., Zu, T., & Dawani, D. (2019). Learning to see the hook: Comparing phishing training approaches. In *International conference on information systems*. AIS.

Harrison, B., Vishwanath, A., & Rao, R. (2016). A user-centered approach to phishing susceptibility: The role of a suspicious personality in protecting against phishing. In *2016 49th Hawaii international conference on system sciences (HICSS)* (pp. 5628–5634). IEEE.

Hetmank, L. (2013). Components and functions of crowdsourcing systems: A systematic literature review. *Wirtschaftsinformatik Proceedings*, *4*, 55–69.

Hiltz, S. R., & Turoff, M. (1985). Structuring computer-mediated communication systems to avoid information overload. *Communications of the ACM*, *28*(7), 680–689.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434.

Hong, J. (2012). The state of phishing attacks. *Communications of the ACM*, *55*(1), 74–81.

Hong, K. W., Kelley, C. M., Tembe, R., Murphy-Hill, E., & Mayhorn, C. B. (2013). Keeping up with the Joneses: Assessing phishing susceptibility in an email task. In *Proceedings of the human factors and ergonomics society annual meeting* (pp. 1012–1016). Sage.

Hossain, M., & Kauranen, I. (2015). Crowdsourcing: A comprehensive literature review. *Strategic Outsourcing: An International Journal*, *8*(1), 2–22.

Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine*, *14*(6), 1–4.

Hurley, R. F. (2006). The decision to trust. *Harvard Business Review*, *84*(9), 55–62.

Hussein, A., Elsawah, S., & Abbass, H. (2020). Towards trust-aware human–automation interaction: An overview of the potential of computational trust models. In *Proceedings of the 53rd Hawaii international conference on system sciences* (pp. 375–384). HICSS.

Iuga, C., Nurse, J. R., & Erola, A. (2016). Baiting the hook: Factors impacting susceptibility to phishing attacks. *Human-Centric Computing and Information Sciences*, *6*(1), 3–20.

Jagatic, T., Johnson, N., & Jackobsson, F. (2007). Social phishing. *Communications of the ACM*, *50*(10), 94–100.

Jansson, K., & von Solms, R. (2013). Phishing for phishing awareness. *Behaviour & Information Technology*, *32*(6), 584–593.

Jensen, M. L., Dinger, M., Wright, R., & Thatcher, J. (2013). Training to mitigate threats from customized phishing attacks. *Credibility Assessment and Information Quality in Government and Business*. IEEE Computer Society.

Jensen, M. L., Dinger, M., Wright, R. T., & Thatcher, J. B. (2017). Training to mitigate phishing attacks using mindfulness techniques. *Journal of Management Information Systems*, *34*(2), 597–626.

Jensen, M. L., Durcikova, A., & Wright, R. T. (2017). Combating phishing attacks: A knowledge management approach. In *Proceedings of the 50th annual Hawaii international conference on system science*. HICSS.

Jiao, J., Zhou, F., Gebraeel, N. Z., & Duffy, V. (2020). Towards augmenting cyber–physical–human collaborative cognition for human–automation interaction in complex manufacturing and operational environments. *International Journal of Production Research*, *58*(16), 1–23.

Karlamangla, S. (2019). Data breach exposes information of thousands of patients in L.A. county. Retrieved from https://www.latimes.com/local/california/la-me-nemadji-breach-20190709-story.html

Karumbaiah, S., Wright, R. T., Durcikova, A., & Jensen, M. L. (2016). Phishing training: A preliminary look at the effects of different types of training. *Proceedings of the 11th Pre- ICIS Workshop on Information Security and Privacy* (pp. 1–10). ICIS.

Kim, J., & Gambino, A. (2016). Do we trust the crowd or information system? Effects of personalization and bandwagon cues on users' attitudes and behavioral intentions toward a restaurant recommendation website. *Computers in Human Behavior*, *65*, 369–379.

Kirlappos, I., Beautement, A., & Sasse, M. A. (2013). "Comply or die" is dead: Long live security-aware principal agents. In *International conference on financial cryptography and data security* (pp. 70–82). Springer.

Kirwan, B., Anderson, B., Eargle, D., Jenkins, J., & Vance, A. (2020). Using fMRI to measure stimulus generalization of software notification to security warnings. In *Information systems and neuroscience* (pp. 93–99). Springer.

Kleitman, S., Law, M. K., & Kay, J. (2018). It's the deceiver and the receiver: Individual differences in phishing susceptibility and false positives with item profiling. *PloS One*, *13*(10), 1–29.

Knop, N., Durward, D., & Blohm, I. (2017). How to design an internal crowdsourcing system. In *Proceedings of the 38th international conference on information systems*. ICIS.

Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M. A., & Pham, T. (2009). School of phish: A real-world evaluation of anti-phishing training. In *SOUPS'09: Proceedings of the 5th symposium on usable privacy and security* (pp. 1–12). ACM.

Kumaraguru, P., Rhee, Y., Hasan, S., Acquisti, A., Cranor, L., & Hong, J. (2007). Getting users to pay attention to anti-phishing education: Evaluation of retention and transfer. In *Proceedings of the APWG 2nd annual ecrime researchers summit* (pp. 70–81). ACM.

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2010). Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology (TOIT)*, *10*(2), 1–31.

Lakhmani, S. G., Wright, J. L., & Chen, J. Y. (2020). Transparent interaction and human–robot collaboration for military operations. In R. Pak, E. J. De Visser, & E. Rovira (Eds.), *Living with robots: Emerging issues on the psychological and social implications of robotics* (pp. 1–19). Academic Press.

Lankton, N. K., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems*, *16*(10), 880–918.

Lankton, N. K., McKnight, D. H., Wright, R. T., & Thatcher, J. B. (2016). Using expectation disconfirmation theory and polynomial modeling to understand trust in technology. *Information Systems Research*, *27*(1), 197–213.

Lawson, P., Pearson, C. J., Crowson, A., & Mayhorn, C. B. (2020). Email phishing and signal detection: How persuasion principles and personality influence response patterns and accuracy. *Applied Ergonomics*, *86*(103), 84.

Lee, J. D. (2008). Review of a pivotal human factors article: "Humans and automation: Use, misuse, disuse, abuse". *Human Factors*, *50*(3), 404–410.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80.

Lim, I.-K., Park, Y.-G., & Lee, J.-K. (2016). Design of security training system for individual users. *Wireless Personal Communications*, *90*(3), 1105–1120.

Lin, Z. (2014). An empirical investigation of user and system recommendations in e-commerce. *Decision Support Systems*, *68*, 111–124.

Liu, G., Xiang, G., Pendleton, B. A., Hong, J. I., & Liu, W. (2011). Smartening the crowds: Computational techniques for improving human verification to fight phishing scams. In *Proceedings of the seventh symposium on usable privacy and security* (pp. 1–13). ACM.

Love, J., & Hirschheim, R. (2017). Crowdsourcing of information systems research. *European Journal of Information Systems*, *26*(3), 315–332.

Lowry, P. B., & Moody, G. D. (2015). Proposing the control-reactance compliance model (CRCM) to explain opposing motivations to comply with organisational information security policies. *Information Systems Journal*, *25*(5), 433–463.

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

Madhavan, P., & Phillips, R. R. (2010). Effects of computer self-efficacy and system reliability on user interaction with decision support systems. *Computers in Human Behavior*, *26*(2), 199–204.

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*, *15*(4), 336–355.

Marsh, S., & Dibben, M. R. (2003). The role of trust in information science and technology. *Annual Review of Information Science and Technology*, *37*(1), 465–498.

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

McCrohan, K. F., Engel, K., & Harvey, J. W. (2010). Influence of awareness and training on cyber security. *Journal of Internet Commerce*, *9*(1), 23–41.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, *13*(3), 334–359.

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors*, *50*(2), 194–210.

Meservy, T. O., Jensen, M. L., & Fadel, K. J. (2014). Evaluation of competing candidate solutions in electronic networks of practice. *Information Systems Research*, *25*(1), 15–34.

Miller, C. A., & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human Factors*, *49*(1), 57–75.

Mohebzada, J., El Zarka, A., BHojani, A. H., & Darwish, A. (2012). Phishing in a university community: Two large scale phishing experiments. In *2012 international conference on innovations in information technology (IIT)* (pp. 249–254). IEEE.

Moody, G. D., Galletta, D. F., & Dunn, B. K. (2017). Which phish get caught? An exploratory study of individuals' susceptibility to phishing. *European Journal of Information Systems*, *26*(6), 564–584.

Moore, T., & Clayton, R. (2008). Evaluating the wisdom of crowds in assessing phishing websites. *Lecture Notes in Computer Science*, *5143*, 16–30.

Muir, B. M. (1988). Trust between humans and machines, and the design of decision aids. In E. Hollnagel, G. Mancini, & D. D. Woods (Eds.), *Cognitive engineering in complex dynamic worlds* (pp. 71–83). Academic.

Musuva, P. M., Getao, K. W., & Chepken, C. K. (2019). A new approach to modelling the effects of cognitive processing and threat detection on phishing susceptibility. *Computers in Human Behavior*, *94*, 154–175.

Natvig, K., Kular, D., and Neumann, R. (2020). Tax refunds & grants anyone? – How not to fall for simple phishing. Retrieved from https://www.latimes.com/local/california/la-me-nemadji-breach-20190709-story.html

Oliveira, D., Rocha, H., Yang, H., Ellis, D., Dommaraju, S., Muradoglu, M., Weir, D., Soliman, A., Lin, T., & Ebner, N. (2017). Dissecting spear phishing emails for older vs. young adults: On the interplay of weapons of influence and life domains in predicting susceptibility to phishing. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 6412–6424). ACM.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230–253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *30*(3), 286–297.

Parsons, K., Butavicius, M., Delfabbro, P., & Lillie, M. (2019). Predicting susceptibility to social influence in phishing emails. *International Journal of Human-Computer Studies*, *128*, 17–26.

Pearson, C. J., Geden, M., & Mayhorn, C. B. (2019). Who's the real expert here? Pedigree's unique bias on trust between human and automated advisers. *Applied Ergonomics*, *81*, 1–10.

Pedersen, J., Kocsis, D., Tripathi, A., Tarrell, A., Weerakoon, A., Tahmasbi, N., Jason Jie Xiong, Wei Deng, Onook Oh & De Vreede, G. J. (2013). Conceptual foundations of crowdsourcing: A review of IS research. In *2013 46th Hawaii international conference on system sciences* (pp. 579–588). IEEE

Pelaez, A., Chen, C. W., & Chen, Y. X. (2019). Effects of perceived risk on intention to purchase: A meta-analysis. *Journal of Computer Information Systems*, *59*(1), 73–84.

Petelka, J., Zou, Y., & Schaub, F. (2019). Put your warning where your link is: Improving and evaluating email phishing warnings. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–15). ACM.

Ramesh, G., Krishnamurthi, I., & Kumar, K. S. S. (2014). An efficacious method for detecting phishing webpages through target domain identification. *Decision Support Systems*, *61*, 12–22.

Rath, J., Ischi, M., & Perkins, D. (2014). Evolution of different dual-use concepts in international and national law and its implications on research ethics and governance. *Science and Engineering Ethics*, *20*(3), 769–790.

Reeder, R. W., Felt, A. P., Consolvo, S., Malkin, N., Thompson, C., & Egelman, S. (2018). An experience sampling study of user reactions to browser warnings in the field. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–13). ACM.

Shaw, R. S., Chen, C. C., Harris, A. L., & Huang, H.-J. (2009). The impact of information richness on information security awareness training effectiveness. *Computers & Education*, *52*(1), 92–100.

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J. (2010). Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of 28th international conference on human factors in computing systems* (pp. 373–382). ACM.

Sjouwerman, S. (2019). Men are twice as likely to fall for phishing attacks. Retrieved from https://blog.knowbe4.com/men-are-twice-as-likely-to-fall-for-phishing-attacks

Syed, R. (2019). Enterprise reputation threats on social media: A case of data breach framing. *The Journal of Strategic Information Systems*, *28*(3), 257–274.

Symantec (2014). Internet security threat report. Retrieved from http://www.symantec.com/content/en/us/enterprise/other_resources/bistr_main_report_v19_21291018.en-us.pdf

Tan, C. L., Chiew, K. L., & Wong, K. (2016). PhishWHO: Phishing webpage detection via identity keywords extraction and target domain name finder. *Decision Support Systems*, *88*, 18–27.

Thatcher, J., Wright, R. T., Sun, H., Klein, R., & Zagenczyk, T. (2018). Mindfulness in information technology use: Conceptual and operational definitions. *MIS Quarterly*, *42*(3), 831–847.

Thuan, N. H., Antunes, P., & Johnstone, D. (2013). Factors influencing the decision to crowdsource. In *Collaboration and technology* (pp. 110–125). Springer.

Vance, A., Jenkins, J. L., Anderson, B. B., Bjornn, D. K., & Kirwan, C. B. (2018). Tuning out security warnings: A longitudinal examination of habituation through fMRI, eye tracking, and field experiments. *MIS Quarterly*, *42*(2), 355–380.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, *46*(2), 186–204.

Verizon RISK (2019). 2019 data breach investigations report Retrieved from https://enterprise.verizon.com/resources/reports/2019-data-breach-investigations-report.pdf

Vishwanath, A. (2015). Examining the distinct antecedents of email habits and its influence on the outcomes of a phishing attack. *Journal of Computer-Mediated Communication*, *20*(5), 570–584.

Vishwanath, A., Harrison, B., & Ng, Y. J. (2018). Suspicion, cognition, and automaticity model of phishing susceptibility. *Communication Research*, *45*(8), 1146–1166.

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

Volkamer, M., Renaud, K., Reinheimer, B., & Kunz, A. (2017). User experiences of TORPEDO: Tooltip-powered phishing email detection. *Computers & Security*, *71*, 100–113.

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

Wang, J., Li, Y., & Rao, H. R. (2016). Overconfidence in phishing email detection. *Journal of the Association for Information Systems*, *17*(11), 759–783.

Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and reliance on an automated combat identification system. *Human Factors*, *51*(3), 281–291.

Wang, W., & Benbasat, I. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems*, *6*(3), 72–101.

Warby, S. C., Wendt, S. L., Welinder, P., Munk, E. G., Carrillo, O., Sorensen, H. B., Jennum, P., Peppard, P. E., Perona, P., & Mignot, E. (2014). Sleep-spindle detection: Crowdsourcing and evaluating performance of experts, non-experts and automated methods. *Nature Methods*, *11*(4), 385–392.

Williams, E. J., Hinds, J., & Joinson, A. N. (2018). Exploring susceptibility to phishing in the workplace. *International Journal of Human–Computer Studies*, *120*, 1–13.

Willison, R., Warkentin, M., & Johnston, A. C. (2018). Examining employee computer abuse intentions: Insights from justice, deterrence and neutralization perspectives. *Information Systems Journal*, *28*(2), 266–293.

Wogalter, M. S., Mayhorn, C. B., & Wogalter, M. S. (2021). Chapter 23: Warnings and hazard communications. In G. Salvendy & W. Karwowski (Eds.), *Handbook of human factors/ergonomics*. Wiley (in press).

Wood, W., & Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology*, *67*, 289–314.

Workman, M. (2008). 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), 662–674.

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

Wright, R. T., & Marett, K. (2010). The influence of experiential and dispositional factors in phishing: An empirical investigation of the deceived. *Journal of Management Information Systems*, *27*(1), 273–303.

Wu, M., Miller, R. C., & Garfunkel, S. L. (2006). Do security toolbars actually prevent phishing attacks? In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 601–610). ACM.

Zielinska, O. A., Welk, A. K., Mayhorn, C. B., & Murphy-Hill, E. (2016). A temporal analysis of persuasion principles in phishing emails. In *Proceedings of the human factors and ergonomics society annual meeting* (pp. 765–769). Sage.

Zuchowski, O., Posegga, O., Schlagwein, D., & Fischbach, K. (2016). Internal crowdsourcing: Conceptual framework, structured review, and research agenda. *Journal of Information Technology*, *31*(2), 166–184.

AUTHOR BIOGRAPHIES

Christopher Nguyen, Ph.D., is a management consultant at LMI. His educational background is in industrial and organisational psychology, where his work focusses on developing assessments and analysing workforce metrics to drive decision-making for diversity and inclusion initiatives. His research interests include diversity and inclusion in organisations, cybersecurity and group and team dynamics, and he has contributed to research published in *Management Communication Quarterly*, *Human Relations* and *the Journal of Applied Communication Research*.

Matthew L. Jensen, Ph.D., is an associate professor of Management Information Systems and a codirector of the Center for Applied Social Research at the University of Oklahoma. His interests include computer-aided decision-making, human-computer interaction and information security. He studies how people attribute credibility in mediated interactions and how people filter and evaluate information they find online. His research has been published in *Information Systems Research*, *Journal of MIS*, *MIS Quarterly* and other journals. He has been primary investigator or co-primary investigator on externally funded research projects totaling more than $8 million.

Alexandra Durcikova, Ph.D., is an associate professor in the MIS Division of the Price College of Business at the University of Oklahoma. Her research focusses on knowledge management systems, end-user security and healthcare IT. Her work is published in outlets such as *MIS Quarterly*, *Information Systems Research*, *Journal of MIS*, *Journal of AIS*, *European Journal of Information Systems* and several additional international journals and professional conferences. She has published more than 50 peer-reviewed journal and conference articles. Her research has been funded by the National Science Foundation, Department of Homeland Security, CITer and IBM.

Ryan T. Wright, MBA, Ph.D., is the C. Coleman McGehee Professor in the McIntire School of Commerce at the University of Virginia. His research interests include IT security and privacy and the diffusion of IT innovations. His research has been funded by the National Science Foundation, the State of Massachusetts, the State of Virginia. He has more than 75 publications in journals, books, cases and conference proceedings. He currently serves as IT Area Coordinator and the Associate Director for the Center of Management of IT at the University of Virginia. He also is presently an Associate Editor at MIS Quarterly and is co-conference chair for the International Conference on Information Systems in 2021.

How to cite this article: Nguyen C, Jensen ML, Durcikova A, Wright RT. A comparison of features in a crowdsourced phishing warning system. *Inf Syst J*. 2021;31:473–513. https://doi.org/10.1111/isj.12318

APPENDIX A

Literature review on susceptibility on phishing

We conducted a thorough literature review that included searching Google Scholar, IEEE, ACM and AIS databases for pertinent literature. We used the keywords 'phishing', 'phishing susceptibility' and 'phishing resilience' to identify 52 papers that empirically studied the factors that influence a user's susceptibility. There were four main categories of studies: (a) efficiency of different trainings or warnings, (b) properties of the phishing website or email and (c) psychographic factors of the recipients. The papers that met our criteria for the literature are listed in Table A1.

TABLE A1 Literature review of phishing susceptibility

Reference	Outlet	Focus	Subjects	Findings
Abbasi et al. (2021)	*Information Systems Research*	Warnings	1273 employees	Developed a susceptibility funnel model, developed warnings and tested the interventions over 3 months
Abawajy (2014)	*Behaviour & Information Technology*	Training	60 participants	Video training was effective, but the subjects preferred combined methods
Akdemir and Lawless (2020)	*Internet Research*	Psychographic	33 participants	Deviant online behaviours increase susceptibility
Alseadoon et al. (2012)	*ACIS Proceedings*	Psychographic	200 students	Subjects that have high openness, are submissive and lack suspicion predict susceptibility

(Continues)

TABLE A1 (Continued)

Reference	Outlet	Focus	Subjects	Findings
Alsharnouby et al. (2016)	*International Journal of HCI*	Warnings	21 students	Eye tracker study which found that security measures are typically not effective, but users that gaze at security measures are less susceptible
Arachchilage et al. (2014)	*Computers in Human Behavior*	Training	161 students	Revealed an interaction effect between conceptual and procedural knowledge positively impacts computer users' self-efficacy. CSE enhances their threat avoidance behaviour
Arachchilage and Love (2013)	*Computers in Human Behavior*	Training	20 students	Subjects exposed to this mobile game self-reported they were more motivated to detect phishing messages
Ayaburi and Andoh-Baidoo (2019)	*ICIS Proceedings*	Email/Web properties	228 students, faculty, staff	Urgency of phishing message influences susceptibility
Bansal (2018)	*MWAIS Proceedings*	Training	853 students, faculty, staff	Video training decreased susceptibility
Benenson et al. (2017)	*International Conference on Financial Cryptography and Data Security*	Email/Web properties	280 student Facebook users; 975 students	Examined susceptibility of Facebook messages over email. Curiosity and expectations influence click click-through as well
Blythe et al. (2011)	*CHI Proceedings*	Email/Web properties	224 students and staff	Found that logos and security language in the email increased susceptibility
Canfield et al. (2016)	*Human Factors*	Psychographic	100 and 162 from MTurk	Two experiments found that confidence and consequences impact actions on phishing messages. Also, performance of detection was different when asked to evaluate vs respond
Caputo et al. (2013)	*IEEE Security & Privacy*	Training	1359 employees	No difference in training categories
Dhamija et al. (2006)	*CHI Proceedings*	Email/Web properties	22 students	90% of users cannot identify a phishing website correctly
Diaz et al. (2020)	*Cryptologia*	Psychographic	1350 students	Correlations resulting in lowered susceptibility were based on college affiliation, academic year progression, cyber training, time spent on the computer and age demographics
Dodge et al. (2007)	*Computers & Security*	Training	4118 students	Concluded the need to continue phishing exercises with increased frequency
Dodge et al. (2012)	*IFIP International Information Security Conference*	Training	892 students	No change in susceptibility after training 10 days prior, but there was a change in susceptible after longer periods (63 days)

TABLE A1 (Continued)

Reference	Outlet	Focus	Subjects	Findings
Downs et al. (2006)	*Proceedings of the Second Symposium on Usable Privacy and Security*	Psychographic	20 community members	Interviews found that risk with familiar technology (email) can help mitigate phishing
Egelman et al. (2008)	*SIGCHI Conference on Human Factors in Computing System*	Training	60 participants	Passive phishing warnings are ineffective, and just-in-time warnings decrease phishing susceptibility
Eminağaoğlu et al. (2009)	*Technical Report*	Training	2900 employees	Education and awareness are the most effective and powerful mechanisms for mitigating information security risks
Gavett et al. (2017)	*PLoS One*	Psychographic	193 students	Education, interaction of age and prior phishing experience and the MAZE psychological assessment predict susceptibility
Goel et al. (2017)	*Journal of the AIS*	Email/Web properties	7225 students	Fear of losing or anticipation of gaining something valuable increased susceptibility to phishing
Harrison et al. (2016)	*HICSS Proceedings*	Psychographic	200 students	Suspicion decreased susceptibility
Harrison et al. (2019)	*ICIS Proceedings*	Training	422 students and staff	Findings suggest that integrated training approaches with mindfulness and cue based are superior
Hong et al. (2013)	*Proceedings of the Human Factors and Ergonomics*	Psychographic	53 students	Trust, extraversion and openness to new experiences are correlated to susceptibility. Also, women are less likely to identify phishing messages
Iuga et al. (2016)	*Human-centric Computing and Information Sciences*	Psychographic	386 Internet subjects	Females and low computer experience increase susceptibility. Pop-up based attacks have greater success
Jagatic et al. (2007)	*Communications of the ACM*	Psychographic and email properties	1731 students	If the receiver knows the identity of the sender, they are more likely to fall for a phishing message. Also, females are more susceptible to phishing attacks
Jansson and von Solms (2013)	*Behaviour & Information Technology*	Training	25 579 students and faculty	Simulated phishing with embedded training reduces likelihood of falling for phishing as shown in a decrease of 'insecure' reaction to phishing emails between two rounds
Jensen, Dinger, et al. (2017)	*Journal of Management Information System*	Training	355 students, faculties and staff	Participants who received mindfulness training were better able to avoid the phishing attack. This work introduces and provides evidence supporting a new approach that may be used to develop anti-phishing training

(Continues)

TABLE A1 (Continued)

Reference	Outlet	Focus	Subjects	Findings
Kirlappos et al. (2013)	*IEEE Security & Privacy*	Training	36 students	The authors argue that user education needs to focus on challenging and correcting the misconceptions that guide current user behaviour
Kleitman et al. (2018)	*PLoS One*	Psychographic	150 students	Perceptions of maliciousness, intelligence, knowledge of phishing and on-task confidence contribute significantly to phishing susceptibility, directly and/or indirectly via perception of maliciousness
Kumaraguru et al. (2009)	*Symposium on Usable Privacy and Security*	Training	515 students	The tool PhishGuru is effective in reducing phishing susceptibility, and individuals who are in 18 to 25 age group are most susceptible to phishing
Kumaraguru et al. (2010)	*ACM Transactions on Internet Technology*	Training	14 participants	The tool PhishGuru and online game Anti-Phishing Phil were effective user education methods for reducing phishing susceptibility
Lawson et al. (2020)	*Applied Ergonomics*	Psychographic and email properties	102 students	Certain persuasion techniques are trusted. Extroversion predicts susceptibility, as well as the interaction between personality and persuasion type
Lim et al. (2016)	*Wireless Personal Communications*	Training	481;1045 trainees	Training decreases clicking rate on phishing messages after four rounds of training
Luo et al. (2013)	*Computers & Security*	Email/Web properties	105 from university community	Argument quality, source credibility, genre conformity and pretexting influenced susceptibility and believability of the emails
Mayhorn and Nyeste (2012)	*Work*	Training	84 students	Comic and video game approaches to phishing training work for the college-aged demographic
McCrohan et al. (2010)	*Journal of Internet Commerce*	Training	180; 216 students	Subjects exposed to high-information lectures experienced an improvement in password security
Mohebzada et al. (2012)	*International Conference on Innovations in Information Technology*	Psychographic and training	10 000 from university community	No correlation between demographics and susceptibility. Emails warning users had no effect
Moody et al. (2017)	*European Journal of Information Systems*	Psychographic and email properties	632 students	Factors that influenced susceptibility included known sources, curiosity, risk, general Internet usage and Internet anxiety
Musuva et al. (2019)	*Computers in Human Behavior*	Psychographic	4483 from university community	Elaboration has a negative effect on susceptibility

TABLE A1 (Continued)

Reference	Outlet	Focus	Subjects	Findings
Parsons et al. (2019)	*International Journal of Human-Computer Interaction*	Email properties	985 participants	Social influence predicted susceptibility. The persuasion technique of consistency and reciprocity were the most successful at garnering a click
Oliveira et al. (2017)	*Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*	Psychographic	158 Internet users	Younger adults are susceptible to scarcity, while older adults are susceptible to reciprocity. Older women are the most susceptible group
Shaw et al. (2009)	*Computers & Education*	Training	240 students	Participants who have the better understanding at the perception and comprehension levels can improve understanding at the projection level
Sheng et al. (2010)	*CHI Proceedings*	Psychographic	1001 mTurk	Individuals in 18 to 25 age group are most susceptible. Women are more susceptible than men. Training reduces susceptibility by 40%
Vishwanath et al. (2011)	*Decision Support Systems*	Psychographic	325 students	Found that email load and relevance of message increase vulnerability to phishing messages
Vishwanath (2015)	*Journal of Computer-Mediated Communication*	Psychographic	200 students	The decision performance of subjects is particularly when openness and conscientiousness are very low
Vishwanath et al. (2018)	*Communication Research*	Psychographic	Students	A dual process modelling of phishing is supported. Risks and self-regulation were related to susceptibility
Volkamer et al. (2017)	*Computers & Security*	Training	43 students	A field study suggests that placing of just-in-time tool tips cut a user's susceptibility in half
Wang et al. (2016)	*Journal of the AIS*	Psychographic	600 Internet subjects	Subjects' overconfidence in detecting phishing emails is driven by cognitive effort but does not predict detection accuracy
Wang et al. (2012)	*IEEE Transactions on Professional Communication*	Psychographic	321 student and staff	Cognitive effort significantly affects response likelihood. 'Attention to visceral triggers, attention to phishing deception indicators and phishing knowledge play critical roles in phishing detection'
Williams et al. (2018)	*International Journal of Human-Computer Studies*	Email/Web properties	62 000 employees	This study found that authority and urgency in the email are found to impact susceptibility
Workman (2008)	*Journal of the American Society for Information Systems*	Psychographic and email properties	612 employees	Normative communication, continuance commitment and trust predict susceptibility
Wright and Marett (2010)	*Journal of MIS*	Psychographic	446 students	Computer self-efficacy, experience with security and Web and suspicion predict phishing susceptibility

(Continues)

TABLE A1 (Continued)

Reference	Outlet	Focus	Subjects	Findings
Wright et al. (2014)	*Information Systems Research*	Email/Web properties	2600 students and staff	Found that different types of persuasion techniques increased susceptibility
Zielinska et al. (2016)	*Proceedings of the Human Factors and Ergonomics Society*	Email properties	887 phishing emails assessed by two reviewers	The use of different types of persuasion techniques in phishing messages has changed over time, with an increase in commitment/consistency and scarcity but a decrease in reciprocation and social proof

APPENDIX B

Example emails

Example phishing emails

From: "Netflix" <noreply@netl.com>
To: noreply@netflix.com
Subject: You need to update your payment method
Date: Mon, 09 Jan 2017 10:11:15 +0730

Update your payment method

We were unable to bill your membership for the current month. To ensure that the service will not be interrupted, please update your payment method.

To update your payment method, click: Sign In to Netflix then you will be prompted to update your payment method.

- The Netflix Team

From: "University Information Technology" <itfeedback@university.edu>
To: noreply@university.edu
Subject: University Email Sign-in Alert
Date: Tue, 10 Jan 2017 02:36:31 +0730

We detected a login attempt with valid password to your university email account from an unrecognized device on Mon Jan 11, 2017 08:11 AM CDT.
Location: Germany (IP = 81.190.142.39)
Note: The location is based on information from your Internet service or wireless carrier provider.

Was this you? If so, you can disregard the rest of this email.

If this wasn't you, please LOGIN HERE to confirm your ownership of this account and to protect your email account information from potential future account compromise.

The office of Information Security will keep this updated if information should change, but we encourage all users to run their updates after the expected release of this patch.

University Information Technology (University IT)

From: "University Canvas" <canvas@university.edu>
To: noreply@university.edu
Subject: Canvas Access: Important Course Error Alert
Date: Wed, 11 Jan 2017 11:26:44 +0730

We detected something unusual about a recent sign-in to your Canvas account. For example, you might be signing in from a new location, device or app.

To keep you safe, we've blocked access to your inbox, contacts list and calendar for that sign-in. Please review your recent activity and we'll help you take correct action. To regain access, you'll need to confirm your identity.

Thanks,
Canvas Administrative.

From: "Dropbox" <no-reply@dropbox.com>
To: noreply@dropbox.com
Subject: Verify your email
Date: Thu, 12 Jan 2017 07:21:12 +0730

Hi,

Someone just shared a document with you via Dropbox. We just need to verify your email address before you can view/share the received file/folders. You are required to sign in with your email address to access your folder.

Verify your email

Thanks!
- The Dropbox Team

From: "Facebook" <security@facebookmail.com>
To: noreply@facebookmail.com
Subject: Facebook account update
Date: Fri, 13 Jan 2017 02:12:14 +0730

Dear Facebook User,

In an effort to make your online experience safer and more enjoyable, Facebook will be implementing a new login system that will affect all Facebook users. These changes will offer new features and increased account security.

Before you are able to use the new login system, you will be required to update your account.
Click here to update your account online now.

If you have any questions, reference our New User Guide.

Thanks,
The Facebook Team

Example legitimate emails

From: "Amazon.com" <account-update@amazon.com>
To: noreply@amazon.com
Subject: Revision to Your Amazon Account
Date: Sat, 14 Jan 2017 12:22:22 +0730

Thanks for visiting Amazon! Per your request, we have successfully changed your password.

Visit your account to view your orders, make changes to any order that hasn't yet entered the shipping process, update your subscriptions, and much more.

Should you need to contact us for any reason, please know that we can give out order information only to the name and email address associated with your account. Thanks again for shopping with us.

From: "Information Technology" <universityit@university.edu>
To: All Students <student@university.edu>
Subject: Your University Student Email is Getting an Upgrade!
Date: Sun, 15 Jan 2017 03:22:33 +0730

Your University Student Email is Getting an Upgrade!

Good news! Your university student email is getting a FREE upgrade to Office 365. This new offering includes access to the latest Office 365 products, which are already available to you at portal.office.com (you can login now with your university email address and password).

Need Help?
If you need assistance, please visit askit.university.edu, call 325-HELP (4357) during normal business hours, or email needhelp@university.edu at any time.

University Information Technology

From: "LinkedIn" <messages-noreply@linkedin.com>
To: messages-noreply@linkedin.com
Subject: People are looking at your LinkedIn profile
Date: Tue, 17 Jan 2017 10:26:41 +0730

The Power of your Profile
See how well your profile stands out from the crowd

A profile view from the right person is a step closer to your next career move.
Who's looking?

If you need assistance or have questions, please contact LinkedIn Customer Service.

From: "Desire2Learn Administrator" <learn@university.edu>
To: noreply@university.edu
Subject: You have been added to a University D2L Course
Date: Mon, 16 Jan 2017 04:52:13 +0730

This is an automatic message sent because you have been added to CAS ONLINE ORIENTATION COURSE at the university's learning management system, Desire2Learn (D2L).

If you are receiving this message, it is because someone has added you manually to a course or because you self-registered in a course inside D2L.

Click here to log in.

If you have problems logging in, please review the 'Login Trouble?' area on the front page under the login box.

Students: This is not a confirmation of your official enrollment at the university. To verify your official enrollment, go to http://ozone.university.edu.

D2L Administrator
learn@university.edu

From: "Google Drive Team" <drive-noreply@google.com>
To: drive-noreply@google.com
Subject: Your 33 files stored in Google Docs are now in Google Drive
Date: Wed, 18 Jan 2017 12:53:41 +0730

Hi,
We're writing to let you know about important changes to Google Docs.

Google Drive is the new home for Google Docs

This means the 10 files that you own and the 23 files that have been shared with you will now be available in Drive, and you can access them anytime here.

You can still do everything you could before, like create, share and collaborate with Google documents, spreadsheets and presentations. Now, you can access your stuff anywhere, find files faster and work with more web apps.

On behalf of files everywhere,
The Google Drive Team

APPENDIX C

Example email identification task instructions and warning message

Title: *Testing a Phishing Warning System*

Description: *This new email system is designed to help you assess email messages by providing warnings of potentially suspicious messages*.

In the following pages, you will be presented with a number of email messages. After reading each email message, you will be asked to identify whether you believe the message is a phishing message or not, as well as how confident you are in your assessment.

In addition, you may receive warning messages for some emails that have been reported as potentially suspicious. These warnings include both the number of reports and the report source. The number of reports refers to the number of times an email or one similar has been reported as possible phishing, and the reports come from an automated anti-phishing software.

Example warning messages may look like the following:

WARNING POSSIBLE PHISHING
Number of Reports: 5
Report Source: Automated Anti-Phishing Software

Please note that based on past performance, these reports are 40% correct.

APPENDIX D

Descriptive statistics and measurement properties

TABLE D1 Measurement properties of study covariates

Item number	Item text	Standardised factor loading	Mean	SD	Coefficient alpha
Perceived internet risk					
Risk1	In general, it would be risky to give my information to online companies	0.81	4.97	1.32	.88
Risk2	There would be a high potential for loss associated with giving my information to online firms	0.77	4.61	1.38	
Risk3	There would be too much uncertainty associated with giving my information to online firms	0.83	4.71	1.42	
Risk4	Providing online firms with my information would involve many unexpected problems	0.80	4.27	1.44	
Disposition to trust					
Trust1	I usually trust people unless they give me a reason not to trust them	0.83	5.00	1.39	.88
Trust2	I generally give people the benefit of the doubt	0.66	5.08	1.18	
Trust3	Most of the time, I believe people until they give me a reason not to believe them	0.85	5.20	1.23	
Trust4	My typical approach is to trust people until they prove I should not trust them	0.85	4.96	1.36	
Mindfulness in information technology use					
Mindfulness1	I am very creative when using technology	0.75	4.67	1.42	.90
Mindfulness2	I am often open to hearing new ways of using technology	0.85	5.68	1.19	
Mindfulness3	I like to figure out different ways of using technology	0.87	5.48	1.27	
Mindfulness4	I 'get involved' when using technology	0.87	5.17	1.38	
Phishing identification expertise					
Phishing1	I know what a phishing message looks like	0.92	4.72	1.68	.85
Phishing2	I am confident in my ability to identify phishing messages	0.95	4.70	1.63	
Phishing3	I know what to do with phishing messages when I detect them	0.63	4.08	1.81	
Phishing4	I am confident that I can identify SPAM when I receive it	0.54	5.40	1.30	
Email expertise					
Email1	I can process new emails in my inbox rapidly	0.88	5.52	1.08	.84
Email2	I can understand new messages in my inbox swiftly	0.82	5.62	0.95	
Email3	I can get through my emails quickly	0.71	5.70	1.09	

(Continues)

TABLE D1 (Continued)

Item number	Item text	Standardised factor loading	Mean	SD	Coefficient alpha
Computer self-efficacy—Internal					
CSE1	…if there was no one around to tell me what to do as I go	0.83	3.57	1.43	.80
CSE2	…if I had never used a package like it before	0.82	3.26	1.33	
CSE3	…if I had only the online help manuals for reference	0.63	4.63	1.33	
Computer efficacy—External					
CSE4	…if I could call someone for help if I got stuck	0.72	5.27	1.33	.80
CSE5	…if someone else had helped me get started	0.91	5.59	1.13	
CSE6	…if someone showed me how to do it first	0.70	6.17	1.03	
Gaming experience					
Game1	I like participating in computer games	0.95	4.64	1.71	.90
Game2	I would be interested in playing a computer game	0.97	4.75	1.73	
Game3	I like to compete in any computer games I play	0.80	4.26	1.84	
Game4	I find most computer games a waste of time	0.62	4.07	1.65	
Game5	I would enjoy playing a computer game at my job	0.66	4.49	1.76	
Anxiety with system use					
Anxiety1	I would feel apprehensive about using a phishing warning system	0.54	3.85	1.27	.75
Anxiety2	It scares me to think that I could lose a lot of information using a phishing warning system by hitting the wrong key	0.59	4.79	1.44	
Anxiety3	I would hesitate to use a phishing warning system for fear of making mistakes I cannot correct	0.79	4.30	1.30	
Anxiety4	The phishing warning system is somewhat intimidating to me	0.69	3.79	1.41	

Note: All items measured on a 1- to 7-Likert scale.

TABLE D2 Means, SDs, reliabilities, and correlations of study variables

Variable	M	SD	1	2	3	4	5	6	7	8	9
1. Perceived internet risk	4.64	1.19	(.88)								
2. Disposition to trust	5.06	1.10	−.14**	(.88)							
3. Mindfulness in technology	5.25	1.15	.02	.09	(.90)						
4. Phishing identification expertise	4.72	1.34	.06	−.08	.24***	(.85)					
5. Email experience	5.61	0.91	.03	.13	.28***	.21***	(.84)				
6. Computer self-efficacy—Internal	3.82	1.15	−.03	−.11*	.38***	.24***	.17**	(.80)			
7. Computer self-efficacy—External	5.68	0.99	.01	.08	.26***	.16**	.29***	.36***	(.80)		
8. Gaming experience	4.44	1.47	.04	.00	.33***	.15**	.13**	.25***	.15**	(.90)	
9. Anxiety with system use	4.18	1.02	.14**	.05	−.11*	−.15**	−.11*	−.11*	−.07	−.13**	(.75)

Note: Diagonals are internal consistencies. All variables measured on a 1- to 7-Likert scale. $N = 438$.
*$P < .05$;
**$P < .01$;
***$P < .001$ (two-tailed).