Chapter 5
Examinations of Email Fraud Susceptibility: Perspectives From Academic Research and Industry Practice

Helen S. Jones
University of Dundee, UK

John Towse
Lancaster University, UK

ABSTRACT

The internet provides an ever-expanding, valuable resource for entertainment, communication, and commerce. However, this comes with the simultaneous advancement and sophistication of cyber-attacks, which have serious implications on both a personal and commercial level, as well as within the criminal justice system. Psychologically, such attacks offer an intriguing, under-exploited arena for the understanding of the decision-making processes leading to online fraud victimisation. In this chapter, the authors focus on approaches taken to understand response behaviour surrounding phishing emails. The chapter outlines how approaches from industry and academic research might work together to more effectively understand and potentially tackle the persistent threat of email fraud. In doing this, the authors address alternative methodological approaches taken to understand susceptibility, key insights drawn from each, how useful these are in working towards preventative security measures, and the usability of each approach. It is hoped that these can contribute to collaborative solutions.

INTRODUCTION

In 2016, the rate of malicious emails being sent to users was at its highest in five years. For example, in relation to one specific type of phishing, approximately 1 in every 131 emails contains malware (Verizon, 2017). Despite efforts from experts in the field, email fraud remains one of the most pertinent cyber security threats. The persistence of this threat indicates a need for reconsideration of mitigation methods

DOI: 10.4018/978-1-5225-4053-3.ch005

in place to protect against this, as those currently employed do not seem to be sufficient to counteract it. In line with this, there is also a need to consider the effectiveness of methods used within a research setting to improve our understanding of how users become victims to social engineering attacks, such as phishing emails. It is crucial that there is an alignment between the theoretical knowledge base gained through academic research, and the practical role this has in industry efforts to tackle email fraud.

Research across multiple disciplines has considered how best to address the threat posed by social engineering attacks. Computer science research is often concerned with systems-based approaches to managing fraud through the use of detection algorithms (Islam & Abawajy, 2013; Salah, Alcarez Calero, Zeadally, Al-Mulla, & Alzaabi, 2013) or automated heuristic filters, which detect machine learned patterns (Abu-Nimeh, Nappa, Wang, & Nair, 2007; Garera, Provos, Chew, & Rubin, 2007) to prevent emails from reaching the user. However, simultaneous advancement in the techniques employed by the fraudster means that these solutions are short-lived, as a work around is often found within a short space of time to circumvent such detection algorithms. In addition to this, these machine learning approaches tend to focus more on the detection of generic phishing emails, with detectable anomalies to legitimate email traffic. They may be less suited to the detection of more sophisticated attacks that either employ a hacked account, or are more personalised to appear believable. In these cases, the attacker is targeting what is often considered the systems' weakest link – the human user (Barrett, 2003; Mitnick & Simon, 2002; Schneier, 2000).

As the 'weakest link' in cyber security, the human user and the decision-making processes they employ in email management must be understood in order to address the threat and reduce system vulnerability. Contributions from psychology have considered how various factors can affect email response behaviour, from individual differences amongst users, to the context in which a specific email is read. Unfortunately, such findings are constrained by limitations in conducting fundamental cyber security research from both a practical and an ethical perspective. On the other hand, industry experts in cyber security conduct training exercises and vulnerability tests within organisations without the same constraints that feature within academic research. This chapter will outline how the approaches taken in academia and industry to understand and address issues relating to email decision-making can complement one another. In doing this, the authors aim to highlight the importance of unity between these two approaches, emphasising the need for continued collaboration in future research in order to maximise the effectiveness of efforts made to tackle the persistent threat of email fraud.

EMAIL FRAUD TYPOLOGY

As most internet users will be aware, phishing emails come in all shapes and sizes, covering an array of subjects from sale of Viagra pills to urgent account updates. As such, providing a specific definition of phishing is not straightforward, although one useful example comes from Myers (2007):

Phishing: A form of social engineering in which an attacker, also known as a phisher, attempts to fraudulently retrieve legitimate users' confidential or sensitive credentials by mimicking electronic communications from a trustworthy or public organisation in an automated fashion. (p. 1)

Across the wide array of fraudulent emails in circulation, there are a number of factors that allow for a broad categorisation. Three main types of phishing emails that commonly exist will be outlined.

These are often successful in deceiving thousands of users. A series of specific real world examples will then be outlined to demonstrate how these different approaches and techniques might be incorporated in genuine phishing attacks.

1. **Deceptive Emails:** In the most generic sense, deceptive emails are distributed to thousands of users in an attempt to gather confidential information. This vast distribution only require a small response rate to be economically worthwhile for the fraudster. These emails usually attempt to solicit account information, passwords, or install malicious software. Most often, users are required to download a file, or click a link embedded within the email. Following a link may then ask them to input their login details to a fake website, purporting to come from a genuine organisation. Additional persuasive tactics may be employed in these emails, such as a sense of urgency, leading users to panic about losing access to accounts if they do not act. Empirical evidence demonstrates the impact of time pressure on decision-making, as discussed in more detail below, emphasising the impact that this type of persuasive approach can have on a user.
2. **Spear Phishing:** As opposed to the generic phishing emails that are distributed to many users, spear phishing is a sophisticated social engineering technique that targets an attack towards a specific user or group of users. By accessing information about the recipient from social media and public web pages, or with access to insider knowledge, these emails incorporate personal or particularly relevant information, in order to make an attack more believable.
3. **Whaling:** One specific type of spear phishing attack involves imitation of a senior executive of a company, known as whaling. This acts as a way of convincing an employee to respond and act in a way that benefits the fraudster. For example, they may be asked to transfer money to a fraudulent account, under the premise of a legitimate business transaction. This approach is typically conducted by gaining access to the executive's account to send the email, or through the use of a convincingly similar domain name to that of the company being targeted. The specificity of information incorporated into these emails means users often overlook the minor details that indicate the fraudulent nature of the message.

REAL WORLD CASE STUDIES

Media reports of cyber security attacks are becoming more frequent, as they increase in scale and sophistication. Particular attention is drawn to attacks on large corporations that hold confidential data on thousands, or even millions, of members of the public as customers. The high profile individuals whose response decisions allow access to this data are often scrutinised, in particular when there is substantial financial loss as a result of the attack. Such cases demonstrate the extent of risk associated with user response to email fraud, and also the diversity in approaches taken by the fraudsters to target these organisations.

In late 2013, US store Target was the victim of a substantial data breach, with credit card information for around 110 million customers stolen (Peterson, 2014), and it all began with a malware infected phishing email (Picchi, 2014). By targeting a heating and air conditioning firm that were subcontracted by Target, the fraudsters were able to obtain network credentials and access confidential sales and customer data. Although the financial impact of this case has not been reported, it demonstrates the vulnerability of companies who have a strong obligation to protect confidential customer information.

During the US election campaign in late 2016, thousands of emails from the personal account of John Podesta, campaign chairman for Hilary Clinton, began appearing on the whistle-blowing website WikiLeaks (Smith, 2016). These shed light on relationships and disputes between party members, as well as campaign strategies. As the story evolved, it became apparent that these emails were accessed following a phishing attack purporting to be from Google, which Podesta had responded to even after seeking advice from colleagues about its legitimacy. The email inferred unauthorised access to the account and provided a link for Podesta to change his password in order to ensure security. The success of this attack demonstrates the potential impact of response not only on the user, but also in this case on the integrity of an entire political campaign.

In early 2017, reports hit the headlines of two US companies being victims of a phishing attack that led to over $100 million being transferred to the fraudster over a two-year period (Yuhas, 2017). By posing as an employee of a manufacturing company in Asia that regularly conducted business with the companies, the fraudster convinced employees to set up numerous multi-million dollar transactions. It was later announced that Google and Facebook were the victimised companies (Roberts, 2017), two of the world's biggest technology companies. This targeted attack, using a deceptive cover and mirroring transactions that employees were familiar with organising, demonstrates the level of sophistication that can be achieved.

The case studies outlined here give an insight into the different types and variations on phishing emails, from generic password changes to sophisticated social engineering attacks. These also highlight the differing impacts that these can have on the victim or organisation, from confidential data breaches to extensive financial loss.

BEHAVIOURAL EXAMINATIONS OF SUSCEPTIBILITY IN INDUSTRY

The case studies highlighted above give only a brief snapshot of the degree to which organisations across a range of industry sectors can demonstrate vulnerability to email fraud, with varying magnitude. As such, many organisations acknowledge that there is a need to address cyber security concerns. It is becoming more common for companies to employ cyber security experts to conduct vulnerability testing, and the need for this is increased by the introduction of new regulations on data privacy, such as the EU General Data Protection Regulation. It is possible that this shift in attitude towards engaging with additional security measures may be due to compliance with legislation, rather than a deeper understanding of the risks faced, but regardless it shows a step in the right direction towards addressing ongoing cyber security threats. In this section, the authors will outline the approaches commonly used in an industry setting to assess vulnerability to email fraud, and consider how these can benefit organisations, whilst going on to consider how they may be enhanced by integration with theoretical insights from academic research.

Methods

When engaging cyber security experts to mitigate risk, there are three common routes an organisation can take. On the most basic level, some organisations opt for a vulnerability test, which provides a basic overview of the weaknesses within a system, based on data from an automated scan that picks up common errors and configuration mistakes (Yeo, 2013). This type of testing does not allow experts to

examine the data that they could potentially obtain through these vulnerabilities. Instead, it highlights surface-level issues, such as missing security patches, that can be fixed promptly and without occupying the time of company employees. It also does not take into account the vulnerability that can result from employee behaviour online.

In addition to this, many organisations employ a penetration testing team to demonstrate the scale of the vulnerability. This allows information to be gathered about whether they are able to infiltrate the organisation's network and systems, and what extent of data they are able to attain. This can involve: an attack on the physical network, to assess whether the devices used by employees can be hacked; attempts to bypass the security solutions that the organisation currently has in place, such as firewalls; and assessment of the effectiveness of attacks targeted at employees of the company, through user-based vulnerability testing, which will be discussed in more detail later in this section.

One major concern in employing external penetration testers, also known as ethical hackers, is their trustworthiness. The organisation is, in a sense, encouraging these experts to hijack their systems, and as a result giving them access to potentially confidential data. A penetration test can either be conducted externally, testing how easily an organisation can be hacked from outside the network, or internally to assess the risk associated with insider attacks. An organisation must have faith that these ethical hackers will not become malicious, taking advantage of an organisation at its most vulnerable for personal gain. The need for this kind of security testing may outweigh the potential cost, and an organisation often has no choice but to trust the ethical hackers (Duke, 2002).

The nature of a penetration test is dependent on the specific needs of an organisation, who will highlight their main concerns to the testers and allow them to conduct simulated attacks as appropriate. For organisations that are concerned with the risk posed by employee behaviour, a penetration test might involve a behavioural assessment through administration of a simulated phishing attack, designed for the purpose of the investigation. The employees who receive the phishing email are kept unaware of the vulnerability testing being conducted, in order to provide a realistic assessment of response behaviour. The rate of response to the simulated attack informs the organisation of the level of risk faced as a result of human decision-making. However, penetration testing only captures the level of vulnerability within an organisation at the given point in time when an attack is simulated. In reality, the level of vulnerability is likely to vary with contextual and organisational changes within the organisation, but multiple simulated attacks across different contexts would be time consuming and may raise suspicion with employees, impairing the validity of the assessment.

In contrast, a comprehensive cyber risk assessment goes a long way to overcome this issue of context, by considering the specific assets that leave an organisation most vulnerable, the likelihood of an attack on these occurring, the impact that could come from this, and the risk management strategies that would be most effective in addressing it (NIST, 2012). Whilst this type of assessment often incorporates vulnerability and penetration testing, it will also go beyond these to explore the risks posed by associated organisations, and shared data networks. In addition, the risk management strategies that result from such an assessment are continually monitored and updated in line with the transient nature of threats faced by the company, and the varying levels of vulnerability encountered.

These three approaches offer differing, but complimentary, methods for understanding the extent of an organisation's vulnerability. Each has an alternative outcome measure, and as such the decision about which to employ is dependent upon the needs of a specific organisation. The key insights that can be attained through the use of such techniques are outlined in more detail below.

Key Insights and Usability

In relation specifically to assessing vulnerability to phishing emails, the use of actual simulated attacks, in which the recipients are unaware of the artificial nature of the email received, means a naturalistic assessment of susceptibility in the work place can be achieved. Simulated phishing attacks as part of a penetration test usually require specific hypotheses that shape the email stimuli, have certain success criterion that demonstrate vulnerability within employees, and require use of a method that can be replicated in future assessments (Barrett, 2003). The rigour involved in the design and implementation of such a test is akin to an empirical experiment carried out as part of an academic research project. Unlike a research study though, there are fewer limitations on the methods used in order to simulate an attack, *"*the imagination of the social engineer is the only limit to the types of approaches that they can present and exploit*"* (Barrett, 2003). These tests demonstrate to organisations that the behaviour of employees can put their organisation at risk during an attack, which often results in additional training on how to detect phishing attacks and reduce vulnerability if an actual attack were to happen.

However, some have argued that there are limitations to the techniques involved with simulated attacks, as well as more broadly with penetration testing, which impair the usefulness of this technique in reducing future user vulnerability within an organisation. It has been noted by experts in the field that the lack of continuity or common taxonomy in the type of simulations conducted (Hudic, et al., 2013) makes replication and comparison of results across organisations difficult, in particular where tailored social engineering attacks are implemented (Barrett, 2003). As mentioned above though, the specific risks faced by an organisation vary dependent on their assets, and also across different contexts. As such, a simulated attack requires an element of tailoring (that has the down side of making it non-replicable) in order to give a valid assessment of vulnerability. It is therefore important to consider the balance of assessment rigour and validity in relation to the specific organisation, in order to optimise the impact and utility of this technique.

As an assessment of an organisation's vulnerability to targeted social engineering attacks, penetration tests are limited by an ethical responsibility to avoid infiltration of third party systems and linked organisations (Barrett, 2003). In this sense, the test may underestimate vulnerability, given the magnitude of additional information that could be gathered from these external sources in order to generate or target an attack that affects the organisation itself. As demonstrated in the earlier Target case study, access to confidential data was breached as a result of someone in a third party organisation with close ties to Target, responding to an email attack. By considering the hypothetical scenarios in which third party organisations might increase vulnerability, a cyber risk assessment also describes the additional risks associated with shared access to confidential data and information systems. This is typically accounted for in the development of a risk mitigation strategy that includes monitoring of changes in who has access to data, and addresses the threat associated with this.

Whilst these methods provide a valuable assessment of vulnerability within a naturalistic environment, and in many cases incorporate actual response behaviour through simulated attacks, they provide little insight into the underlying behavioural processes that influence employee behaviour. In relation to penetration testing, this means that there is no assessment of the context in which an email is received and read, which may have an impact on the response decision, as shown in academic research, discussed in more detail later in this chapter. Whilst risk assessments are designed to consider a broader range of vulnerabilities within the organisation, they still lack the depth to understand why some employees may be more susceptible to attack than others. In cases where individual employees *are* recognised for their

demonstration of vulnerability during a simulated attack, this can lead to disciplinary actions (Barrett, 2003) rather than attempts to understand and modify behaviour. Where training is invoked upon employees, this can range from warnings and response instructions, to more informative educational and support programmes. The former has limited success in reducing response behaviour though (Junger, Montoya, & Overink, 2017), whilst the latter is less common due to time constraints and cost of implementation. Instead of encouraging secure behaviour in the future, the use of simulated phishing attacks and training may have an unforeseen effect of employee disgruntlement with the process (Murdoch & Sasse, 2017).

By conducting simulated attacks in the work place, there is an overarching ethical concern that these are in some ways designed to 'catch out' employees. The discontent that results from this means that employees may not acknowledge the benefits that could be elicited through more secure behaviour in the future. Compliance in general is a concern, in relation to risk mitigation strategies that come from a cyber risk assessment also. It is therefore important to consider how such risk mitigation strategies might be informed by our understanding of human behaviour and decision-making processes to better understand both vulnerability, and behaviour change around security protocols. Rather than seeking to punish and patronise employees, this understanding would allow effective user-centric security solutions. This is where insights from academic research might be able to better inform the development of security initiatives taken within industry organisations, in an effort to enhance effective response behaviour surrounding cyber threats.

BEHAVIOURAL EXAMINATIONS OF SUSCEPTIBILITY IN RESEARCH

Psychological approaches from academic research offer an alternative perspective to cyber security behaviour, by attempting to understand the underlying mechanisms that influence behaviour. Focusing on the decisions made by human users, such research has attempted to understand *why* certain users demonstrate a higher level of susceptibility to email fraud. Reports suggest a fraudulent email response rate of approximately 5% (Norton, 2014), meaning that 95% of users who receive the same email do not respond. This may be because some did not see the email, some may have deleted it without reading it, but a large proportion are likely to have read (at least some of) the email and made an explicit decision not to respond. This might be because it was not relevant to them, or because they recognised its fraudulent nature. In all cases, the processes underlying such decisions offer an interesting set of clues that may be eventually help optimise secure online behaviour (Fischer, Lee, & Evans, 2013).

Most research studies are conducted within a controlled environment, allowing the researcher to manipulate various aspects of the stimuli and context in which this is viewed, to understand specific behaviours. However, this also means that the naturalistic assessment of susceptibility is jeopardised in many cases. It is therefore important, for this emerging field of psychology, to understand how this level of experimental control affects decision-making behaviour, and consider ways to enhance the validity of research, whilst maintaining ethical integrity.

Methods

Whilst approaches from industry are most often concerned with the immediate and applied need to manage organisational risk by reducing employee vulnerability, academic research places greater emphasis on identifying conceptual relationships between variables, testing and validating theories of causal processes.

Research has adopted a variety of approaches to assess human behaviour surrounding phishing emails, which perhaps contributes to the lack of consistency. In terms of behavioural assessments of susceptibility, methods have ranged from explicit legitimacy judgments of email stimulus, to simulated attacks whereby participants have given consent to an alternative research study. The use of explicit judgment tasks allows for control of situational influences on response behaviour, such as time pressure (Yan & Gozu, 2010; Jones, Towse, Race, & Harrison, submitted), whilst simulated emails provide a more realistic measure that still allows for some level of control over the stimulus that participants are responding to.

In an explicit judgment task, participants are typically asked to make a decision about the legitimacy of a series of email stimulus, or asked how they would respond to each of these. Yan and Gozu (2012) showed participants a set of 36 emails, all of which were genuine phishing emails, and asked them whether they would 'read' or 'delete' each of these. In this particular task, results may be limited by the use of phishing emails only, meaning that participants who were looking to differentiate between phishing and legitimate emails may have demonstrated an expectancy bias. Alternative versions of this type of task have incorporated a mixture of phishing and legitimate emails (Jones, et al., submitted; Nicholson, Coventry, & Briggs, 2017), allowing for a more representative example of what a participant might see in their own inbox. However, such methodologies are still limited by the explicit nature of the legitimacy judgments, which does not reflect the complexity of the decision-making process that likely occurs in real life. Instead, participants are encouraged to engage in certain behaviours, such as employing more rational decision-making strategies, which results in an artificially high accurate response rate to the email stimulus (Yan & Gozu, 2012; Harrison, Vishwanath, & Rao, 2016).

Alternative approaches to the use of explicit judgment tasks have incorporated a role-play element, whereby participants are asked to interact with and manage the inbox of a fictional employee, and report how they would respond to a series of emails (e.g. Downs, Holbrook, & Cranor, 2007; Hong et al., 2013). This variation on the task allows for an assessment of how users interact with an actual inbox, without being alerted to the nature of the task. However, it is possible that the nature of the task itself, asking participants to make response decisions to a set of emails, may alert them to the purpose of the research. This being said, Parsons et al. (2013) demonstrated that participants showed higher accuracy in a role-play scenario like this when they were aware of the nature of the task, compared to participants who were only told the purpose of the study after completing that task. This difference in response behaviour suggests that the naïve participants were not alerted to the purpose of the study whilst completing it. This study raises an important point about the accuracy of response rates seen in lab-based studies of email behaviour though, demonstrating that making explicit judgments may elicit an artificially low level of susceptibility. However, there is little evidence to indicate how this artificial response behaviour relates to real world susceptibility.

An additional limitation to such lab-based assessments of response behaviour is the lack of consequence associated with responding, in comparison to real life. When responding through a simulated account, the participant has nothing to lose in choosing to respond or not respond. This is in contrast to a genuine attack within industry, where an error in judgment can lead to data leaks, espionage, and financial loss. Within the lab environment on the other hand, participants may receive a reward for accurate responses, but the consequences for inaccurate responses do not equate. As it stands, lab-based studies provide an ethically sound alternative to simulated attacks, but with potentially limited validity.

One way of addressing the uncertainty in the validity of judgment tasks as a measure of susceptibility is to consider working with past victims of email fraud (e.g. Button, Nicholls, Kerr, & Owen, 2014; Modic & Lea, 2011; Whitty & Buchanan, 2012). Typically, this research is conducted in a qualitative

capacity, gathering insights into the decision-making processes that led to victimisation. This methodology provides a sample that has self-evidently demonstrated susceptibility to email fraud. However, such opportunities come with a cost - in terms of the inability to retrospectively capture the psychological influences of this susceptibility, specifically the relevant contextual influences at the time of the attack. This is reliant on the recall of the participant about the exact scenario and external factors that were in place at the time they received and responded to the email. Dependent on the period of time that has passed since the incident, this can prove difficult. Additional assessments of a participant's cognitive make-up may have changed as a result of the incident, making them less representative of the individual differences between users at the point when they became a victim. Finally, this methodology is reliant on the availability of a sample of past victims, as well as a comparable control group who have not demonstrated susceptibility. Ideally this control group will have been recipients of similar incoming emails as the victim, but this is very difficult to control for and unrealistic as a method for precisely comparing response behaviour between the two groups. In line with this approach, research has also analysed the content of genuine past phishing emails, to establish linguistic patterns and persuasion techniques that may encourage a response from the recipient (e.g. Freiermuth, 2011).

Some researchers have employed simulated phishing attacks as an assessment of susceptibility, in a manner comparable with penetration testing in industry assessments. These clearly provide the most ecologically valid behavioural assessment, but are also the most ethically challenging, in that recipients may feel upset or angered at being 'tricked' into responding, without having given consent prior to the attack. Unlike industry penetration testing, simulations as part of academic research are designed to assess the influence of specific factors, rather than a baseline measure of vulnerability. In order to do this, stimuli may be designed in a specific way to emulate persuasive techniques that might be employed in a phishing email, such as authority (Guéguen & Jacob, 2002) or familiarity (Jagatic, Johnson, Jakobsson, & Menczer, 2005). Alternatively, these may include additional measures to assess influences such as security knowledge (Wright & Marett, 2010).

As with industry penetration testing, one major limitation of these simulated attacks is the inability to control or monitor the context in which an email is received. This means that little insight can be gained, directly, about the situational factors influencing response behaviour, as this is reliant on the recall of the user themselves. It also becomes difficult to assess how the effect of different persuasive techniques varies between users, as sending a multitude of target messages, which assess each of these, to the same person may raise suspicion. On the other hand, if a single target email is sent, sample sizes would need to be large enough to account for individual differences between users. This type of methodology requires some further development in order to address these issues and provide a useful assessment of response behaviour.

In order to demonstrate how experimental design and consent issues may interact with the potential validity of a research study, it is worth describing an undergraduate student research project from our lab (Mack, 2014). One of the project objectives was to empirically study the consequences of informed consent about a phishing attack. The project, which had been approved by the institutional ethics committee as well as the institutional network support team, involved a modest sample size (N=30). Half the participants signed up to a study titled "*Reasoning and judgements made in an online capacity"*. These individuals provided informed consent to take part in the study, and as such, they were informed that they would be sent simulated phishing emails in the subsequent 7 days. The other participants signed up for a study, "*A study of human-computer interaction"*. They were sent the same simulated phishing

emails, but prior to the study slot they signed up for, which was actually used to inform them of the study purpose and to then seek post-event consent for their involvement in the study.

Participants received two different emails from two bespoke email accounts created for study purposes with the knowledge of the institutional IT services. These emails came from "Lancaster" accounts (where the study was conducted), adopting a phisher's "spoofing" attack vector, and the sender was unfamiliar to participants. One email carried a warning message (account verification) and the other a competition incentive (win an iPad!). Both emails requested a reply response, though at that stage no confidential information was requested or retained. Later, only participants (both responders and non-responders) who agreed to have their data retained were retained in the study (although as it happened, no-one withdrew). The study showed that for those participants who provided informed consent, 40% showed some vulnerability in responding to one of the two phishing emails, whilst no-one responded to both. For those participants who had no prior warning via informed consent, 80% responded to at least one phishing email, and 20% responded to both. As well as revealing the extent of users' susceptibility to attack, these data suggests a dependency between the form of consent and response patterns. In other words, standard experimental design issues such as obtaining prior informed consent can have a material impact on how users will behave. This is in line with research from Parsons et al. (2013), reported above, which demonstrated similar effects as a result of gaining informed consent from participants for a role play study. Using an alternative research study as a mechanism for gaining partial consent is a concept supported by Resnik and Finn (2017). However, rather than opting in to participate, they argue that participants should be given the opportunity to opt-out of a study on email behaviour, avoiding a sample bias towards those who are more security conscious. Combined with privacy protection and appropriate debriefing, Resnik and Finn believe that simulated attacks can be conducted in an ethical manner, allowing for valuable data to be gathered.

Many ethical review boards are still reluctant to approve a simulated phishing attack without informed consent from the participant though, given the principle that participants should be fully informed before willingly volunteering to participate in research. Studies of this type also require the cooperation of relevant IT support, who need to be aware of the study and how to appropriately handle queries from individuals. It may be argued that seeking post-consent leads participants to feel under pressure to comply, given that the data has already been collected and it would otherwise be wasted. On the other hand, the study by Mack (2014) suggests that obtaining prior consent could compromise the integrity of a study. Although the evidence on the effect of priming is inconsistent to date, one alternative solution to the consent issue is to gain informed consent for a simulated phishing attack that will happen at some point in the future. With enough time between sign-up and the event, the effects of pre-warning may have dissipated, and thus the response behaviour elicited provides more naturalistic data. However, the authors are not aware of research that can pinpoint the delay period required for such a procedure to "work".

There have been a small number of studies that have used simulated attacks where the participants are naïve to the purpose of the study. Jagatic et al.'s (2005) study incorporated a simulation of a targeted phishing attack, using information gained about participants online without their consent. Following this study, Finn and Jakobsson (2007) reported that 30 out of the 1700 participants targeted complained about the research, with 7 asking for their data to be withdrawn. Although this is a small proportion of the overall sample size, disgruntlement amongst any number of participants is of concern to researchers ethically. A further example, although indirectly related, comes from a study conducted by Facebook, in which they manipulated newsfeed content to see if they could affect user's emotions in their own posts (Kramer, Guillory, & Hancock, 2014). Although there is a clause in the Terms and Conditions of

having a Facebook account that legally allows this type of research, the organisation received extensive backlash from users and industry experts for the lack of informed consent (Arthur, 2014). If we wish to understand and model real-world fraud events, we need to consider how on-going genuine research and ethical issues can be accommodated without distorting the integrity of the study itself.

Key Insights and Usability

As mentioned above, the focus of academic research into susceptibility to phishing varies in terms of its aims and the theoretical implications of these, and as such the methodologies adopted vary as well. Most prominently, research considers three potential sources of influence – the content and persuasive techniques employed within the email itself, contextual factors at the time the email is received, and individual differences between the users who are receiving the email (see Jones, Towse, & Race, 2015, for a comprehensive overview).

It is becoming more common for sophisticated phishing emails to be targeted towards a specific recipient or group of recipients, in order to make the email more believable and thus increase response likelihood. This is a process that can be automated, with data gathered across a number of online sources to maximise the plausibility of the attack (Edwards, Larson, Green, Rashid, & Baron, 2017). Jagatic et al. (2005) used publicly available information from social media to develop emails that purported to come from someone known to the recipient and found an increased response rate to these, demonstrating the influence of familiarity and social compliance on response likelihood. Research into the link between social media usage and cyber crime victimisation (through phishing, as well as other attack methods), has demonstrated though that specific types of social networking sites are more likely to increase victimisation. Specifically, knowledge exchange sites, such as LinkedIn and Flickr, where users share a greater amount of personal information in order to maximise networking opportunities, are associated with higher levels of victimisation (Saridakis, Benson, Ezingeard, & Tennakoon, 2015).

Alternative considerations of persuasive techniques have examined the influence of authority. Guéguen and Jacob (2002) again used a simulated phishing attack, targeting users who signed on to network computers in a university building being monitored by the researchers. Participants were either sent an email from a low-status individual (another student) or a professor from the university, deemed to be of a higher authoritative status, with results demonstrating greater response likelihood for the email sent from the high-authority figure. These findings are in line with theoretical perspectives on the psychology of persuasion, outlining the influence of factors such as authority, social proof, and scarcity (Cialdini, 1993). This example is similar to whaling attacks, a type of phishing outlined earlier in the chapter, in which the fraudster sends an email from the hacked or imitated account of a senior executive within an organisation, to induce a response through a purported authoritative identity. Such persuasive factors are thought to lead users to overlook cues that would otherwise have indicated the illegitimacy of an email (Langenderfer & Shimp, 2001). Further support for this notion comes from Freiermuth (2011), who's analysis of email content demonstrated the presence of a number of mechanisms intended to invoke a response from the recipient. This research emphasises consistently used techniques in Nigerian 419 scams, such as building resonance with the scammer, offering rewards, and emulating a sense of urgency.

Additional contextual factors may lead to similar oversights in terms of the cues available within an email, for example when users are distracted or overly concerned with other tasks. Yan and Gozu (2012) demonstrated this when they asked participants to make email legitimacy judgments either as quickly as possible, or to take their time over decisions. When participants spent longer assessing the emails they

demonstrated lower susceptibility. Further to this, Jones (2016) demonstrated that when participants were asked to complete a secondary verbal task (counting backwards aloud) simultaneously with an email legitimacy task, their accuracy was lower than a control and a secondary motor task condition. This study intended to emulate a scenario where users are multi-tasking whilst managing emails, typical of a daily office scenario, such as talking on the phone with a colleague or client.

An alternative, or complimentary, approach to understanding response behaviour considers whether individual differences between users may make some more susceptible than others (Williams, Beardmore, & Joinson, 2017). Factors that indicate a reliance on intuitive responses have been shown to increase susceptibility to fraud, such as self-control (Holtfreter, Reisig, Piquero, & Piquero, 2010), cognitive reflection, and inhibition (Jones et al., submitted). The influence of cognitive reflection was replicated across email legitimacy tasks, as well as an office simulation task in which participants were naïve to the true nature of the research (Jones, 2016), supporting the validity of this finding.

It is possible that there is a crossover between these different explanations of susceptibility, for example, users may be more inclined to rely on intuitive decision-making, thus demonstrating higher susceptibility in certain scenarios and in response to certain persuasive techniques that are employed. At this point though, this is an area that requires further exploration.

All of the insights highlighted here are currently limited by the unknown validity of the measures taken to assess susceptibility. It is therefore important that on-going work considers the development of an assessment tool and method that allows for control of factors being measured, whilst also ensuring maximum possible validity in measuring real world susceptibility. Below, the authors describe some potential directions that could unify approaches taken across industry and academic research in an attempt to reach this goal.

BUILDING A UNITED FRONT

Research gathered under controlled conditions in a lab setting provides valuable knowledge on how specified factors can influence perceptions of phishing emails. However, the methodologies employed in these settings mean that it is hard to know to what extent these factors relate to real world response behaviour and susceptibility. Ultimately – just because an influencing factor is significant in the lab doesn't mean that it is having the same effect in the real world.

Practical and ethical constraints in research make it difficult to assess susceptibility in a naturalistic environment. By enhancing collaborations between industry and academia, we will be one step closer to understanding how user decisions are influenced in the real world. Whilst industry approaches focus on appraising the vulnerability within an organisation, highlighting potential threats and identifying users that are more likely to respond, academic approaches take a more in depth look at why certain users respond whilst others do not. Ultimately, the combination of these approaches can allow for the development of novel techniques and effective training mechanisms to reduce susceptibility.

Psychological research has much to offer in addressing methodological and validity concerns associated with lab-based studies, but there are a number of inconsistencies still to be ironed out. For example, across a series of studies conducted by the authors, cognitive and situational influences were assessed through both an explicit judgment task, and an office simulation where email responses were monitored

without participants being aware of the study purpose (Jones, 2016). Results were partially replicated across these, demonstrating the influence of cognitive reflection. However, a number of factors (e.g. inhibition, time pressure) were not replicated, bringing into question the alignment of the two methodologies in terms of how well they are assessing susceptibility. Without further investigation, it is difficult to establish which methodology provides a better representation of real world response behaviour, and as such which influential factors should be acted upon.

One of the potential ways in which research can be harnessed to interact more closely with real-world security concerns is to focus on the development of risk mitigation strategies that incorporate research insights on user behaviour. At the present time, as discussed, behavioural models of fraud vulnerability are not well developed. For this reason, a lot of the training currently available focuses on issues such as improving the ability to differentiate between genuine and fraudulent web sites and images, based on generic visual cues (Moreno-Fernandez, Blanco, Garaizar, & Malute, 2017). Whilst there is no evidence for or against the effectiveness of such training mechanisms in an organisational setting, some empirical research has demonstrated that priming (Jones, 2016; Junger, Montoya, & Overink, 2017) and knowledge of basic security cues (Downs, Holbrook, & Cranor, 2006) alone is not enough to reduce response likelihood to phishing emails. Even if an improvement were seen shortly after training on these cues, the saliency of these in an actual phishing email during a moment of regular day-to-day behaviour is unlikely to replicate this. Bullée, Montoya Morales, Junger, and Hartel (2016) examined the effectiveness of priming in relation to telephone-based social engineering attacks. Although an improvement in detection was seen one week after the intervention, this effect was lost when participants were tested again two weeks later, and in fact susceptibility was shown to increase.

The increasing sophistication of phishing emails means that generic visual cues that users are told to look out for, such as spelling mistakes and fake email addresses (as seen on the Citizen's Advice Bureau website, 2016), are often irrelevant in many cases. The more generic phishing emails are often picked up by spam filters these days, and so focus needs to be drawn to the more advanced emails designed to trick even the most security conscious user.

By understanding the factors that influence susceptibility, training programmes could be tailored to educate users about these, and to target the most susceptible individuals within an organisation, as part of a broader cyber risk mitigation programme. Therefore, the authors would advocate a gradual transition towards more empirically grounded and theoretically inspired training techniques, which can draw from a greater body of research knowledge in the design of interventions. Moreover, the authors emphasise the importance of assessing the effectiveness of these training methods over multiple time scales and contexts. The continued advancement in the techniques used by fraudsters means that training methods must do more than tell users what cues to look out for. Training programmes must be designed to transition alongside these changes in order to maintain their effectiveness. At the most advanced level, this may mean training users to understand the underlying mechanisms and motivations behind the development of phishing emails, to help users see through the malicious intentions of a sender when they are reading emails. But at its most basic level, this might simply mean that training programmes are kept as up to date as possible with the most sophisticated techniques and persuasive mechanisms used to manipulate the user. In line with current trends in social engineering attacks, one example might be to incorporate advice and information about the unseen harm on both a personal and an organisational level that can result from an employee posting information publicly on social media.

CONCLUSION

As outlined here, it is clear that there are differences in the approaches taken to tackling email fraud between industry and academia, although both are working towards the same goal of tackling the threat posed by email fraud. Whilst industry is concerned with managing risk to protect valuable assets, with immediate solutions to address the issue and prevent future financial loss for organisations, academic research is more focused on understanding the theoretical principles underlying response behaviour in order to develop long-term solutions. Although the process may be more drawn out, given the unknowns that require examination, this more in-depth understanding will benefit all invested parties in the future. For solutions to have an on-going impact on secure behaviour, these must ensure users are able to transition their knowledge in line with the development and increased sophistication of phishing attacks.

There are clear parallels between these two approaches, both of which have advantages and disadvantages in terms of the methods currently employed. Collaborations between industry and academia are becoming more common, and the authors believe that further progression in this direction can only benefit on-going efforts to build a united front against persistent cyber security threats. Whilst industry offers access to a real-world sample that can be studied in a naturalistic environment, academia works towards the most ethical and theoretically grounded methods to harness the potential from this. It is hoped that consideration of industry impact will help academic researchers orient their research to elicit maximum benefit for industry partners, whilst also demonstrating to industry the importance of considering the impact human decision-making can have on cyber security. The transition within research settings to use more naturalistic assessments of email response behaviour will allow for the development of more effective training solutions that are relevant to real world behaviour, have a long-term effect on susceptibility, and as such can decrease the risk of victimisation for organisations and individual users.

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. However, the authors would like to acknowledge the cooperation of Lancaster University Information Systems Services for enabling and facilitating the undergraduate student project described in this chapter (Mack, 2014).

REFERENCES

Abu-Nimeh, S., Nappa, D., Wang, X., & Nair, S. (2007). A comparison of machine learning techniques for phishing detection. *Proceedings of The Anti-Phishing Working Group's Second Annual eCrime Researchers Summit*, 60-69. doi:10.1145/1299015.1299021

Arthur, C. (2014). Facebook study breached ethical guidelines, researchers say. *The Guardian*. Retrieved June 19, 2017, from: https://www.theguardian.com/technology/2014/jun/30/facebook-emotion-study-breached-ethical-guidelines-researchers-say

Barrett, N. (2003). Penetration testing and social engineering: Hacking the weakest link. *Information Security Technical Report*, *8*(4), 56–64. doi:10.1016/S1363-4127(03)00007-4

Bullée, J. H., Montoya Morales, A. L., Junger, M., & Hartel, P. H. (2016). Telephone-based social engineering attacks: An experiment testing the success and time decay of an intervention. In *Proceedings of the Singapore Cyber-Security Conference (SG-CRC) 2016* (pp. 107-114). IOS Press.

Button, M., Nicholls, C. M., Kerr, J., & Owen, R. (2014). Online frauds: Learning from victims why they fall for these scams. *Australian and New Zealand Journal of Criminology*, *47*(3), 391–408. doi:10.1177/0004865814521224

Cialdini, R. B. (1993). *Influence: The Psychology of Persuasion*. New York: Quill William Morrow.

Citizen's Advice Bureau. (2016). *Phishing – spam emails and fake websites*. Retrieved August 1, 2016, from: https://www.citizensadvice.org.uk/consumer/scams/scams/common-scams/computer-and-online-scams/phishing-spam-emails-and-fake-websites/

Downs, J. S., Holbrook, M., & 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. doi:10.1145/1143120.1143131

Downs, J. S., Holbrook, M., & Cranor, L. F. (2007). Behavioural response to phishing risk. *Proceedings of the Anti-Phishing Working Groups Second Annual eCrime Researchers Summit*, 37-44. doi:10.1145/1299015.1299019

Duke, D. (2002). Ethical hackers – can we trust them? *Network Security*, *3*, 3.

Edwards, M., Larson, R., Green, B., Rashid, A., & Baron, A. (2017). Panning for gold: Automatically analysing online social engineering attack surfaces. *Computers & Security*, *69*, 18–34. doi:10.1016/j.cose.2016.12.013

Finn, P., & Jakobsson, M. (2007). Designing ethical phishing experiments. *Technology and Society Magazine, IEEE*, *26*(1), 46–58. doi:10.1109/MTAS.2007.335565

Fischer, P., Lea, S. E., & Evans, K. M. (2013). Why do individuals respond to fraudulent scam communications and lose money? The psychological determinants of scam compliance. *Journal of Applied Social Psychology*, *43*(10), 2060–2072. doi:10.1111/jasp.12158

Freiermuth, M. R. (2011). Text, lies, and electronic bait: An analysis of email fraud and the decisions of the unsuspecting. *Discourse & Communication*, *5*(2), 123–145. doi:10.1177/1750481310395448

Garera, S., Provos, N., Chew, M., & Rubin, A. D. (2007). A framework for detection and measurement of phishing attacks. *Proceedings of the 2007 ACM Workshop on Recurring Malcode*, 1-8. doi:10.1145/1314389.1314391

Guéguen, N., & Jacob, C. (2002). Solicitation by e-mail and solicitor's status: A field study of social influence on the web. *Cyberpsychology & Behavior*, *5*(4), 377–383. doi:10.1089/109493102760275626 PMID:12216702

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.

Holtfreter, K., Reisig, M. D., Piquero, N. L., & Piquero, A. R. (2010). Low self-control and fraud offending, victimization, and their overlap. *Criminal Justice and Behavior*, *37*(2), 188–203. doi:10.1177/0093854809354977

Hong, K. W., Kelley, C. M., Tembe, R., Murphy-Hill, E., & Mayhorn, C. B. (2013, September). Keeping up with the Joneses: Assessing phishing susceptibility in an email task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 1012–1016. doi:10.1177/1541931213571226

Hudic, A., Zechner, L., Islam, S., Krieg, C., Weippl, E. R., Winkler, S., & Hable, R. (2012). Towards a unified penetration testing taxonomy. *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on Social Computing*, 811-812.

Islam, R., & Abawajy, J. (2013). A multi-tier phishing detection and filtering approach. *Journal of Network and Computer Applications*, *36*(1), 324–335. doi:10.1016/j.jnca.2012.05.009

Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2005). Social phishing. *Communications of the ACM*, *50*(10), 94–100. doi:10.1145/1290958.1290968

Jones, H., Towse, J., & Race, N. (2015). Susceptibility to email fraud: A review of psychological perspectives, data-collection methods, and ethical considerations. *International Journal of Cyber Behavior, Psychology and Learning*, *5*(3), 13–29. doi:10.4018/IJCBPL.2015070102

Jones, H. S. (2016). *What makes people click: Assessing individual differences in susceptibility to email fraud* (Unpublished doctoral thesis). Lancaster University, UK.

Jones, H. S., Towse, J., Race, N., & Harrison, T. (submitted). *Email fraud – the search for psychological markers of susceptibility*.

Junger, M., Montoya, L., & Overink, F. J. (2017). Priming and warnings are not effective to prevent social engineering attacks. *Computers in Human Behavior*, *66*, 75–87. doi:10.1016/j.chb.2016.09.012

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(24), 8788–8790. doi:10.1073/pnas.1320040111 PMID:24889601

Langenderfer, J., & Shimp, T. A. (2001). Consumer vulnerability to scams, swindles, and fraud: A new theory of visceral influences on persuasion. *Psychology and Marketing*, *18*(7), 763–783. doi:10.1002/mar.1029

Mack, S. (2014). *Reasoning and judgements made in an online capacity. An exploration of how phishing emails influence decision making strategies* (Unpublished undergraduate dissertation). Lancaster University, Lancaster, UK.

Mitnick, K. D., & Simon, W. L. (2002). *The Art of Deception*. Indianapolis, IN: Wiley Publishing, Inc.

Modic, D., & Lea, S. E. G. (2011). *How neurotic are scam victims, really? The big five and Internet scams*. Paper presented at the 2011 Conference of the International Confederation for the Advancement of Behavioral Economics and Economic Psychology, Exeter, UK.

Moreno-Fernández, M. M., Blanco, F., Garaizar, P., & Matute, H. (2017). Fishing for phishers. Improving Internet users' sensitivity to visual deception cues to prevent electronic fraud. *Computers in Human Behavior*, *69*, 421–436. doi:10.1016/j.chb.2016.12.044

Murdoch, S. J., & Sasse, M. A. (2017). Should you really phish your own employees. *NS Tech*. Retrieved September 28, 2017, from http://tech.newstatesman.com/guest-opinion/phishing-employees

Myers, S. (2007). Introduction to phishing. In M. Jakobsson & S. Myers (Eds.), *Phishing and Countermeasures* (pp. 1–29). John Wiley & Sons, Inc.

National Institute of Standards and Technology (NIST). (2012). *Guide for conducting risk assessments*. NIST Special Publication 800-30, Revision 1. Retrieved September 28, 2017, from http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-30r1.pdf

Nicholson, J., Coventry, L., & Briggs, P. (2017). Can we fight social engineering attacks by social means? Assessing social salience as a means to improve phish detection. In *Proceedings of the Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017).* Santa Clara, CA: USENIX.

Norton. (2014). *Online fraud: Phishing*. Retrieved July 12, 2014, from http://uk.norton.com/cybercrime-phishing

Parsons, K., McCormac, A., Pattinson, M., Butavicius, M., & Jerram, C. (2013). Phishing for the truth: A scenario-based study of users' behavioural response to emails. In *Proceedings of IFIP International Information Security Conference* (pp. 366-378). Berlin: Springer.

Peterson, H. (2014). Target's massive data breach originated with a single phishing email. *Business Insider*. Retrieved May 18, 2017, from: http://www.businessinsider.com/target-hack-traced-to-phishing-email-2014-2?IR=T

Picchi, A. (2014). Target breach may have started with email phishing. *CBS News*. Retrieved May 18, 2017, from: http://www.cbsnews.com/news/target-breach-may-have-started-with-email-phishing/

Resnik, D. B., & Finn, P. R. (2017). Ethics and phishing experiments. *Science and Engineering Ethics*. doi:10.1007/s11948-017-9952-9 PMID:28812222

Roberts, J. J. (2017, April 27). Exclusive: Facebook and Google were victims of $100, payment scam. *Fortune*. Retrieved May 16, 2017, from: http://fortune.com/2017/04/27/facebook- google-rimasauskas/

Salah, K., Alcaraz Calero, J. M., Zeadally, S., Al-Mulla, S., & Alzaabi, M. (2013). Using cloud computing to implement a security overlay network. *IEEE Security and Privacy*, *11*(1), 44–53.

Saridakis, G., Benson, V., Ezingeard, J., & Tennakoon, H. (2015). Individual information security, user behaviour and cyber victimisation: An empirical study of social networking users. *Technological Forecasting and Social Change*, *102*, 320–330. doi:10.1016/j.techfore.2015.08.012

Schneier, B. (2000). *Secrets & Lies: Digital Security in a Networked World*. Indianapolis, IN: Wiley Publishing Inc.

Smith, D. (2016, November 6). WikiLeaks emails: What they revealed about the Clinton campaign's mechanics. *The Guardian*. Retrieved May 16, 2017, from: https://www.theguardian.com/us- news/2016/nov/06/wikileaks-emails-hillary-clinton-campaign-john-podesta

Verizon. (2017). *2017 Data Breach Investigations Report*. Retrieved May 18, 2017, from: http://www.verizonenterprise.com/verizon-insights-lab/dbir/2017/

Whitty, M. T., & Buchanan, T. (2012). The online romance scam: A serious cybercrime. *Cyberpsychology, Behavior, and Social Networking*, *15*(3), 181–183. doi:10.1089/cyber.2011.0352 PMID:22304401

Williams, E. J., Beardmore, A., & Joinson, A. N. (2017). Individual differences in susceptibility to online influence: A theoretical review. *Computers in Human Behavior*, *72*, 412–421. doi:10.1016/j.chb.2017.03.002

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. doi:10.2753/MIS0742-1222270111

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

Yeo, J. (2013). Using penetration testing to enhance your company's security. *Computer Fraud & Security*, *4*(4), 17–20. doi:10.1016/S1361-3723(13)70039-3

Yuhas, A. (2017, March 22). Lithuanian man's phishing tricked US tech companies into wiring over $100m. *The Guardian*. Retrieved May 16, 2017, from: https://www.theguardian.com/technology/2017/mar/22/phishing-scam-us-tech-companies-tricked-100-million-lithuanian-man

KEY TERMS AND DEFINITIONS

Collaboration: Bringing together alternative approaches and perspectives to develop novel ideas and solutions.

Email Fraud: The use of email as a means of deceiving users for personal or financial gain.

Individual Differences: Variations in human behaviour as a result of a specific trait or traits.

Informed Consent: The participant agrees to participate in research, with a full understanding of the research purpose and tasks involved.

Penetration Testing: Assessment of vulnerabilities within an organisation's computer system or network, including human users.

Persuasive Techniques: Factors that can be manipulated to influence human behaviour.

Susceptibility: A likelihood to be more easily affected or influenced by a specific thing.