



RHODES UNIVERSITY
Where leaders learn

**A PERSONALITY-BASED BEHAVIOURAL MODEL:
SUSCEPTIBILITY TO PHISHING ON SOCIAL
NETWORKING SITES**

by

Edwin Donald Frauenstein

**A PERSONALITY-BASED BEHAVIOURAL MODEL:
SUSCEPTIBILITY TO PHISHING ON SOCIAL
NETWORKING SITES**

by

Edwin Donald Frauenstein

18F9034

 <https://orcid.org/0000-0002-9658-587X>

THESIS

submitted in fulfilment of the requirements for the degree

DOCTOR OF PHILOSOPHY

in

INFORMATION SYSTEMS

in the

FACULTY OF COMMERCE

of

RHODES UNIVERSITY

Supervisor: Prof. Stephen Flowerday

August 2021

ABSTRACT

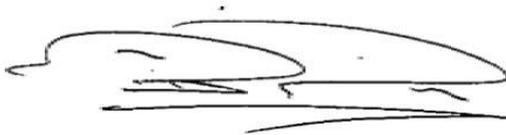
The worldwide popularity of social networking sites (SNSs) and the technical features they offer users have created many opportunities for malicious individuals to exploit the behavioral tendencies of their users via social engineering tactics. The self-representation and social interactions on SNSs encourage users to reveal their personalities in a way which characterises their behaviour. Frequent engagement on SNSs may also reinforce the performance of certain activities, such as sharing and clicking on links, at a “habitual” level on these sites. Subsequently, this may also influence users to overlook phishing posts and messages on SNSs and thus not apply sufficient cognitive effort in their decision-making. As users do not expect phishing threats on these sites, they may become accustomed to behaving in this manner which may consequently put them at risk of such attacks. Using an online survey, primary data was collected from 215 final-year undergraduate students. Employing structural equation modelling techniques, the associations between the Big Five personality traits, habits and information processing were examined with the aim to identify users susceptible to phishing on SNSs. Moreover, other behavioural factors such as social norms, computer self-efficacy and perceived risk were examined in terms of their influence on phishing susceptibility. The results of the analysis revealed the following key findings: 1) users with the personality traits of *extraversion*, *agreeableness* and *neuroticism* are more likely to perform habitual behaviour, while *conscientious* users are least likely; 2) users who perform certain behaviours out of *habit* are directly susceptible to phishing attacks; 3) users who behave out of habit are likely to apply a *heuristic* mode of processing and are therefore more susceptible to phishing attacks on SNSs than those who apply *systematic processing*; 4) users with higher *computer self-efficacy* are less susceptible to phishing; and 5) users who are influenced by *social norms* are at greater risk of phishing. This study makes a contribution to scholarship and to practice, as it is the first empirical study to investigate, in one comprehensive model, the relationship between personality traits, habit and their effect on information processing which may influence susceptibility to phishing on SNSs. The findings of this study may assist organisations in the customisation of an individual anti-phishing training programme to target specific dispositional factors in vulnerable users. By using a similar instrument to the one used in this study, pre-assessments could determine and classify certain risk profiles that make users vulnerable to phishing attacks.

DECLARATION

I, Edwin Donald Frauenstein, hereby declare that:

- The work presented in this thesis is my own work.
- All sources used or referred to have been documented and acknowledged.
- This thesis has not previously been submitted in full or partial fulfilment of the requirements for an equivalent or higher qualification.
- I am fully aware of the Rhodes University's policy on plagiarism and have taken every precaution to adhere to this regulation.
- Ethical approval for this study was obtained from the university's Ethics Committee with certificate reference number: FLO071SFRA01.

Signed:



Date: 16 August 2021

ACKNOWLEDGEMENTS

My sincerest gratitude to the following people and institutions who contributed to the successful completion of this study:

- Prof. Stephen Flowerday, my research supervisor, for his expertise, guidance and patience which helped me complete this study.
- The Govan Mbeki Research and Development Centre (GMRDC) at the University of Fort Hare in East London which partly supported my study fees.
- Walter Sisulu University (WSU) which financially supported my study fees throughout the duration of this study.
- Prof. Syden Mishi, from Nelson Mandela University (NMU), for assisting with the statistical data analyses.
- Ms. Alexa Barnby for language editing this dissertation.

GLOSSARY

Term	Description
Information processing	A cognitive process in which people typically evaluate messages in one of two modes: systematically or heuristically (Trumbo, 2006). Systematic processing requires more effort and more attention to the characteristics and content of the information and considers information based on its facts and comparisons to prior knowledge (Bohner et al., 1995). Heuristic processing requires less effort and focuses on simple cues embedded (i.e. heuristic cues) in the context of the message (Bohner et al., 1995).
Dispositional factors	Dispositional factors, also known as internal factors, are individual characteristics that can influence the behaviour and actions in an individual. These factors, which are relatively stable over time, include personality traits, Dark Triad traits, propensity to trust, temperament, cognitive style, self-esteem and other traits (Johnston et al., 2016).
The Big Five personality traits	The Big Five is a widely accepted personality theory which categorises an individual's personality into five traits. These traits, sometimes referred by the acronym CANOE or OCEAN, are Openness, Conscientiousness, Agreeableness, Extraversion and Neuroticism (John & Srivastava, 1999).
Social norms	A social norm is a type of subjective norm. Social norms are typically defined as “rules and standards that are understood by members of a group, and that guide or constrain social behaviors without the force of law” (Cialdini & Trost, 1998). Social norms develop and evolve as a result of the interaction between individuals in social groups (Cialdini & Trost, 1998) and users may adjust their behaviour based on the perceived expectations of others.

Social engineering	Social engineering can be described as the “art” of deceiving people in order to gain valuable information from them, or to persuade them to perform an action(s) that will ultimately benefit the attacker in some way (Mitnick & Simon, 2002).
Computer self-efficacy	Refers to the user's judgement of or belief in their capability to use computers to achieve a particular purpose and is influenced by computer knowledge and prior computer experience (Potosky, 2002).
Habit	The “learned sequences of acts that have become automatic responses to specific cues” (Verplanken & Aarts, 1999). In the context of this study, habit refers to the extent to which social media users automatically click and share a link/posts that are requested by their friends.
Perceived risk	A dispositional factor defined as the “subjective belief that there is some probability of suffering a loss in pursuit of the desired outcome” (Pavlou, 2003).
Structural equation modelling (SEM)	A multivariate statistical analysis technique that is used to analyse structural relationships. This technique uses a combination of confirmatory factor analysis and path analysis to analyse the structural relationship between measured variables and latent constructs (Fan et al., 2016).

TABLE OF CONTENTS

ABSTRACT	iii
DECLARATION	iv
ACKNOWLEDGEMENTS	v
GLOSSARY	vi
LIST OF FIGURES	xiii
LIST OF TABLES.....	xiv
1 INTRODUCTION.....	1
1.1 Introduction.....	1
1.2 Background.....	1
1.2.1 Approaches to mitigate phishing	4
1.3 Background to the problem.....	8
1.4 Problem statement	14
1.5 Research questions	14
1.6 Research objective.....	17
1.7 Research design and methodological approach.....	18
1.8 Contribution	21
1.9 Ethical considerations	22
1.10 Thesis layout	22
2 SOCIAL ENGINEERING AND THE HUMAN FACTOR	25
2.1 Introduction.....	25
2.2 Social engineering	26
2.2.1 Taxonomies, models and frameworks	27
2.2.2 Persuasion principles	31
2.2.3 User behaviour toward warnings and security indicators	33
2.2.4 Social networking sites.....	35
2.2.5 Ethics in social engineering experiments	36
2.3 Social engineering approaches	37
2.3.1 Physical approach	38
2.3.2 Social approach.....	38
2.3.3 Reverse social engineering.....	38
2.3.4 Technical approach.....	39
2.3.5 Socio-technical approach	39

2.4	Persuasion techniques on social network sites	40
2.4.1	Authority	41
2.4.2	Conformity or social proof.....	43
2.4.3	Reciprocity	44
2.4.4	Commitment or consistency.....	45
2.4.5	Scarcity	46
2.4.6	Liking.....	48
2.4.7	Other persuasion techniques and threats on social network sites	52
2.5	Habit.....	54
2.5.1	Antecedents of habit	56
2.6	Perceived risk.....	59
2.7	Computer self-efficacy.....	61
2.8	Social norms	62
2.9	Summary	63
3	PERSONALITY TRAITS AND THEIR INFLUENCE ON HUMAN BEHAVIOUR	65
3.1	Introduction.....	65
3.2	Personality traits	66
3.2.1	Brief history of the Five Factor Model.....	67
3.2.2	Description of the Big Five personality traits	69
3.3	Background on personality traits relating to information security behaviour	71
3.3.1	Influence of personality traits on employee security behaviour	71
3.3.2	Influence of personality traits on internet dependency.....	72
3.3.3	Influence of persuasion strategies on personality traits.....	73
3.3.4	Influence of personality traits and gender on phishing detection	76
3.3.5	Influence of personality traits on social media user behaviour	77
3.3.6	Influence of personality traits on user susceptibility to social engineering and phishing.....	79
3.4	Summary	80
4	INFORMATION PROCESSING AND ITS INFLUENCE ON USERS' ABILITY TO DETECT PHISHING	81
4.1	Introduction.....	81
4.2	Psychological frameworks used in phishing research.....	82
4.3	The Heuristic-Systematic Model	85
4.3.1	Systematic processing.....	86

4.3.2	Heuristic processing.....	87
4.3.3	Influences and barriers to information processing.....	87
4.4	Background on the Heuristic-Systematic Model in the context of phishing	93
4.5	Summary	97
5	HYPOTHESIS DEVELOPMENT AND MODEL	99
5.1	Introduction.....	99
5.2	Theory development and hypotheses	99
5.2.1	Influence of the Big Five personality traits on habit	100
5.2.2	Influence of habit on information processing	103
5.2.3	Influence of habit on phishing susceptibility	104
5.2.4	Influence of information processing on phishing susceptibility	104
5.2.5	Influence of social norms on phishing susceptibility	105
5.2.6	Influence of computer self-efficacy on phishing susceptibility	105
5.2.7	Influence of perceived risk on phishing susceptibility	106
5.3	Demography.....	108
5.4	Summary	108
6	RESEARCH DESIGN AND METHODOLOGY	109
6.1	Introduction.....	109
6.2	Research process.....	109
6.3	Research design	111
6.4	Philosophical paradigms	112
6.4.1	Positivism	114
6.4.2	Post-positivism.....	114
6.4.3	Interpretivism	116
6.4.4	Critical realism	116
6.4.5	Pragmatism.....	117
6.5	Paradigm adopted in this study.....	117
6.6	Research approach.....	117
6.7	Literature review process	120
6.8	Population and sample	124
6.9	Variable descriptions and measures	125
6.9.1	Personality traits.....	125
6.9.2	Habit	126
6.9.3	Information processing.....	126

6.9.4	Phishing susceptibility	129
6.9.5	Social norms	130
6.9.6	Computer self-efficacy	130
6.9.7	Perceived risk	131
6.9.8	Demography	132
6.10	Data analysis	132
6.10.1	SEM approaches and techniques	133
6.10.2	Reliability	134
6.10.3	Validity	135
6.11	Ethical considerations.....	135
6.12	Summary	137
7	RESULTS	138
7.1	Introduction.....	138
7.2	Results of the pilot study.....	138
7.2.1	Sample size and demography	138
7.3	Reliability assessment	139
7.4	Results of the main study	140
7.4.1	Sample size and demography	140
7.4.2	Reliability and validity assessment.....	140
7.4.3	Univariate analysis.....	144
7.5	Measurement model evaluation.....	149
7.5.1	Common method variance	152
7.6	Structural model evaluation	153
7.6.1	Path analysis and hypotheses outcomes	154
7.6.2	Evaluation of model fit.....	158
7.7	Summary	159
8	DISCUSSION	160
8.1	Introduction.....	160
8.2	The extent to which the Big Five personality traits influence habit.....	160
8.2.1	Extraversion and its influence on Habit	161
8.2.2	Agreeableness and its influence on Habit	162
8.2.3	Conscientiousness and its influence on Habit.....	162
8.2.4	Neuroticism and its influence on Habit	163
8.2.5	Openness and its influence on Habit	164

8.3	The extent to which Habit influences Information processing.....	164
8.4	The extent to which Information processing influences phishing susceptibility	165
8.5	The influence of dispositional factors on phishing susceptibility	166
8.6	Summary	168
9	CONCLUSION.....	170
9.1	Introduction.....	170
9.2	Chapter summaries	170
9.3	Research questions revisited.....	175
9.4	Contributions of the study.....	178
9.4.1	Contribution to theory	179
9.4.2	Contribution to practice.....	180
9.5	Limitations and future research directions.....	182
9.5.1	Sample size and demography.....	182
9.5.2	Context	183
9.5.3	Experimental design	183
9.5.4	Persuasion principles	184
9.6	Publications emanating from the study	185
9.7	Conclusion	186
	REFERENCES	188
	APPENDIX A: CONSTRUCT DESCRIPTIVE STATISTICS	221
	APPENDIX B: MEASUREMENT ITEMS OF THE STUDY	224
	APPENDIX C: INFORMED CONSENT FORM	228

LIST OF FIGURES

Figure 1.1: Phishing email impersonating a banking institution	9
Figure 1.2: Phishing post on Facebook which leads to spoofed website	11
Figure 1.3: The seven stages involved in the application of the scientific method (adapted from Zikmund et al., 2013)	19
Figure 1.4: Chapter layout of the study	24
Figure 2.1: Social engineering taxonomy classifying attack characteristics and attack scenarios (Krombholz et al., 2015).....	28
Figure 2.2: Combination of technical approaches in a phishing attack with intersections represented by the nodes (Chiew et al., 2018).....	29
Figure 2.3: An ontological model of a social engineering attack (Mouton et al., 2014)	30
Figure 2.4: Authority principle applied in Facebook	42
Figure 2.5: Social proof principle applied in Facebook	43
Figure 2.6: Reciprocity principle applied in Facebook Messenger	45
Figure 2.7: Scarcity principle applied in Facebook	47
Figure 2.8: Legitimate post using the scarcity principle applied in Facebook	48
Figure 2.9: Curiosity principle applied in Facebook Messenger	50
Figure 2.10: Curiosity principle applied in Facebook	51
Figure 4.1: Features in an email typically identified by users	89
Figure 4.2: Comparison of a Facebook phishing post versus legitimate post	91
Figure 5.1: Proposed model	107
Figure 6.1: The research onion (adapted from Saunders et al., 2016, p. 124)	112
Figure 6.2: Typology of ontological assumptions on a continuum of paradigms (adapted from Morgan & Smircich, 1980, p. 492).....	113
Figure 6.3: Deductive reasoning (adapted from Trochim, 2006)	119
Figure 6.4: The funnel method used to structure a literature review (adapted from Hofstee, 2006, p. 96)	123
Figure 6.5: Phishing email purportedly originating from Facebook	130
Figure 7.1: The Personality-Habit-Information Processing (PHIP) model.....	157

LIST OF TABLES

Table 1.1: Constructs identified for the study.....	20
Table 2.1: Principles for influencing users into performing actions (adapted from Ferreira et al., 2015).....	31
Table 3.1: Summary of Big Five personality trait characteristics (adapted from Costa & McCrae, 1992b; John & Srivastava, 1999; Zhang, 2006).....	70
Table 6.1: Categories of literature work in the identification and screening process	122
Table 6.2: Description of stimuli used to test information processing.....	128
Table 7.1: Summary of the results of the reliability tests for the pilot study.....	139
Table 7.2: Results of the reliability tests for personality traits	141
Table 7.3: Results of the validity test for Extraversion.....	142
Table 7.4: Results of the validity test for Agreeableness.....	142
Table 7.5: Results of the validity test for Conscientiousness	142
Table 7.6: Results of the validity test for Neuroticism.....	142
Table 7.7: Results of the validity test for Openness	143
Table 7.8: Results of the reliability tests for the Information processing construct	143
Table 7.9: Summary of the results of the reliability tests for the main study	144
Table 7.10: Distribution of the items related to the construct Habit	145
Table 7.11: Distribution of the items related to the construct Social norms.....	146
Table 7.12: Distribution of the items related to the construct Computer self-efficacy	147
Table 7.13: Distribution of the items related to the construct Perceived risk	148
Table 7.14: Univariate results of construct Phishing susceptibility	148
Table 7.15: Descriptive statistics and correlations	151
Table 7.16: Path analysis and hypotheses outcomes	156
Table 7.17: Fit indices of the model.....	159

1

INTRODUCTION

1.1 Introduction

Popular social networking sites (SNSs), such as Facebook, provide a means for users to communicate with friends, family, colleagues, and customers and to share their views, thoughts and opinions on various topics of interest. Users can also follow or subscribe to profiles, pages and groups. It is this same environment which can also be used by phishers to launch attacks on individuals by leveraging on certain behavioural factors of these users. It is these behavioural factors which presents the main underlying problem of this study, which warrants further exploration in this area. This chapter introduces the background to the problem and the approaches used to combat phishing. The chapter also puts forward the study's main objective and research questions, and moreover the methodological approach that was followed.

1.2 Background

The proliferation of Web 2.0 technologies has introduced new forms of internet communication, including instant messengers (IM), video conferencing, cloud services, blogs, really simple syndication (RSS), wikis, podcasting and social networking sites. This has led to a shift where information is user-generated with considerably more interaction and collaboration amongst users (Grabner-Kräuter, 2009). In particular, the explosion of SNSs, with an estimated 3.6 billion people using SNSs worldwide (Statista, 2020), has given rise to a virtual community that fosters a culture of information sharing among its members. Some popular online SNSs include Facebook, Twitter, Instagram, LinkedIn, SnapChat, Pinterest, TikTok, YouTube, ResearchGate, Xing and many more. The most popular SNS is Facebook which currently has an estimated 2.7 billion active monthly users (Statista, 2020). SNSs are

not restricted by any geographical boundaries and attract members from different cultures and backgrounds, each possessing their own religious beliefs, ethnicity, education, political views and social class. SNSs encourage the members of their online community to create a public profile and share their interests, ideas, photos, music and videos with other registered users. By design, this social interaction satisfies an innate need to belong (James et al., 2017) and a desire to be part of a community (Cheung et al., 2011).

Apart from receiving information primarily generated by its users, some SNSs can receive information externally from various outlets such as news blogs, sponsored pages and the like, allowing users to share this information with others and thereby act as “opinion leaders”, encouraging discussion between multiple parties and potentially increasing their involvement in current events (Oeldorf-Hirsch & Sundar, 2015). As a result, SNS users are more likely to receive current news and events on SNSs before viewing them on other mainstream media outlets. This poses concerns, as users will become accustomed to accepting this information without further validation because either the users trust the sources sharing the information, or it is considered too time-consuming, or they are not capable of doing so. To elaborate, studies have shown that competent social media users (i.e. social media self-efficacy) consider SNSs to be a trustworthy source of information (Hocevar et al., 2014). Moreover, users with a particular political ideology and political interest tend to perceive SNSs as a credible and trusted source of information (Johnson & Kaye, 2014). However, SNSs are increasingly being used as a channel for the diffusion of misinformation and hoaxes (Tacchini et al., 2017). On this subject, it was found that during the 2016 United States presidential elections, fake profiles and bots were used on Twitter to spread fake news in the form of millions of tweets, thereby influencing the views of some of the public during the election process (Bovet & Makse, 2019). This emphasises the powerful influence SNSs has on individuals as well as the public at large. On the positive side, SNSs are also becoming accepted in the workplace, as a study by Song et al. (2019) found that certain work-oriented and socialization-oriented social media apps complemented team and employee performance in the workplace. As SNSs are popular and have afforded users these opportunities, this same environment give social engineers and phishers an opportunity to influence users into actions that allow them to obtain our personal information through phishing attacks.

The Anti-Phishing Working Group, a global consortium, defines phishing as “a crime employing both *social engineering* and *technical subterfuge* to steal consumers’ personal

identity data and financial account credentials” (APWG, 2021). Following a comprehensive literature study, designed to achieve a consensual definition for the term “phishing”, Lastdrager (2014) defined it as “a scalable act of deception whereby impersonation is used to obtain information from a target”. Phishing can be viewed as a confidence trick whereby users are lured by social engineering (SE) techniques to perform certain actions such as clicking and sharing of links, downloading attachments, and logging on spoofed websites. SE uses “influence and persuasion to deceive people by convincing them that the social engineer is someone he is not, or by manipulation” (Mitnick & Simon, 2002). In this regard, persuasion strategies are key to the success of phishing and are discussed in further detail in Section 2.4. Phishing is regarded as a type of internet fraud typically carried out through email with the source usually impersonating a financial institution or individual recognisable by the target (Blythe et al., 2011). The user is persuaded, by a convincing story described in the email message, to click on a hyperlink or a malicious file contained in the email. The link subsequently diverts the user to an external phoney webpage, unbeknown to the user, imitating the original website. The user submits their login information on the spoofed website which is captured by the phisher and subsequently used to carry out their criminal objectives (Vishwanath, 2015a). Phishing emails can also include attachments, which users may unknowingly install, that contain malware that data-mines the victim’s computer for usernames, passwords and credit card information (Harrison et al., 2015).

The term “phishing” was coined in January 1996 when attempts were made by hackers to steal passwords from the accounts of America Online (AOL) users (Ollmann, 2002). Since then, annual reports released by various government agencies, cyber security companies (e.g. Verizon, Sophos) and IT security analysts continue to emphasise the growth and impact that phishing has on numerous industries and their customers (APWG, 2021), reinforcing its position as one of the most pertinent security threat agents facing organisations today (Furnell et al., 2019). Phishing attacks continue to affect millions of Internet users, causing substantial monetary losses for both organisations and individuals (Chen et al. 2021). In 2019, phishing was reported as the fifth most common primary cause of security incidents, and ranked as the main cause of data breaches with the highest success rate of any threat vector (Verizon, 2020). This outlook did not improve as, in 2020, 85% of phishing attacks were conducted on various channels such as instant messaging, gaming, social and productivity apps, while also seeing an added rise in smishing and vishing (Verizon, 2020). As organisations have enabled greater

remote working during the Covid-19 period, there has been more exposure and reliance on internet apps and services. Phishers have taken advantage of this and the public's fear and uncertainty over the coronavirus (Covid-19) global pandemic. According to security firm Barracuda Networks (Muncaster, 2020), since the start of January 2020, Covid-19-related email phishing attacks have seen a steady increase with a sudden surge of 667% by the end of February. The Covid-19-themed phishing attacks took the form of scams (54%), brand impersonation (34%), blackmail (11%) and business emails (1%). In April 2020, Google blocked more than 18 million Covid-19-related emails consisting of malware and phishing and 240 million Covid-19-related daily spam messages (Kumaran & Lugani, 2020). Cyberattacks referencing Covid-19 became so rampant that it led to a collective effort of more than 4000 information security analysts who formed the Covid-19 Cyber Threat Coalition (CCTC) in a show of defiance (Sophos, 2021). Covid-19-related scams have cost Americans more than \$160 million since the start of 2020 (Edward, 2020). These reports emphasise the ever-changing approaches used by phishers that take advantage of current events (Mansfield-Devine, 2018), and moreover suggest that current approaches are inadequate for mitigating this threat. Appropriately, the next subsection discusses the various approaches used to combat phishing.

1.2.1 Approaches to mitigate phishing

The growing body of research has focused mainly on two approaches to protect users from phishing, namely a technological approach and a behavioural approach (Vishwanath et al., 2011). For several years, countermeasures used to mitigate phishing have largely focused on technological controls which have continued to make gains in phishing prevention for end users. According to Jensen et al. (2017), organisations often rely on automated removal or quarantine of phishing messages and corresponding websites and automated warning mechanisms that notify individuals when individuals encounter a suspicious message or website. Anti-virus programs, often seen as a solution to most threats, do not offer adequate protection against phishing (Mansfield-Devine, 2018). This has been elaborated by Furnell et al. (2019) who state that anti-phishing technology solutions can be an effective approach to block or blacklist malicious messages, using domain and IP address blocking, from "known" sources, however they do not offer the same protection if messages were received from unknown and multiple sources. Moreover, organisations have found a gateway-based single-pass inspection model to be inadequate against evolving phishing threats (Jang-Jaccard &

Nepal, 2014). As such, organisations are moving in the direction of a layered security model that employs continuous email monitoring and detection (Jang-Jaccard & Nepal, 2014).

From a scholarly perspective, researchers have focused on the testing and measurement of a variety of anti-phishing detection tools such as web browser toolbar add-ons (Marchal et al., 2017; Zhang, Hong et al., 2007), email detection filters (Fette et al., 2007), and URL detection (Abu-Nimeh et al., 2007; Garera et al., 2007; Wenyin et al., 2005; Zhang, Egelman et al., 2007). Purkait (2015) evaluated the effectiveness of popular web browsers, third-party phishing toolbar add-ons and anti-virus programs in terms of their capability to detect spoofed websites. Fire et al. (2014) focused predominately on a wide array of software-based solutions to combat a variety of threat agents on SNSs. Of these technological controls, from an end-user perspective, web browser plug-ins have for decades been the most popular in the fight against phishing (Raffetseder et al., 2007). For example, Dewan and Kumaraguru (2017) developed Facebook Inspector (FBI), a free browser plug-in for both Google Chrome and Mozilla Firefox web browsers, which can be used to identify malicious content on Facebook in real time. Similarly, Volkamer et al. (2016) created the Tooltip-powered Phish Email Detection (TOrPeDO), a browser add-on for the Thunderbird browser, which can detect suspicious links embedded in emails, provide real-time tooltips and display the obfuscated URL. Researchers have also developed software tools to educate users on identifying and protecting themselves against phishing, including online games such as Anti-Phishing Phil (Sheng et al., 2007), email-based systems such as PhishGuru (Kumaraguru et al., 2010) and game-based smartphone apps such as NoPhish (Canova et al., 2014).

As technical measures might not always be effective, this necessitates the need for users to have the knowledge and skills to identify phishing (Verkijika, 2019). Because much of the responsibility for detecting phishing activities still remains with the end user, scholars have also investigated whether users are able to correctly interpret security warnings or whether they may simply choose to ignore them. In this regard, prior research has focused on user behaviour by investigating their responses in how they interpret web browser warnings (Dhamija et al., 2006; Downs et al., 2006; Egelman et al., 2008; Petelka et al., 2019; Wu et al., 2006). While, to a certain extent, these technological measures can help mitigate phishing, relying solely on technical measures has been widely criticised as insufficient because phishing techniques evolve as technology evolves (Mansfield-Devine, 2018; Vishwanath et al., 2011). Generally, cyber threats attempt to exploit software flaws in computer systems to install

malware (Jang-Jaccard & Nepal, 2014; Kesan & Hayes, 2016). This, coupled with users' reliance on technology to "do the job", has placed too much reliance on technology controls, overshadowing the need for users to also play a role in the process. As result, users may become lax in their security behaviour and not be alert to phishing.

While having a trusting disposition is generally a good characteristic for people to possess, as it facilitates communication (Levine, 2014), such qualities appear to be frowned on in the information security domain and pointed out as a "weakness" that requires fixing (Schneier, 2016). The "blame-the-victim" approach has resulted in scholars frequently citing humans as the "weakest link" in the security chain (Johnston et al., 2016; West et al., 2009) and, as such, has led researchers to turn their attention to exploring behavioural vulnerabilities and their interrelationships and to develop interventions to address these vulnerabilities (Briggs et al., 2017). As a result, research efforts were directed towards educating the human element to change their current behaviour (Burns et al., 2013; Kirlappos & Sasse, 2012).

From an *organisation* perspective, to address behavioural vulnerabilities in particular knowledge, attitudes and behaviour, organisations typically focus on creating employee security education, training and awareness (SETA) programmes (Warkentin et al., 2012). To address phishing, SETA interventions are typically aimed at helping users identify particular cues employed in phishing emails that can help reveal them to be deceptive (Jampen et al., 2020; Vishwanath et al., 2011). As such, SETA programmes often follow a "rule-based" training approach, wherein trainers incorporate rules to help individuals to identify certain cues incorporated in phishing attacks and to take protective action (Jensen et al., 2017). Some researchers have proposed SETA frameworks with the aim of creating an organisational information security culture (Da Veiga & Eloff, 2010; Tsohou et al., 2015). In the context of SNSs, Facebook offers no phishing countermeasures to protect its users, apart from general security advice through their online Help Centre. Under the section "What steps can I take to protect myself from phishing on Facebook?" users are informed to: "Look out for suspicious emails or messages, Don't click suspicious links, Don't respond to these emails, and Get alerts for unrecognized logins" (Facebook, 2021). Coincidentally, exactly the same advice is provided for Instagram users. To a certain extent, it appears that these interventions have not yielded the desired results, as studies have shown that users who consider themselves to be aware of phishing have not demonstrated this in their behaviour (Oliveira et al., 2017).

Employees have more workload and expectations than before and, as such, are prone to making mistakes and being deceived (Bhardwaj et al., 2020). Moreover, some are not motivated and perceive training as a secondary task outside of their normal work (Schuetz et al., 2016). As SETA programmes are usually conducted once-off, ongoing and integrated programmes are needed to have a longer-lasting effect on the recipient's awareness, especially given the ever-changing nature of the threat landscape (Jampen et al., 2020; Mansfield-Devine, 2018). Other measures organisations employ include the enforcement of security policies in the belief that employees will comply, as there are implications in the form of penalties for any disobedience (Colwill, 2009; Herath & Rao, 2009; Warkentin et al., 2012). In addition, technical controls also do not prevent employees from violating information security policies (Johnston et al., 2016) and users are quite willing to accept risk in return for convenience (Vance et al., 2013; Workman, 2008). On the other hand, motivational factors such as rewards have also been used to improve compliance (Bulgurcu et al., 2010). While the aforementioned technical and educational approaches and methods have, to a certain extent, been shown to be effective, it takes just one careless employee clicking on a malicious phishing link to potentially put an entire organisation at risk (Kesan & Hayes, 2016). In view of this, other behavioural factors that can hinder these efforts have warranted further exploration of this area.

Security scholars have identified, used or adapted popular behavioural theories pertaining to attitudinal change such as the theory of reasoned action (TRA), the theory of planned behaviour (TPB), protection motivation theory (PMT), technology acceptance model (TAM), social learning theory (SLT), social cognitive theory (SCT) and general deterrence theory (GDT) to gain new insights into the behavioural problems of users (Bulgurcu et al., 2010; Bullée et al., 2015; Kearney & Kruger, 2016; Ophoff & Robinson, 2014; Williams et al., 2018). However, as pointed out by Kearney and Kruger (2016), behavioural theories on their own do not provide adequate solutions to adverse behaviour. Because individuals each possess their own unique set of vulnerabilities, addressing these by means of a one-size-fits-all approach also has limitations. As such, this has prompted many researchers to explore individual differences in characteristics and to examine their influence on user susceptibility to SE and phishing (Albladi & Weir, 2018; Alseadoon et al., 2015; Goel et al., 2017; Kaptein et al., 2009; Mayhorn et al., 2015; Moody et al., 2017; Vishwanath et al., 2018; Workman, 2007, 2008; Wright & Marett, 2010). In particular, prior literature has shown that certain people with personality traits exhibit behaviours that can make some users more vulnerable to phishing

than others (Cho et al., 2016; Cusack & Adedokun, 2018; Halevi et al., 2013). This is to be expected, as prior research has shown that different personalities exhibit certain internet behaviours (Amichai-Hamburger, 2002). Gratification can also result from performing certain behaviours on SNSs (Quan-Haase & Young, 2010). In this regard, as users frequently engage on SNSs, this could lead to the formation of certain habits which might make some users vulnerable to deception (Vishwanath, 2015b). Prior literature has also shown that the mode in which users cognitively process information could also put them at risk of phishing attacks (Harrison et al., 2015; Luo et al., 2013; Seazzu, 2013; Valecha et al., 2015; Xu & Zhang, 2012). In this regard, certain habits may potentially influence the mode in which the user will cognitively process the information they receive on these sites. All of the aforementioned aspects highlights the susceptibility of social media users to phishing. This is discussed next.

1.3 Background to the problem

The ubiquitous nature of SNSs has increased online interactions and communications in the personal and business environment, thereby creating new methods and opportunities for SE attacks (Krombholz et al., 2015). This has also simultaneously given rise to research aimed at redefining the taxonomies associated with social engineering and phishing (Chiew et al., 2018; Heartfield & Loukas, 2015). Vishwanath (2017) alludes to email-based phishing attacks as simple and one-dimensional, as they typically involve a one-step attack process. Accordingly, training and awareness interventions to mitigate phishing typically focus on making users aware of the following: 1) to not trust emails that do not specifically mention the recipient by name, 2) to look out for spelling and grammatical errors and the domain name, 3) to understand that personal or sensitive information (i.e. account numbers, passwords) should not be given out via email or phone, 4) to check whether the website connection is secure by looking out for Hypertext Transfer Protocol Secure (HTTPS) and the padlock icon in the web address bar, 5) to hover the cursor over a suspicious link to see where it is directed to, and 6) not click on links or attachments originating from unknown sources. In addition, Binks (2019) puts forward the most obvious point (but seemingly not so obvious to the victim), that is, when an offer is too good to be true, it almost certainly is a scam. However, some of these areas have their own limitations. For example, Schneier (2016) states that educating users not to click on links is futile, as for decades users have been trained to click on links and this is necessary for navigating websites. Vishwanath et al. (2018) state that education and training may be ineffective when users' online habits determine their likelihood of being deceived. Moreover,

educating users to positively identify legitimate websites by looking out for HTTPS can be a pointless exercise. Phishers are aware of this and thus more than half of all phishing sites (as of end of first quarter of 2019) employ the use of HTTPS (PhishLabs, 2019). This point is substantiated in Figure 1.1, which shows an actual phishing email purportedly impersonating a well-known bank and displays most of the characteristics of a legitimate email, thereby making it almost impossible for one to easily recognise it as phishing. All of the above precautions are also predicated on the premise that the user's intention is to look out for these characteristics.



Figure 1.1: Phishing email impersonating a banking institution

To elaborate, the email in Figure 1.1 shows that it originates from inContact@fnb.co.za which is a legitimate email address for receiving notifications and which members associated with this bank are familiar with. The hyperlink begins with HTTPS and the email incorporates the original corporate logos associated with the bank. Users are educated to be mindful of certain organisations that are frequently impersonated, particularly financial institutions, government agencies and e-commerce organisations. A study conducted by Frauenstein (2018) found that

participants, despite receiving no training on phishing, were more suspicious of phishing emails impersonating financial institutions than of emails impersonating Facebook. The email in Figure 1.1 also uses persuasion, incorporating fear and showing that an unauthorised transaction occurred on the account. As such, with regard to Figure 1.1, suspicion, stemming from prior knowledge, about banks being victims to phishing might be the only factor that could prevent an individual from falling victim to the email.

As SNSs promote social relationships between members, phishers have found that the platform presents them with opportunities, features and techniques with which to exploit the behavioural vulnerabilities of SNS members (Algarni et al., 2017). In contrast to email-based phishing, phishing attacks conducted on SNSs are multistaged, with users receiving friend requests that are often followed by messages (Vishwanath, 2015b). More recently, Vishwanath (2017) has described these attacks as involving a two-level attack process. SNSs have fewer limitations on who can be impersonated, as an attacker can impersonate a legitimate user, page or group. Much like we typically judge the trustworthiness of people based on their physical appearance and other cues, users react similarly to SNS posts based on their appearance and are further biased by persuasion principles. Cialdini (2007) identified six key principles of persuasion, namely, reciprocity, commitment or consistency, social proof or conformity, authority, liking, and scarcity (see Section 2.4). Certain persuasion principles incorporated in email-based phishing have been shown to be effective with certain individuals (Parsons et al., 2019) and can also be executed on SNSs (Algarni, Xu, & Chan, 2014; Frauenstein & Flowerday, 2020). Phishers can use these aforementioned principles to trigger a user to engage in risky behaviour such as sharing malicious links, as well as to entice users to “like”, click and share malicious posts. This is illustrated in Figure 1.2, which shows a Facebook post purportedly originating from News24.com, a prominent online news source for South African users. The post shows internationally known Hollywood actress Charlize Theron purportedly involved in reporting something controversial about banking institutions on the Late Night Show. The catchy headline triggers the users’ interest to look further.

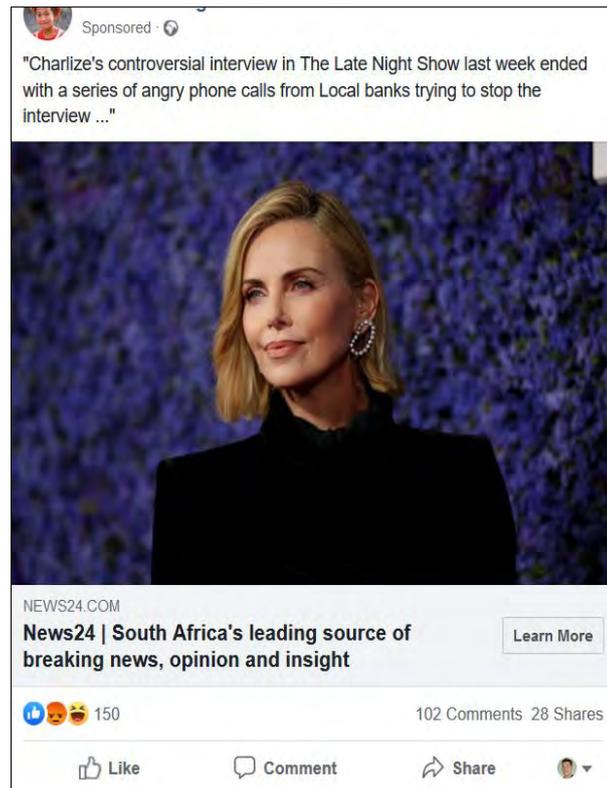


Figure 1.2: Phishing post on Facebook which leads to spoofed website

The appearance of the post in Figure 1.2 makes it difficult to distinguish this as fake, as News24.com is a legitimate website and a well-recognised source of online news. Moreover, the post is strengthened by the “authority” persuasion technique, as it features a celebrity recognisable to most people. It also draws in the unsuspecting with a headline that provokes “curiosity” and potential financial opportunity. If the user clicks on the link in the Facebook post they are subsequently directed to a phoney News24 website. This example shows the effectiveness of a two-level attack process (Vishwanath, 2017), as the phishers went a step further to imitate the News24 website as well.

Users engage across SNSs with little concern for their privacy and the security of their information, as their trust during these online interactions is based mainly on the visual attributes of other users, pages and groups. While the default tendency for people to believe that others are honest helps facilitate cooperation and communication between people, it can also make people vulnerable to the occasional deception (Levine, 2014). Accordingly, users will be more willing to help others and disclose personal information (Junger, et al., 2017; Luo et al., 2011). It has also been found that there is little difference between the level of users’

trust in both online and offline environments (Mesch, 2012). The latter creates opportunities for phishers to target users using *physical* or *technical* approaches or both (discussed in Section 2.3).

Since Facebook's inception in the international market in September 2006, its once plain and minimalistic design has continued to evolve, incorporating new features to further enhance the user experience while also employing persuasive techniques to increase interaction and user retention (Alutaybi et al., 2019). Over the past fifteen years, some prominent features introduced into the Facebook interface include advertisements and Pages (i.e. public profiles) (November 2007), the "like" function (February 2009), instant messenger (August 2011), timeline feed (September 2011), safety check (October 2014), like reactions/emoticons (February 2016), the marketplace (October 2016) and Messenger rooms (July 2020). As there are no restrictions on geographical boundaries on these platforms, the videos, comments and pictures that are posted allow phishers to reach a much broader audience, made possible by the sharing feature. While these aforementioned technical features enhance the user experience, they could be used effectively by phishers to create and spread deceitful posts such as controversial videos, conspiracy theories, fake news and advertisements that are of interest or benefit to the user (Sushama et al., 2021; Tacchini et al., 2017). Moreover, they are also used to create fake accounts, fake apps and groups, and to distribute malicious content (Fire et al., 2014).

While the technological features embedded in SNS platforms can be leveraged by phishers, these features also bring out certain behaviours of their user base that can also be exploited. Owing to individual characteristics like *personality traits*, the posts on SNSs can effectively trigger the emotions of certain users, thereby luring them into performing actions that can put them at risk of phishing (Pattinson et al., 2012). As each individual possesses different personality traits, this makes addressing user behaviour and designing solutions to such an even more challenging task. The frequent use of clicking and sharing of links, liking posts, copying and posting messages, and uploading and downloading media content may have the potential to unintentionally lead to the formation of undesirable *habits* (Turel & Serenko, 2011). Subsequently, a "habitual mind-set" can make an individual less attentive to new information and courses of action (Verplanken & Aarts, 1999). As a result, as users are repeatedly performing a behaviour out of habit on these sites, they could become lax and unintentionally click on or share malicious posts because they did not apply sufficient

cognitive processing (Vishwanath, 2017). To reduce phishing susceptibility, users must apply a *systematic* mode of information processing when presented with phishing messages. However, as users receive a proliferation of information on the SNSs, originating from different sources at any given time, users are inclined to apply a *heuristic* mode of information processing to resourcefully save time and effort. Social media users might not *perceive any risks* on these sites (Krasnova et al., 2009) or they may think that risks will be more likely to happen to others (Kim & Hancock, 2015). As such, users might not make an effort to verify the validity and authenticity of the origins of messages. In this regard, prior research has shown that the Big Five personality traits influence information-seeking behaviour (Heinström, 2003). Moreover, some users may not have the aptitude to seek verification. To elaborate, some users may not have the required computer experience or *computer self-efficacy* to go outside the realm of Facebook to search other websites in order to verify the origins of the message. Owing to *social norms*, Facebook users typically want to be perceived by others in favourable ways and sometimes do not think about the reason for their sharing or posting messages or content that may unintentionally be offensive to others (Wang et al., 2011) and the potential to damage organisations' reputations (Langheinrich & Karjoth, 2010).

Each of these potential areas of concern subsequently increases the risk of overlooking phishing on SNSs. A systematic review and meta-analysis conducted by Parker and Flowerday (2020) confirmed the aforementioned variables of personality traits, online habits and information processing to be factors that can make users susceptible to phishing on SNSs. Moreover, a study by Alotaibi (2019) also considered these components can make LinkedIn users susceptible to SE attacks. Essentially, these factors form the theoretical foundation for this study. In addition, this study considers the effect other dispositional factors such as perceived risk, social norms and computer self-efficacy have on phishing susceptibility on SNSs.

From the background of the problem, it is evident that current methods are inadequate to address phishing, especially in the case of phishing conducted on SNSs. Prior literature has, to a certain extent, mainly posited that phishing victimisation, much like responding to scams, stems from a lack of awareness or knowledge. However, literature reveals that despite users receiving training on phishing, after a certain period they will return to their "old ways" (Dodge et al., 2007). However, the behavioural priming of users' frequent engagement on SNSs, which could develop into habits, is another factor that can influence a user's judgement, leading to

overlooking certain malicious messages on SNS. Accordingly, behavioural vulnerabilities, stemming from dispositional factors, could overpower rational thinking with regard to spending more time deliberating the authenticity of a message (i.e. phishing message).

1.4 Problem statement

The problem statement for this study can therefore be stated as follows:

Certain dispositional factors, such as personality traits, habits and information processing, influences users' ability to detect phishing messages on social networking sites, thus putting organisations and other users at risk.

This section has specified the problem statement that this study addresses.

1.5 Research questions

From the problem statement, this study implies that certain users on SNSs are more at risk from phishing than others. As such, it follows that it is necessary to investigate the behavioural influence of users pertaining to their personality traits, habits and information processing, and how they respond when presented with various phishing messages found on SNSs. To address the problem statement, the following main research question is postulated:

Main question. *To what extent do certain dispositional factors, such as personality traits, habits and information processing, influence a user's response to phishing messages on social networking sites?*

To address each of the dispositional factors contained in the main research question, the following sub-questions were put forward:

- **Sub-question one.** *What type of messages are social media users exposed to which can put them at risk of phishing attacks?*

In order to answer this question, a literature review was conducted to identify the common phishing messages or stimuli employed on Facebook. In achieving this, the study makes use of examples (in the form of screenshots) to subsequently identify how each personality trait (in sub-question two) would respond to such threats in the context of SNSs.

- **Sub-question two.** *Users with which personality traits are more susceptible to phishing attacks than others?*

The main personality traits recognised in this study are known as the “Big Five” and are categorised as openness, conscientiousness, extraversion, agreeableness and neuroticism (Gosling et al., 2003). Prior phishing research has established that users who possess a particular personality trait are more susceptible to falling prey to phishing attacks than others (Cho et al., 2016; Pattinson et al., 2012). As a result, it is important to determine what personality traits a user possesses, as it can help identify to what extent each of the five personality traits are more at risk than others to phishing on SNSs.

- **Sub-question three.** *To what extent do personality traits influence habit when users are presented with phishing messages on social networking sites?*

Sub-question three attempts to identify which Big Five personality traits have an influence on the way users typically respond when confronted with phishing messages. Specific characteristics associated with each trait may influence a user to perform certain behaviours out of habit. Such habit might emanate from characteristics in certain personality traits. For example, an individual who possesses an openness trait is typically curious about the world and other people and also tends to be adventurous. Thus, they might frequently use SNSs to interact with other people, search for new opportunities, or be willing to take risks in order to gain new experiences. In the quest to satisfy these needs, they might develop a habit. Much like the act of clicking on hyperlinks in a phishing email, this study recognises the behaviour of clicking or sharing a social media post, requested by a friend, as a habit (depending on its extent) and could potentially put users at risk of phishing attacks. The aim of this question was to evaluate hypothesis 1a–e of the study, as depicted in Figure 5.1 in the proposed model.

- **Sub-question four.** *To what extent does habit influence information processing when users are presented with phishing messages on social networking sites?*

Users spend a great deal of time on SNSs performing many activities that are of interest or benefit to them. The habitual behaviour of clicking or sharing a social media post, requested by a friend, could influence a user not to perform a more thorough evaluation or assessment of messages that could potentially be malicious. In this regard, this study considers “cognitive” assessment as information processing, based on the Heuristic-Systematic Model (HSM) of information processing, and serves as the theoretical foundation of the study and one of the main factors contributing to phishing deception (Luo et al., 2013). Thus, the aim of this question is to determine to what extent habit influences information processing, specifically the mode of information processing – either heuristic or systematic. Moreover, the aim of this question was to evaluate hypothesis 2a and b of the study, as shown in Figure 5.1 in the proposed model.

- **Sub-question five.** *How does habit influence susceptibility to phishing?*

In order to distinguish whether the relationship of habits to information processing is directly part of the same process or not, it is necessary to also investigate the effect of habits on phishing susceptibility. By doing so, it may confirm whether performing a habit involves a mode of information processing or not. The aim of this question was to evaluate hypothesis 3 of the study, as shown in Figure 5.1 in the proposed model.

- **Sub-question six.** *How does information processing influence phishing susceptibility when users are presented with phishing messages on social networking sites?*

This question is to determine in which mode of processing users will respond, either heuristic processing or systematic processing, when presented with phishing messages found on Facebook. The choice of mode has an influence on the outcome of a phishing attack. The aim of this question was to evaluate hypotheses 4 and 5 of the study, as shown in Figure 5.1 in the proposed model.

- **Sub-question seven.** *To what extent do the dispositional factors of social norms, computer self-efficacy and perceived risk influence phishing susceptibility?*

The literature revealed several dispositional or behavioural factors that can also influence phishing susceptibility. These factors are social norms, perceived risk and computer self-efficacy. The aim of this question was to determine what role each of these factors plays in phishing susceptibility. This question relates to evaluating hypotheses 6–8 of the study, as shown in Figure 5.1 in the proposed model.

Overall, the aforementioned research questions examine the extent to which these variables influence users' responses to phishing on SNSs. Figure 5.1 illustrates the model diagram with the associated hypotheses related to the research questions above.

1.6 Research objective

In order to address the research questions above, the primary objective of this study is:

To develop a model that identifies users who are susceptible to phishing on social networking sites.

A model is “a set of propositions or statements expressing relationships among constructs” (March & Smith, 1995). The model serves to explain the underlying variables that influence user behaviour when presented with phishing messages on SNS. This is important, as it allows investigation into and reasoning about the phenomena described by the model. The constructs required in the model were identified through a literature study, paying particular attention to existing phishing models that aim to prevent, explain or reduce user susceptibility to phishing. In this regard, the approach used to identify the constructs is briefly discussed in the next section.

1.7 Research design and methodological approach

This study follows a *post-positivist* paradigm, as it concerns both an objective epistemological and critical realist ontological position (Levers, 2013). To elaborate, this study adapted a deductive approach as it utilised an existing theoretical model (i.e. the HSM) and designed a research strategy to test the hypotheses. Moreover, owing to the exploratory nature of identifying the appropriate constructs in the literature, an inductive approach contributed to the development of the hypotheses, which subsequently led to the construction of the theoretical model. As depicted in Figure 1.3, the research process involves a sequence of actions or stages necessary to carry out the study effectively.

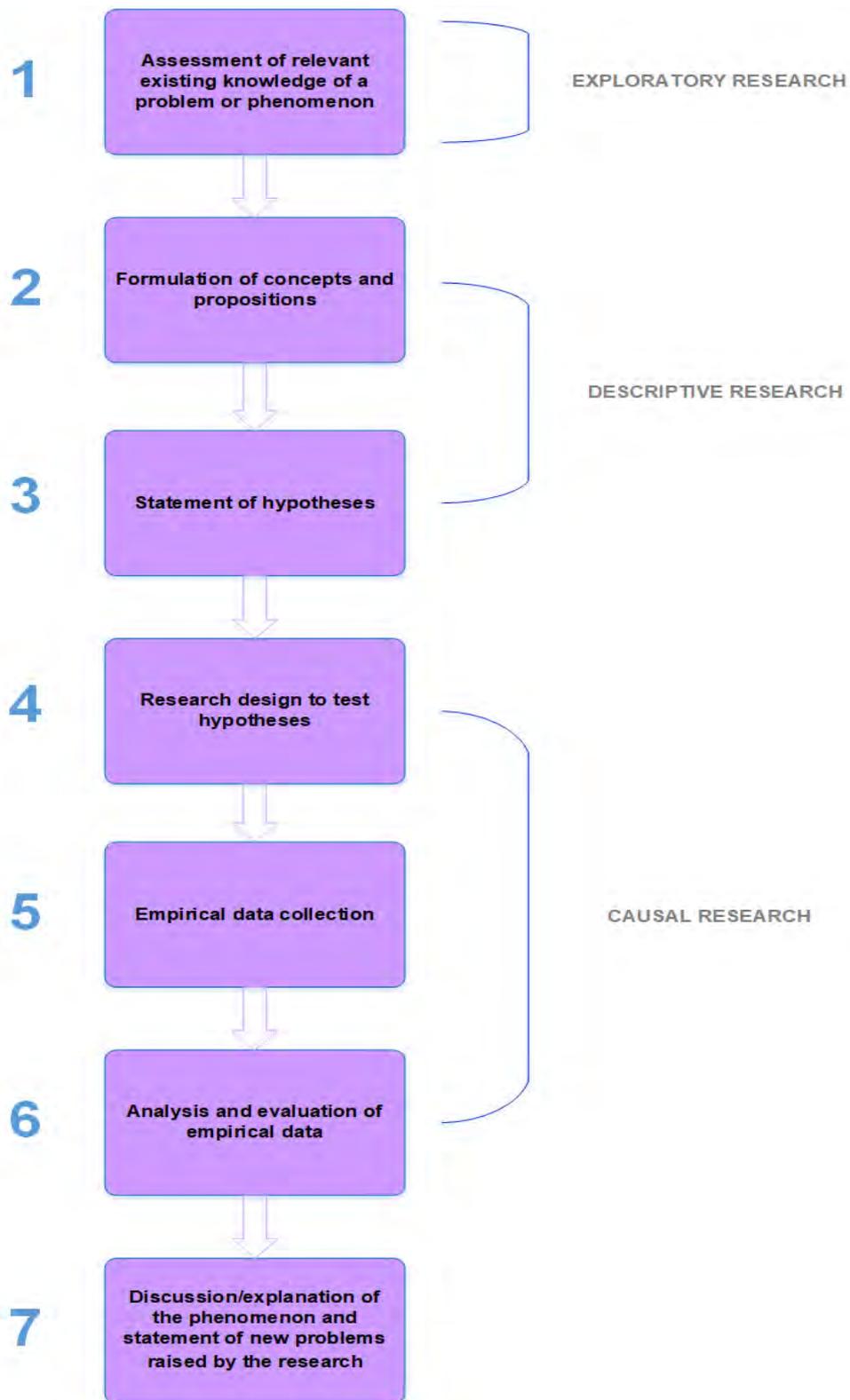


Figure 1.3: The seven stages involved in the application of the scientific method (adapted from Zikmund et al., 2013)

This study adopted a survey strategy and collected primary data by means of an online questionnaire using SurveyMonkey. To qualify for the survey, respondents had to be final-year university students and active SNS users at a South African university. A pilot survey was initially used to ascertain whether the instrument was valid and reliable, following which the final questionnaire was used to collect data for the main study. The final questionnaire consisted of a total of 122 items and, where possible, the items were either taken directly from other sources or adapted accordingly. The distribution of these items in relation to each construct is as follows:

Table 1.1: Constructs identified for the study

Construct	Items	Source(s)
Personality traits (Big Five)	44	John & Srivastava (1999)
Habit	12	Verplanken & Orbell (2003)
Information processing*	7	Griffin et al. (2002) Vishwanath et al. (2011)
Social norms	6	-
Computer self-efficacy	8	-
Perceived risk	4	-
Phishing susceptibility	6	-

*Note: Six stimuli were used to test information processing, with a combined 42 items for this construct.

The population for this study ($N = 477$) consisted of university students in their final year of study located across three university campuses in South Africa. The population was made up of students of Information Technology, Analytical Chemistry, Business Management, Public Relations, Financial Accounting, and Mechanical Engineering. A total of 215 responses were collected. Following data collection, reliability and validity tests were conducted. The *reliability* of the items was determined by the widely used Cronbach's alpha coefficient (α) which was used to test the internal consistency of multiple-item scales and correlate these items to the related construct (Collis & Hussey, 2014, p. 275). In addition, to ensure that the scales measured what they were expected to measure, the *validity* of the constructs was tested using factor analysis. Factor analysis examines the relationship between pairs of variables measured on a rating scale (Collis & Hussey, 2014, p. 276). To measure the linear associations between the various constructs, Pearson's correlation coefficient was used (Collis & Hussey, 2014, p.

275). Any shortcomings in each of these requirements led to the elimination of several items. Using a deductive approach, the hypotheses were statistically evaluated by employing several multivariate analysis techniques known as structural equation modelling (SEM). This was a suitable approach for conducting the analysis for this study, given the number of items (in the constructs) and considering the complexity of the relationships required to form a single structural model. A more detailed discussion on the data analysis techniques employed in this study is presented in Section 6.10.

1.8 Contribution

Prior studies conducted in SE and phishing have focused mainly on email-based phishing and not on phishing in the context of SNSs (Algarni et al., 2017). As users least expect to be phished on SNSs (Volkman, 2019b), their behaviours on these sites pose risks to both general end users and organisations alike. As SNSs are popular the world over and most internet users have at least one SNS account, this study will be of significance to anyone or any organisation that wishes to strengthen their “human firewall” to protect against phishing attacks (Mansfield-Devine, 2018). From an *organisation* perspective, Chi (2011) points out that organisations must recognise the security threats associated with SNSs and put in place sound policies to reduce such threats. In this regard, this study also makes an indirect contribution to organisations and can assist them to identify specific “human-related” areas phishers exploit, thus reducing the various risks brought on by these threat agents. Identifying these factors and how they relate to each other can offer organisations insights to these vulnerabilities through focused control measures such as technical controls, security education, training and awareness, and improving their existing security policies. From a *scholarly* perspective, the contribution of this study is novel as no prior study has combined these different variables into a single model. By doing so, this study provides a comprehensive view by characterising individuals who are at risk to phishing. As such, this research makes a novel contribution to the body of knowledge, in the form of a theory, in the behavioural information security domain (Vaishnavi & Kuechler, 2015). This study has benefits for future researchers who may then adopt a similar model or theory by incorporating more constructs and it may also be extended to other online applications. Moreover, the study offers insights to scholars, as real-world phishing stimuli found on Facebook were utilised. A more detailed discussion on the study’s contribution to theory and practice is provided in Section 9.4.

1.9 Ethical considerations

Ethics is concerned with the manner in which data is collected and how the findings are reported (Collis & Hussey, 2014, p. 30). Phishing-related experiments, in particular, are well known to bear relevant ethical concerns and requirements for information security researchers (Finn & Jakobsson, 2007; Mouton et al., 2013). However, in this study, the data collection process does not attempt to conduct any form of phishing attack against the participants of this study, nor does it put the university and the participants at any risk or loss – whether financially, emotionally or physically. In abiding by the ethical principles as recommended by Collis and Hussey (2014, p. 31), research participants were informed of the purpose of the study and that they could withdraw their participation from the study at any time (see Appendix C). Collis and Hussey (2014, p. 31), recommend that data collected from participants should offer *anonymity* and *confidentiality*. In accordance with this requirement, anonymity was assured as the participants and the organisation were not named in this research. Confidentiality is concerned with ensuring that information provided by the participants and organisations cannot be traceable to the organisation or participants providing it (Collis & Hussey, 2014). This requirement was also assured as the respondents were not required to provide their identities on the survey instrument. A more detailed discussion on the ethical process can be found in Section 6.11 and the informed consent form can be found in Appendix C.

1.10 Thesis layout

The study is organised into nine chapters. This chapter provided an introduction to the study by describing the background, the problem statement and the study objectives, as well as giving a brief summary of the research process, the research design and the methodology.

The literature review, which spans Chapters 2 to 4, serves as the theoretical foundation of the study and provides support for the hypotheses. *Chapter 2* discusses prior literature on SE and persuasion strategies and how they can effectively influence users to become vulnerable to phishing attacks on SNSs. This is important as certain persuasion strategies are key to the success of phishing on SNSs. This is followed by a discussion on dispositional/interpersonal factors (i.e. habit, perceived risk, computer self-efficacy, and social norms) that can influence the outcome of a phishing attack.

Chapter 3 introduces another major dispositional factor of the study, the Big Five personality traits, and presents a literature review which demonstrates that certain traits may be more or less susceptible to phishing attacks than others. In addition, the literature also shows that personality traits may be influenced by certain environments (i.e. workplace, SNSs) and by gender and persuasion strategies, and may potentially be susceptible to forming an internet dependency or addiction.

Chapter 4 introduces information processing, utilising the Heuristic-Systematic Model (HSM) as an underlying factor of phishing susceptibility. Moreover, it serves as the theoretical framework for this study as it describes why some users give insufficient attention to detail when presented with phishing messages on SNSs.

This is followed by *Chapter 5*, which formally presents each of the hypotheses derived from the resultant literature review. The chapter also formally illustrates the conceptual model for this study with the associated hypotheses. *Chapter 6* provides a detailed discussion as to the research design and methodology applied in this study, while *Chapter 7* presents the results of the multivariate analysis conducted in this study. Importantly, it illustrates the theoretical model (i.e. structural model) of the study in the form of a path diagram. The path diagram shows the significant relationships each of the constructs have with each other. It also provides evidence to suggest that the model is structurally sound, reliable and valid.

Chapter 8 provides a detailed discussion based on the results of each of the hypothesis tests with the aim of addressing each of the research questions. *Chapter 9* formally concludes the study by summarising the earlier chapters and ensuring that the research questions have been answered and the resultant research objective has been achieved. The chapter also presents publications stemming from the research study. Figure 1.4 illustrates the layout of the chapters for this study.

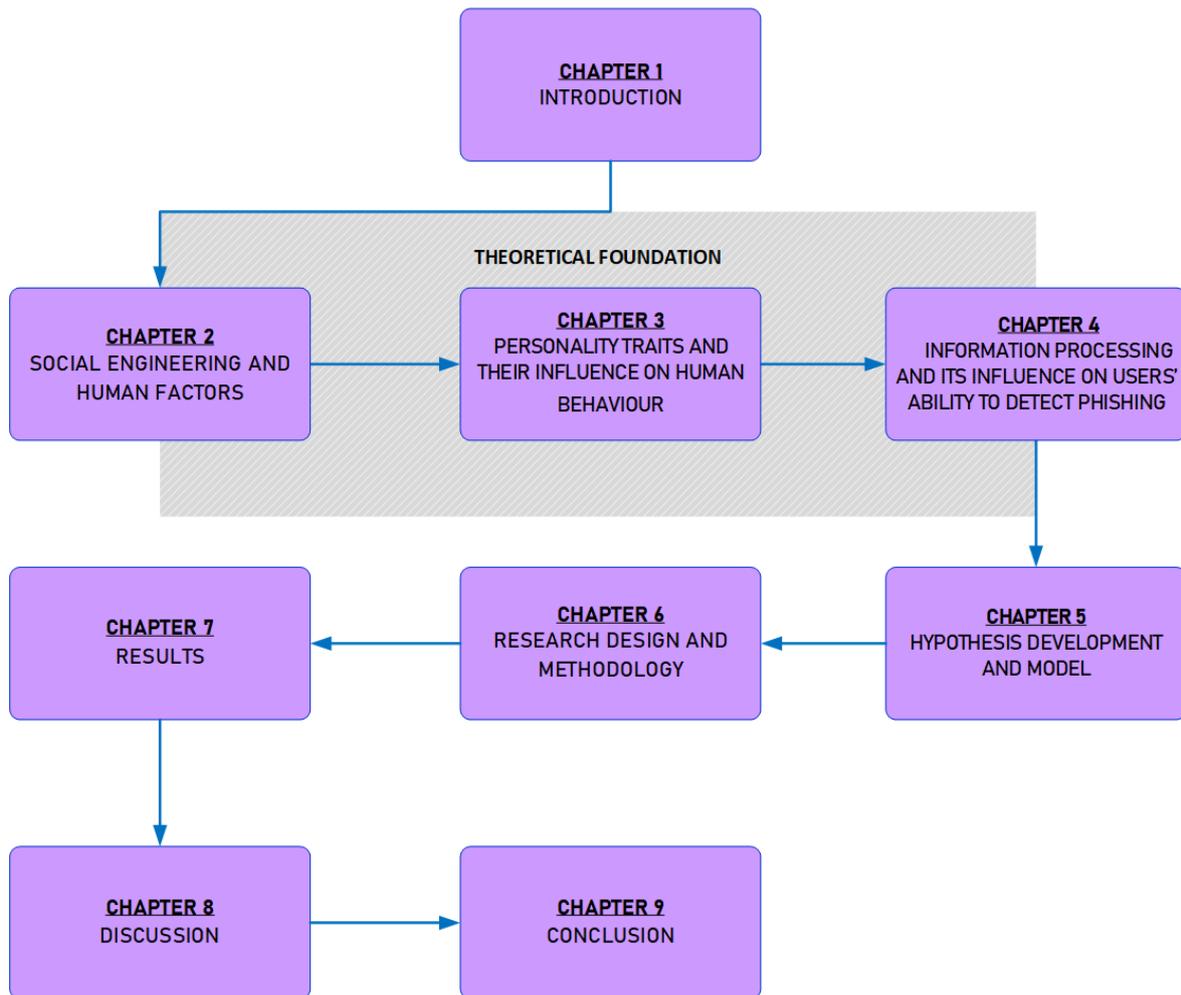


Figure 1.4: Chapter layout of the study

SOCIAL ENGINEERING AND THE HUMAN FACTOR

2.1 Introduction

Technology as a means of protection against security threat agents has been available for many years and yet information security attacks and threats remain a risk for both organisations and individuals. While security countermeasures can be automated through technology, the protection of information no longer relies solely on technological solutions as it once did. With every new and improved technological control, security threats continue to seek new ways in which to exploit the technological vulnerabilities in order to gain access to the valuable information contained on various information systems. Alternatively, security threat agents have found it easier to exploit human vulnerabilities instead of focusing on circumventing technological controls (Luo et al., 2011).

The inception of the internet age has resulted in security efforts being directed towards addressing the human factor, as users themselves are considered the most vulnerable part of an information system (Mann, 2008; Mitnick & Simon, 2002). This is because it is humans that design, implement, manage and operate information systems. With this involvement, information systems and data are exposed to various risks such as design flaws, theft, weak passwords and the behavioural vulnerabilities of users. These behavioural vulnerabilities, the main focus of this study, comprise a multifaceted area that continues to be explored today by information security researchers. In the case of phishing, despite the magnitude of efforts made to warn users of imminent risks, prior literature has determined that users ignore security toolbars and web browser warnings (Egelman et al., 2008; Junger et al., 2017; Petelka et al., 2019; Wu et al., 2006). This might be because users consider the warnings to be an

inconvenience as they are interrupting them when focusing on their primary work-related tasks (Bada et al., 2015). While some of these behaviours may be unintentional, people may also, as a result of their attitudes, deliberately misuse or abuse information systems and, by doing so, unintentionally expose the organisation to information security threats.

As this study focuses primarily on the behavioural vulnerabilities of users that put themselves at risk for phishing, the aim of this chapter is to discuss social engineering (SE) approaches and persuasion principles, as they underpin the effectiveness of phishing deception (Ferreira & Teles, 2019). By doing so, this chapter aims to address the first research sub-question of the study. The chapter also provides a review of prior literature on SE and discusses influential interpersonal factors that affect human behaviour which subsequently can make users more vulnerable to phishing attacks on SNSs. Moreover, these interpersonal factors form part of the theoretical foundation for this study. This next section presents a review of literature in the area of social engineering (SE) and phishing.

2.2 Social engineering

Frank Abagnale, an American former con man and cheque forger, stated: “There is no technology today that cannot be defeated by social engineering” (BrainyQuote, 2019). More recently, when asked how much easier cybercrime is today than 50 years ago, Abagnale replied that it is “4000 times easier” due to the proliferation of technology accessible today that was not available in the past (Roby, 2019). The term “social engineering” in the cybersecurity domain started as the “phone phreaking” phenomenon in the late 1950s and continued into the early 1970s (Hatfield, 2018). Social engineering (SE) is defined as a form of psychological manipulation consisting of techniques used to influence and persuade potential victims to perform actions or divulge confidential information that would in some way be of benefit to the attacker (Mitnick & Simon, 2002). Mouton et al. (2014) define SE as “the science of using social interaction as a means to persuade an individual or an organization to comply with a specific request from an attacker where either the social interaction, the persuasion or the request involves a computer-related entity”. SE is synonymous with such terms as a “human confidence game”, “confidence trick” and “human hacking”. Mitnick and Simon (2002) popularised the term and stated that from practical experience it is far easier to obtain information directly from people, by means of persuasion, than to hack information systems.

Each of the persuasion techniques and how they can be executed on SNSs is discussed further in Section 2.4. The outcomes of an SE attack can be used illegally for financial gain and range from identity theft to the selling of bulk personal data such as driver licences, credit card numbers and identification numbers on the black market (Workman, 2007). According to Krombholz et al. (2015), SE remains the most successful strategy for compromising large-scale organisations. SE techniques are effective because they evoke human emotions such as excitement, greed, sympathy, fear, trust and commitment (Algarni et al., 2013b). The next subsection categorises the literature pertaining to SE.

2.2.1 Taxonomies, models and frameworks

Owing to the ever-changing nature of phishing attacks and social engineers' methods and attacks, scholars have found a need to develop taxonomies and frameworks aligned to various mitigation strategies (Chiew et al., 2018). The primary aim of such taxonomies is to identify the SE threat landscape and the associated attack types, approaches and vectors (Alabdan, 2020). This in turn would be useful in helping organisations to identify their vulnerabilities and to develop effective countermeasures to combat SE attacks. Chiew et al. (2018) identify three components required to execute phishing: the medium of phishing, the vector to transmit the attack and the technical approaches used during the attack. Krombholz et al. (2015) present a taxonomy, as shown in Figure 2.1, classifying SE attacks by channel, operator and type.

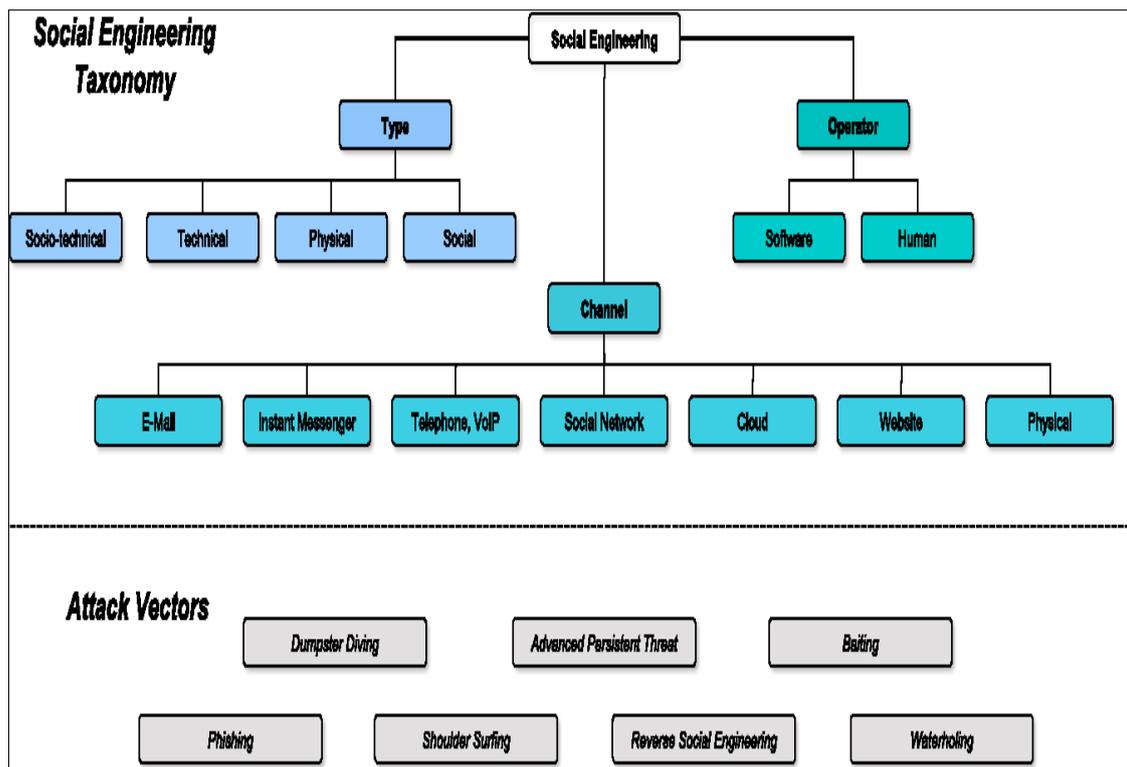


Figure 2.1: Social engineering taxonomy classifying attack characteristics and attack scenarios (Krombholz et al., 2015)

As shown in Figure 2.1, Krombholz et al. (2015) state that SE attacks can be performed through various *channels* such as email, instant messenger, telephone/VOIP, social networks, cloud, website and physical. Similarly, in a phishing context, Chiew et al. (2018) considered email, eFax, instant messaging, SNSs and websites to be *vectors*. Krombholz et al. (2015) classifies the *operator* category into human and software, and SE attacks into four types, physical, technical, social and socio-technical. Krombholz et al. (2015) further classify the attack vectors as phishing, shoulder surfing, dumpster diving, advanced persistent threats, baiting and waterholing. Krombholz et al. (2015) consider SNSs to be platforms for launching SE attacks, with **social phishing** and fake profiles considered attack vectors within this area. Chiew et al. (2018) consider phishing to be merely deploying a technical-based approach, including browser vulnerabilities, clickjacking, cross-site scripting (XSS) attacks, drive-by download, man-in-the-middle, javascript obfuscation, SQL injection, malvertising, session fixation, SE, tabnapping, typo-squatting, sound-squatting, spear phishing, whaling, wiphishing, mobile phone and phishing kits. Chiew et al. (2018) go on to say that these technical approaches could

function independently or in combination (e.g. clickjacking and an XSS attack) as depicted in Figure 2.2.

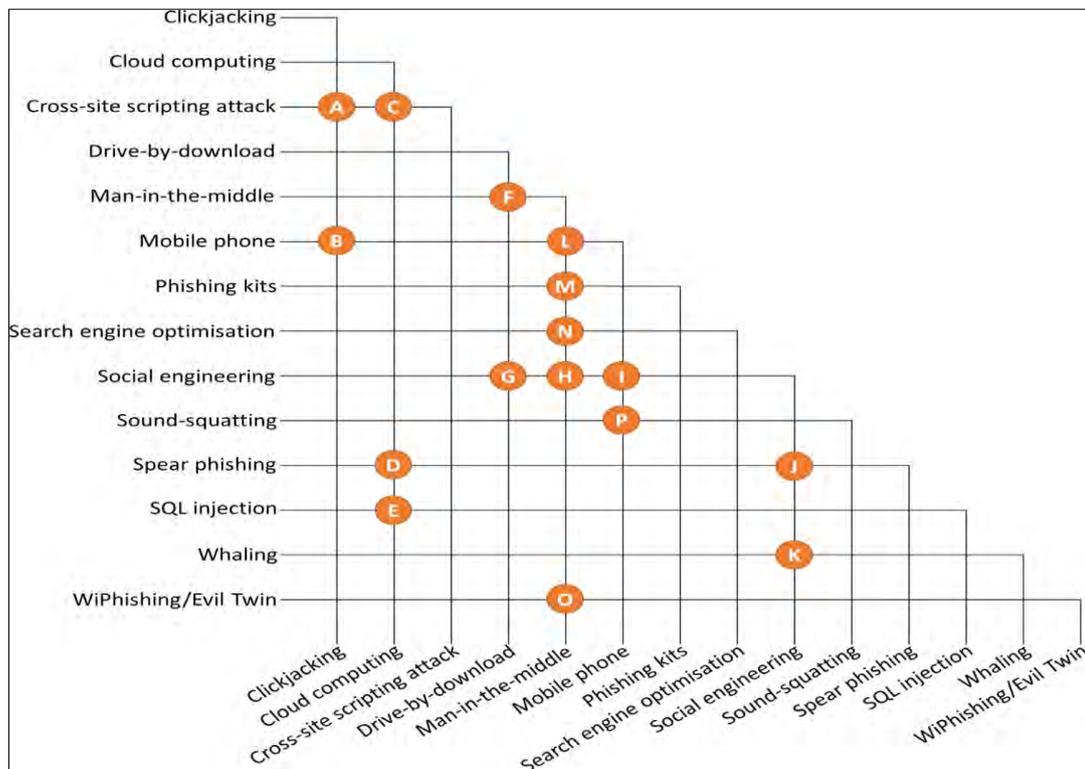


Figure 2.2: Combination of technical approaches in a phishing attack with intersections represented by the nodes (Chiew et al., 2018)

Heartfield and Loukas (2015) presented a taxonomy of semantic SE attacks and associated defences to mitigate the attacks in the form of a single comparative matrix, which helped identify areas that needed to be addressed. Ivaturi and Janczewski (2011) classified their taxonomy according to two major categories: person to person and person to person via media. The latter is the largest category as it includes phishing, SMSishing, vishing, social networking, cross-site request forgery (CSRF), malware, email, pop-ups, and search engine poisoning. The disparities between taxonomies was noted by Mouton et al. (2014, 2016), who state that current literature on SE attacks is limited because it does not include all the attack steps and phases. Mouton et al. (2016) attempted to address this problem by introducing detailed SE attack templates derived from real-world SE examples and aligned them to an SE attack framework. Mouton et al. (2014) note that the phases of Mitnick’s attack model are not explained in enough detail and, as a result, they developed an SE attack framework that

expands on each phase of Mitnick’s model. In the complex ontological model by Mouton et al. (2014), the authors state that an SE attack “employs either direct communication or indirect communication, and includes a social engineer, a target, a medium, a goal, one or more compliance principles and one or more techniques”. According to their model, the attack can be divided into more than one attack phase, and each phase is handled as a new attack. Mouton et al.'s (2014) model is depicted in Figure 2.3.

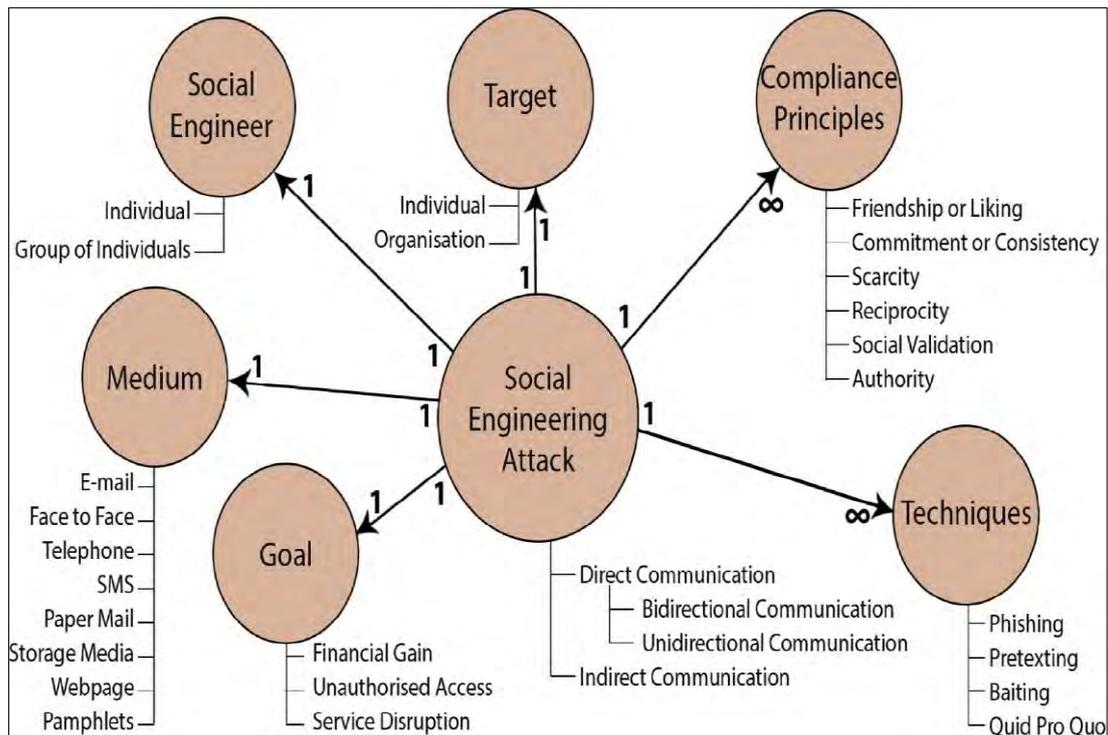


Figure 2.3: An ontological model of a social engineering attack (Mouton et al., 2014)

Mouton et al. (2014) did not include SNSs under the category of medium. On the other hand, following a literature investigation, Algarni et al. (2013a) identified different entities and sub-entities that affect SE-based attacks on SNSs. For social network users, they identified socio-psychological factors that are affected by personality types, demographics, and motivation and drive. Aleroud and Zhou (2017) propose a phishing taxonomy that considers attack techniques, countermeasures, environments and communication media. Their taxonomy provides guidance on the design of effective techniques for phishing detection and prevention in various types of environments.

2.2.2 Persuasion principles

Persuasion strategies and their effect on user behaviour play a central role in the effectiveness of phishing attacks and continue to be an area explored by security scholars (Ferreira & Teles, 2019; Ferreira et al., 2015; Kaptein & Eckles, 2012; Kaptein et al., 2009; Lawson et al., 2017; Lawson et al., 2018; Lawson et al., 2020; Oliveira et al., 2017; Parsons et al., 2019). The effectiveness of persuasion strategies varies from one person to another (Kaptein & Eckles, 2012). Kaptein et al. (2009) found that incorporating a persuasive cue can indeed increase compliance with a persuasive request. Akbar (2014) investigated the relationships between persuasion techniques and generic characteristics of phishing emails and found that *authority* is the most effective persuasion principle regardless of the target and motives behind using it. Bullée et al. (2015) investigated to what extent an awareness campaign influenced participants to comply with an SE request employing the *authority* principle. They found that in their sample, the odds of being susceptible to SE were 2.84 times higher when participants had not received any awareness intervention. Akbar (2014) also found that the principles of *scarcity*, *consistency* and *liking* were dependent on the type of target and the motive. The *scarcity* principle has a high involvement with the administrator target type and account-related concerns. Ferreira et al. (2015) pointed out that Cialdini’s six principles of persuasion are not the only persuasion principles relevant to SE attacks. Accordingly, they propose a reviewed list of persuasion principles related to SE.

Table 2.1: Principles for influencing users into performing actions (adapted from Ferreira et al., 2015)

Principles of Persuasion (Cialdini, 2007)	Psychological Triggers (Gragg, 2003)	Principles of Scams (Stajano & Wilson, 2011)
Authority	Authority	Social Compliance
Social Proof	Diffusion of Responsibility	Herd
Liking & Similarity	Deceptive Relationship	Deception
Commitment & Consistency	Integrity & Consistency	Dishonesty
Scarcity	Overloading	Time
Reciprocation	Reciprocation	Need & Greed
	Strong Affect	Distraction

As evident in Table 2.1, there are common techniques overlapping each of the principles used by social engineers. For example, the principle of *authority* is common to both the work by Cialdini (2007) and Gragg (2003). Furthermore, it should not be unexpected if, to increase deception, attackers incorporate more than one principle in a message. As Lawson et al. (2017) found, a combination of *authority* and *scarcity* persuasion principles is most likely to arouse suspicion in phishing emails.

Workman (2007, 2008) developed a theoretical framework from an empirical field study investigating SE attacks. Workman (2007) identified several factors used in their framework, namely perceptions of threat severity and vulnerability, reciprocity, likeability, trust, fear, authority, scarcity and commitment. Inspired by the work of Mitnick, Workman's study was underpinned by the Elaboration Likelihood Model (Petty & Cacioppo, 1986). Subsequently, Workman (2007) found the three types of commitment (normative, continuance, affective) as having a pivotal role in the success of SE attempts and related each of these types of commitment to Cialdini's persuasion principles. Workman (2007) found that people who are high in normative commitment will feel obligated to *reciprocate*. People with high continuance commitment will give up increasingly sensitive information if it benefits their goals, and high affective commitment was also found to contribute to SE attempts. People who were trusting were more likely to fall victim to SE than those who were distrusting. Workman (2007) tested authoritative commands and fear strategies in relation to the levels of people's obedience to *authority* and reaction to *scarcity*. Higher degrees of obedience to *authority* were important factors in whether people responded to these types of SE attacks; however, the same result was not found for *scarcity*.

Butavicius et al. (2015) examined the influence of *authority*, *scarcity* and *social proof* on users' ability to distinguish between legitimate, phishing and spear-phishing emails. Of the three persuasion principles, *authority* was the most effective in convincing users that a link in an email was safe. When detecting phishing and spear-phishing emails, users performed worst when the emails used the *authority* principle and performed best when *social proof* was present. Overall, Butavicius et al.'s (2015) study found that users struggled to distinguish between genuine and spear-phishing emails. Finally, users who were less impulsive in making decisions were less likely to judge a link as safe in the fraudulent emails. Oliveira et al. (2017) investigated the extent to which age can be a factor in the successfulness of persuasion

strategies in spear-phishing emails. They found younger adults were most susceptible to *scarcity* while older adults were most susceptible to *reciprocation*.

It is thus evident, as this section shows, that not all individuals react in the same way to each of the persuasion principles. As a result, substantial research has turned to personality traits to identify whether certain traits are more susceptible than others to the particular persuasion principles employed by phishers and social engineers. For example, persuasive messages that incorporate the *urgency* principle may be more effective with impulsive users, thus targeting users who possess the *extraversion* trait. A more detailed discussion on this focus area can be found in Section 3.3.3.

2.2.3 User behaviour toward warnings and security indicators

More than a decade ago, prior work by Dhamija et al. (2006) found that when participants were shown 20 spoofed websites, 23% of them did not look at browser-based cues such as the address bar, status bar and the security indicators. Instead, participants evaluated the authenticity of the website based on its content. Egelman et al. (2008) investigated sixty participants' responses to browser warnings related to a simulated spear phishing attack and found that 97% of participants visited the spoofed website by following the link contained in the phishing email. Moreover, Egelman et al. (2008) also found differences in how users would respond to warnings stemming from different browsers (i.e. Firefox and Internet Explorer). On the other hand, Zhang, Egelman et al. (2007) tested the effectiveness of anti-phishing detection tools using 200 verified phishing URLs from two sources and 516 legitimate URLs to test the effectiveness of 10 popular anti-phishing tools. Zhang, Egelman et al. (2007) found that only one tool was consistently able to identify more than 90% of phishing URLs correctly; however, it also incorrectly identified 42% of legitimate URLs as phishing. Similarly, Wu et al. (2006) evaluated three types of security toolbars, as well as browser address and status bars, to test their effectiveness at preventing phishing attacks. They found that all failed to prevent users from being spoofed by high-quality phishing attacks. More recently, when participants were presented with a Mozilla Firefox browser warning reporting a malicious webpage, they either did not trust the warning or did not understand why they received the warning (Frauenstein, 2018). Most of the participants (46.8%) chose to enquire why the page was blocked, while 18.3% ignored the warning. The remainder chose the option: "I know the webpage I want to

go to is safe, I don't understand why this annoying message has come up.” Only 32.5% obeyed the warning (Frauenstein, 2018).

With the aim of protecting users against SE attacks, Junger et al. (2017) investigated the effectiveness of training users by means of cues to raise awareness and to warn against the disclosure of personal information. Despite these interventions, the study found that 79.1% of the participants disclosed their email address, and 43.5% provided bank account information. Among the online shoppers, 89.8% of participants provided information on the type of products they purchased and 91.4% the name of the online shop from which they purchased their products. Junger et al.'s (2017) analysis revealed that neither priming questions, nor a warning had a significant influence on whether users disclosed their personal information or not.

Social engineers have found opportunities to exploit two factor authentication (2FA) mechanisms and coerce users to submit authentication codes to them (Siadati et al., 2017). Siadati et al. (2017) investigated how SE attacks, specifically verification code forwarding attacks (VCFA) on SMS-based out-of-band authentication, could be mitigated by improving the messages used during the authentication process. They looked closely at the message associated with the authentication code and determined that alternative messages can help to reduce coercion. Siadati et al. (2017) found that when a warning contained “Please ignore this message if you did not request a code” precedes the authentication code, it thwarts SE attempts better than any other tested method, resulting in a susceptibility to attack of just 8%. Burns et al. (2019) found that incorporating individual-loss-framed messaging into a training programme increases the effectiveness of the awareness in spear phishing.

This section highlights the fact that although warning indicators and training can assist users to identify phishing, they have their limitations as users can choose to ignore them, or might not understand them, thus indicating that there is possibly other variables hindering the effectiveness of this approach.

2.2.4 Social networking sites

Many scholars have highlighted the various security threats and risks social network users are exposed to (Adewole et al., 2017; Aleroud & Zhou, 2017; Algarni, Xu, & Chan, 2014; Algarni et al., 2017; Chen et al., 2017; Fu et al., 2018; Jang-Jaccard & Nepal, 2014; Jin et al., 2013; Rathore et al., 2017; Silic & Back, 2016; Trivedi et al., 2016; Turel & Serenko, 2012; Wilcox & Bhattacharya, 2015). Jang-Jaccard and Nepal (2014) consider SNSs, such as Facebook and Twitter, as emerging threats that pose several risks to users. Jang-Jaccard and Nepal (2014) state that compromised social network accounts used to distribute malicious links that pose the risk of SE attacks are used to spread malware. According to Algarni et al. (2013a), owing to users' inability to detect SE threats, SNSs is an environment are used by social engineers to exploit their vulnerabilities. Aleroud and Zhou (2017) point out that phishing attacks are a growing problem on SNSs for several reasons: it is not difficult for phishers to impersonate profiles, users' willingness to trust, and SNSs are extremely popular. Moreover, popular SNS like Facebook create opportunities to exploit SE techniques such as posts, tags, applications, games, impersonation using fake profiles, and persuasion through other channels such as the instant messaging platform (Algarni et al., 2017).

Vishwanath (2017) classifies email-based phishing as usually a one-step process where a phisher sends out millions of email requests in the hope that somebody will take the "bait". However, in contrast to email-based phishing, Vishwanath (2017) explains that social network phishing may be viewed as a two-level attack. In a level-1 attack, the perpetrator attempts to connect with or friend a person using a fake persona. In SNSs, this is more effective than emailing, as information about the victim's Facebook friends and even their friends could be used for subsequent phishing attacks (Vishwanath, 2017). A level-2 attack usually involves requesting personal or sensitive information from the victims using Facebook's messaging platform. In some cases, such messages include a hyperlink or attachment, which when clicked installs malware on the victim's computer or could lead to a malicious website. Users who receive a level-2 request may be more likely to respond because it was sent by a Facebook friend associated to the victim. Vishwanath (2017) states that social network phishing attacks present fewer identifiable cues than email-based phishing attacks. Furthermore, the majority of SNS users have access through their smartphones and as such this increases the risk to phishing as the amount of textual information that can be displayed on smartphones is limited.

As a result, this puts more emphasis on users to resort to heuristic cues to make their judgements (Vishwanath, 2016, 2017).

Using a mapping study research technique, Waheed et al. (2017) identified the behavioural characteristics of SNS users. They initially identified 2681 studies that contained the search string “user behavior” after which they concluded with 116 studies eligible for analysis. This included frequency of use, information control, social affiliation, self-orientation, reciprocity, social boldness, and social investigation. The context within which user behaviour was discussed included trust, privacy, age, culture, gender, information sharing, and distance.

According to Halevi et al. (2013), the relationship between psychological factors and social network behaviour has not been thoroughly explored. Given the nature of SNSs as a platform that encourages users to share information about themselves, it is not surprising that privacy is an area that social engineers take advantage of. Gross and Acquisti (2005) conducted a study consisting of more than 4000 Facebook users at Carnegie Mellon University and found that users were unconcerned about their online privacy as they appeared to provide large amounts of personal information about themselves on their profiles and as such expose themselves to various physical and cyber risks.

Algarni et al. (2016) developed an instrument to measure source credibility in terms of SE on Facebook, and presented four dimensions pertaining to source credibility, namely, sincerity, competence, attraction and worthiness. Later, Algarni et al. (2017) investigated the impact of source characteristics on users’ ability to detect SE techniques on Facebook and identified source credibility dimensions, source characteristics and mediation effects as dimensions that play into susceptibility to SE victimization.

2.2.5 Ethics in social engineering experiments

Ideally, when investigating behavioural aspects in information security research, researchers wish to obtain “real” data from their participants instead of relying on theoretical assumptions. This is a double-edged sword, as experiments must be conducted in an ethical way that prevents harm to participants. However, this approach could also skew the results as the findings would not represent reality sufficiently or give appropriate predictive power

(Jakobsson & Ratkiewicz, 2006). As a result, earlier work by Finn and Jakobsson (2007) and Jakobsson and Ratkiewicz (2006) described the ethical and procedural aspects of developing and conducting phishing experiments. SE-based research presents several ethical concerns and requirements for researchers in terms of protecting participants (Mouton et al., 2015). In this regard, Mouton et al. (2015) note that researchers may be unaware of the challenges associated with SE research due to the fact that the ethical requirements have not yet been formalised. They identified concerns related to SE in different environments, particularly public communication, penetration testing and SE research. Mouton et al. (2015) discuss their concerns based on normative ethics approaches, namely, virtue ethics, utilitarianism and deontology, as well as providing ethical guidelines on each of these three approaches and how they can be used in practice.

From the models and taxonomies discussed here, it is evident that phishing is evolving, making it a complex task for researchers to formulate a holistic taxonomy, model or framework that all can agree on. Phishing conducted on SNSs (i.e. social phishing) is regarded as a sophisticated attack as it is included in some of the above-mentioned SE taxonomies. In addition, this section pointed out several focus areas in SE research and discussed some concerns in each of these areas. Personality traits and how they can be influenced by persuasion principles is another area that has been identified for this study and is discussed in more detail in the next chapter. The next section describes various approaches used to carry out SE attacks.

2.3 Social engineering approaches

According to Cusack and Adedokun (2018), SE attacks can be classified into the targeted and a target of opportunity. The targeted refers to when a specific victim is intentionally chosen to be attacked. A target of opportunity is when the attacker broadens the attack to wider audience, much like casting a fishing net, in the anticipation of acquiring a response from as many victims as possible (Cusack & Adedokun, 2018). Krombholz et al. (2015) categorised SE attacks into physical, technical, social and socio-technical approaches, which can be used at different stages of an attack and over a variety of different channels. Using the approaches of Krombholz et al. as a guide, the section below discusses each of these SE approaches.

2.3.1 Physical approach

The physical approach takes place when the attacker physically gathers information on the victim. One of the popular methods to achieve this is known as “dumpster diving” in which the attacker will physically look through the organisation’s trash for any useful and valuable information (Krombholz et al., 2015). Sought-after information can range from personal data related to employees and suppliers, organisation manuals, policies, memos, customer information and even login credentials. The attacker can then use this information to create a more targeted attack on its victims such as a spear phishing attack. Alternatively, the attacker could pose as a trustworthy person in an attempt to physically gain access to the organisation. They would potentially look for unlocked computer devices and passwords written on post-it notes or to convince an employee within the organisation to grant them access to information (Krombholz et al., 2015).

2.3.2 Social approach

Social approaches rely largely on psychological aspects to trick victims into giving out personal information. This could be achieved through the use of persuasion principles (see Section 2.4). One of the most common techniques to accomplish this is over the phone and is known as “vishing”. In this way, the social engineer could be friendly and may establish a relationship with the victim in order to build trust (Krombholz et al., 2015). Moreover, the attacker may also spoof the number to make the call appear to originate from a source that is known or trusted by the user.

2.3.3 Reverse social engineering

As the word “reverse” suggests, instead of the attacker attempting to engage directly with the victim, the victim approaches the attacker (Hatfield, 2018). This indirect approach is accomplished by creating a false situation which will cause the victim to seek assistance from the attacker. This can be achieved by sabotaging an organisational network or by compromising software applications used by the victim. The attacker uses the opportunity to present themselves to the victim as being helpful in resolving the problem. In order to resolve the problem created by the attacker, the attacker requests password credentials or requests the

victim to install certain software which the victim then does. In this situation, the attacker uses the persuasion technique of *reciprocity* (see Section 2.4.3).

2.3.4 Technical approach

Technical attacks are typically carried out over the internet. Like the physical approach, the social engineer can gather information on their target. SNSs and search engines can be used as a means to acquire information on the target (Huber et al., 2009). This is made easier if the victims have made their information publicly available on such websites. Data mining tools, such as Maltego, can be used for open-source intelligence and forensics to gather information from multiple sources on the target (Edwards et al., 2017; Huber et al., 2009). If the victim has used the same password on other web accounts, they increase the risk of compromising all their accounts should the social engineer gain access to one of them. In an experiment, Rößling and Müller (2009) used data from XING, a European career-oriented social networking site, to find more information on who their target was connected to within their organisation of employment. Moreover, they used Facebook to acquire family photos of their targets.

2.3.5 Socio-technical approach

As the term “socio-technical” suggests, it is a combination of the characteristics that form part of the aforementioned social and technical approaches and it is regarded as the most effective approach. Phishing is an example that uses this approach. Jagatic et al. (2007) showed the effectiveness of the socio-technical approach in carrying out a phishing attack. Indicative of the *technical approach*, Jagatic et al. used SNSs to first gather information on their target, in their case students, and subsequently sent them a phishing email made to appear as though it originated from a friend. In this instance, this was regarded as a spear phishing attack due to the specific information related to the target contained in the email. As the current study investigates phishing susceptibility on SNSs, it is aligned more to the socio-technical approach.

As discussed, social engineers can choose various approaches to help them achieve their goal. Besides the *social-technical approach*, other approaches can also be used in combination. SE attacks typically consist of four phases, referred to as “the cycle” which consists of information gathering, relationship development, exploitation and execution (Allen, 2006).

The “information gathering” phase is similar to the characteristics discussed in the technical approach. The “relationship development” phase concerns the SE attempt to establish the trust of the victim in order to develop a connection with them. Following the establishment of a relationship between the attacker and the victim, the attacker attempts to manipulate the victim into revealing confidential information or performing certain actions that will benefit the attacker. Finally, the cycle is complete in the execution phase when the victim has fulfilled the task set by the attacker.

Prospective phishers can also download and purchase sophisticated phishing kits on distribution sites advertised by underground websites mainly found on the Dark Web (Oest et al., 2018). Those that can be downloaded for free often contain backdoor viruses. These kits are in an easy-to-deploy format and contain scripts (developed in basic HTML and PHP) to mimic complete phishing websites such as Amazon, eBay, Google, Instagram, Office 365, and PayPal. SE attacks can also be conducted using software tools such as Kali Linux (formerly BackTrack) which contains a Social Engineering Toolkit (SET) to launch more refined attacks.

This section discussed various approaches social engineers can use in order to carry out their SE attacks successfully. State-of-the-art SE attacks are conducted on SNSs (Krombholz et al., 2015). The socio-technical approach was subsequently found to fit this study. The next section discusses the persuasion techniques that social engineers typically use in phishing emails and how these can also be used on SNSs to effectively target psychological aspects of the user in order to manipulate them into responding.

2.4 Persuasion techniques on social network sites

In discussing the various SE approaches in the previous section, it was found that the effectiveness of such lies predominantly in the social psychology domain, particularly as highlighted in the work of Cialdini (2007). Educating users on how to identify phishing by means of grammatical inconsistencies, spelling errors and information misplacement (Blythe et al., 2011) has been effective to a certain extent, however these protection methods have not changed the fact that users continue to fall victim to phishing attacks (Ferreira et al., 2015). Krombholz et al. (2015) point out that users’ awareness of SE in SNSs is still comparatively low as compared to emails. According to Gragg (2003) any education on SE must include

psychology and persuasion in order to understand SE and counter attacks. Like in email-based phishing, persuasion principles can also be effectively executed on SNSs. There has been a lack of research describing and illustrating examples on how phishing can be carried out on SNSs, as most studies related to SE have focused on email and not on SNS (Algarni et al., 2017). Persuasion is a key element in how people choose to respond to messages (Workman, 2007) and humans by nature are susceptible to persuasion and, depending on the circumstances, resisting it is almost impossible (Bullée et al., 2015). The use of persuasion to influence users on SNSs does not necessarily have to be perceived as only used by phishers. The nature of SNSs encourages their members to self-promote or self-brand, acting in a similar way to celebrities for other users to follow (Khamis et al., 2017). This has given rise to the term “social media influencers”, who are prevalent on SNSs (Khamis et al., 2017).

Regarding the phishing messages found on Facebook, this section discusses and illustrates Cialdini’s (2007) six key principles of persuasion, namely, reciprocity, commitment or consistency, conformity or social proof, authority, liking, and scarcity. As the current study investigates social network phishing, the phishing messages or stimuli presented next demonstrate how effective each of these principles can be in persuading users to perform certain actions on SNSs. The posts on SNSs can be used easily and effectively to attract victims to respond to phishing links (Vishwanath, 2014). The images used to represent each of the persuasion principles in this section are real-world cases personally received by the researcher and are explained next.

2.4.1 Authority

The authority principle is the most used persuasion technique in phishing (Akbar, 2014) and is found to be the most effective strategy (Butavicius et al., 2015; Lin et al., 2016). People have a tendency to trust and obey the requests of authoritative figures (Bullée et al., 2015). As such, messages designed to appear as if they originate from an authoritative or trustworthy entity (e.g. a bank, or from the recipient’s employer or a friend) may persuade users to feel obligated to obey or respond to requests. This is because social learning encourages people not to question authority and therefore they are conditioned or may feel obligated to respond (Ferreira et al., 2015). Butavicius et al. (2015) and Williams et al. (2018) found that the presence of authority cues increased the likelihood that a user would click on a suspicious link contained

in a phishing email. On SNSs, this technique may be effective if the attacker has created an attractive profile or page, with fabricated information intended to make it appear legitimate. The fake profile may also have many followers, mutual friends, recent updates and interesting photos, thus increasing the user's trust. A study by Valecha et al. (2015) found that messages (i.e. Tweets) posted by Twitters users determined as having a high source credibility were more likely to be shared. Alternatively, the attacker could impersonate a public figure or organisation, clone a profile or pretend to be someone that the victim might trust or be interested in (Stajano & Wilson, 2011). An earlier study by Jagatic et al. (2007) found that subjects were more likely to respond when the phishing email appeared to have been sent by a friend.

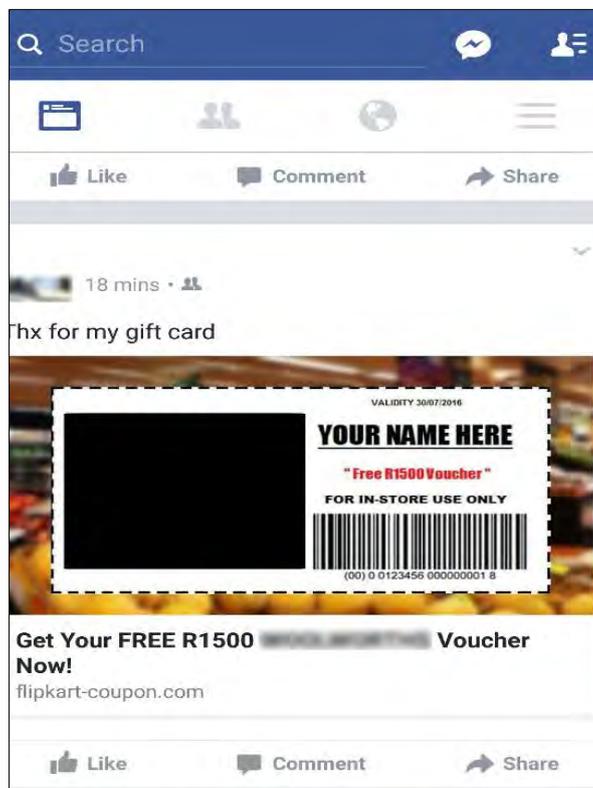


Figure 2.4: Authority principle applied in Facebook

Perhaps another reason why people would comply or respond to an authority is the awareness of potential consequences (i.e. the fear factor) that can follow from disobedience (Workman, 2008). In some instances, these consequences could be breaking the law or not complying with an organisation's policy.

2.4.2 Conformity or social proof

People tend to be influenced by others and will thus attempt to mimic what the majority of people do or are perceived to be doing (Bullée et al., 2015; Ferreira et al., 2015). The tendency to imitate the behaviour of members of a group is known as social proof. People will comply with a request if they see others have also complied (Cialdini, 2007). This may be in order to also accept choices that are perceived to be correct in order to fit in or be accepted by others in the group. For example, the user could share a post that appears to have good intentions for the recipient and broader network or it could be a post generating awareness of something. People generally will have higher levels of trust in people who share similar opinions to them especially in ambiguous situations (Uebelacker & Quiel, 2014). If others are behaving in a certain way, possibly a risky or a negative way, they may believe that by doing the same they cannot be held solely accountable for their actions (Ferreira et al., 2015). This is a typical example of the “safety in numbers” perception. In South Africa, the sharing of a helpful post associated with scholarships, internships and career opportunities is often accompanied with the catchphrase “puff and pass”.



Figure 2.5: Social proof principle applied in Facebook

In Figure 2.5, the message preys on users' Christian beliefs, as the message includes an image of Jesus and requests the user to "share me". In view of the sentiments expressed in the message, it may go against subjective norms for users to choose to ignore such a request. The overlapping usage of the *reciprocity* technique is also evident in this example.

2.4.3 Reciprocity

The reciprocity persuasion principle states that people are typically bound or driven to repay a debt or favours (Cialdini, 2007). Reciprocity is a strong social norm; when someone does something for you, one may feel obligated or have a sense that one must do the same for them in return (Uebelacker & Quiel, 2014). This principle can be effective if the victim is offered something first so that they are made to feel indebted to the person. For example, a message is made to appear helpful and warns the user that there is a possibility that their Facebook account could be hacked. This may be even more effective if the offer is perceived as something exclusive or personal to the victim such as a prize or reward. In addition, the message might also request the user to share this warning with others. Facebook responses such as complimenting, commenting on or liking another user's posts can contribute towards developing a relationship between users, thus encouraging them to accept each other's requests (Algarni et al., 2014). For example, a notice of a death on Facebook draws the attention of friends who give their sympathies and condolences. Others might feel obligated to do the same for fear that if they do not, they are being insensitive thus lending itself to the *social proof* technique. Figure 2.6 gives an example of how this technique is used in Facebook messenger.

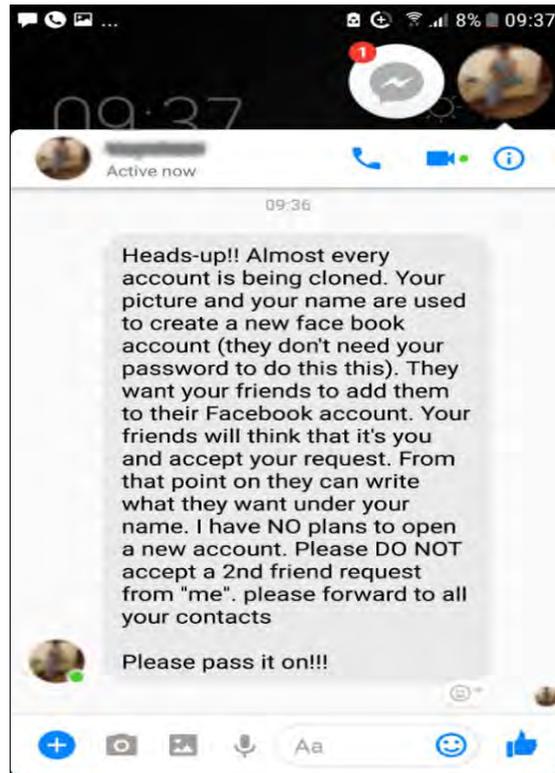


Figure 2.6: Reciprocity principle applied in Facebook Messenger

Figure 2.6 is not considered an example of social network phishing but of the hoax messages typically found on Facebook. However, such hoax messages could effectively lead to phishing if the user is requested to click on a link with a message stating, “Click on the link to see if your profile has been hacked too”. Users may comply with the instruction especially if the message originates from a trusted friend.

2.4.4 Commitment or consistency

For convenience, it is human nature to make a single decision and then adhere to that choice for all subsequent related choices. The commitment principle refers to the likelihood of dedicating oneself to a cause or idea after making a promise or agreeing to something (Cialdini, 2007). Our choice to commit to something compels us to remain consistent in our original decision because we will encounter personal and interpersonal pressure to behave consistently in line with that commitment (Cialdini, 2007). Moreover, people will be more confident in their decision to commit especially if they make it known publicly (Ferreira et al., 2015). In

this regard, SNSs could be perceived by users as the appropriate platform to make known to others their commitment to something. In this way, being committed to that choice may arouse a sense of personal fulfilment. In addition, new commitments can also be built on prior commitments made.

2.4.5 Scarcity

People by nature assume that when products, services or items are scarce or difficult to obtain, they are usually better than those that are easily available. As pointed out by Uebelacker and Quiel (2014), this corresponds with reactance theory which suggests people react only because their freedom to choose has become rare or limited (Brehm, 1966). To trigger a reaction from people, “urgency” cues are usually incorporated in the scarcity technique. By doing so, this puts the user under pressure to act or respond quickly and distracts the user from the main motive behind the message. Ferreira et al. (2015) include “distraction” under this principle, for example to have the availability of a product or service limited by its quantity, price or period of time. Typically, the user may respond in order to avoid missing out on an opportunity, a product, a service, or information, especially if it has limited availability (Bullée et al., 2015). Vouchers, coupons and one-of-a-kind special offers are just some of the ways this technique can influence users. Alternatively, phishers could use “fear” cues to create urgency (Workman, 2008). In this regard, urgency can be enhanced by adding a consequence or a timeframe to the message (e.g. a special discount or a prize valid for a certain period), as seen in Figure 2.4

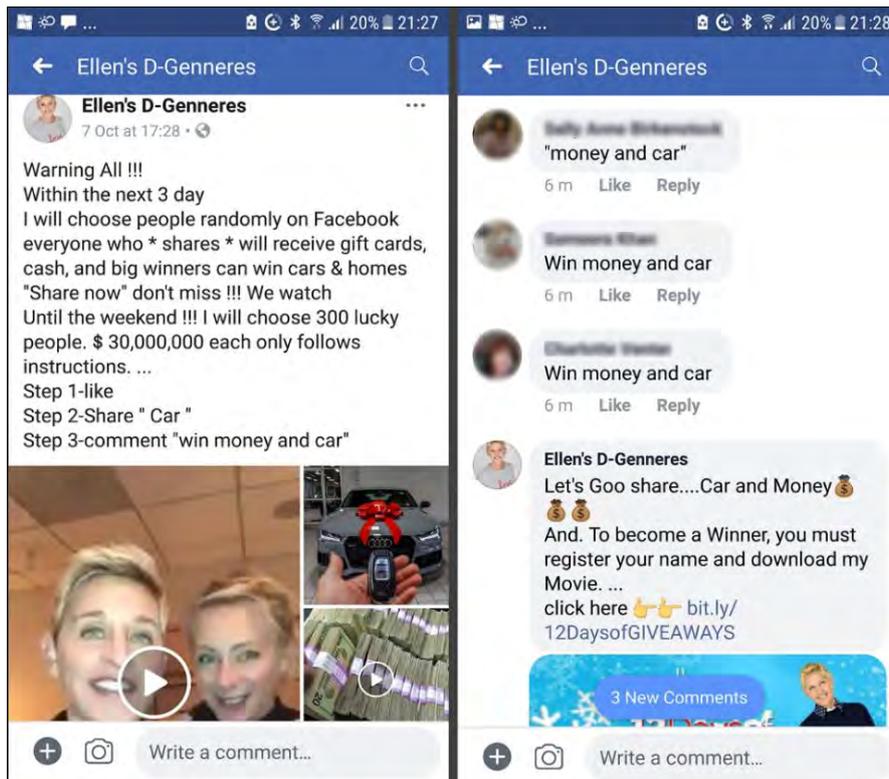


Figure 2.7: Scarcity principle applied in Facebook

The Facebook post in Figure 2.7 is further enhanced by impersonating the internationally known American comedian, Ellen DeGeneres, thus taking advantage of the *authority* principle. Social network users complied with this request by commenting and received further instructions to register their name and to download a movie. In order to become a winner, users were required to click on the shortened URL link concealing the site that the user was directed to. It is apparent that the incorrect spelling of Ellen's name and the lack of the verification badge did not affect the trust of the respondents. Accordingly, these users responded quickly (in the form of comments) as indicated by the response times. In this particular example, the level-2 attack described by Vishwanath (2017) is evident, as the link provided in the message could lead to malware installed by the victim. However, educating users that scams typically originate from celebrities offering money or offers too good to be true also has its limitations. For example, Figure 2.8 shows a legitimate Facebook post by famous Hollywood actor Patrick Dempsey offering an opportunity to win a luxury vehicle and \$20 000.



Figure 2.8: Legitimate post using the scarcity principle applied in Facebook

In Figure 2.8 the verification badge/mark, which is a blue checkmark icon next to the profile name, is one of the indicators that distinguish the official accounts of public figures, organisations or brands from fake accounts.

2.4.6 Liking

People typically like or prefer to be associated with people who are similar to them in terms of personal interests, attitudes, beliefs, opinions, backgrounds, personality types and so on (Bullée et al., 2015). A study by Lawson et al. (2017) found the liking persuasion principle was considered trustworthy in both phishing and legitimate emails. As users typically judge others by what they see, attributes such as physical attractiveness may be associated with other traits such as trustworthiness, friendliness and honesty. In terms of persuasion, people may be persuaded to obey others if they display any of the aforementioned characteristics that they regard as favourable or are familiar with (Ferreira et al., 2015). For example, a Facebook user may receive an invitation to accept a friend request but before accepting the request, he or she may seek information on the sender in relation to the number of friends they have in common,

photo albums, occupation and where they live. Developing persuasive messages that are of interest is one of the starting points in getting a user's attention. For example, in a phishing experiment, Chowdhury and Chakraborty (2014) used popular movies, incorporated in a phishing email and spoofed website, to effectively deceive university students into responding. SNSs provide an environment that encourages "liking", as there are built-in features that allow the user to indicate their support for posts by means of a reaction, such as "like" or emotion indicators or emoji's, that is visible to other members. According to Mattke et al. (2020), when a user observes sponsored content on SNSs, he or she notices who has "liked" the content which subsequently influences them to click, especially if the person is a close friend. The "like" feature is familiar to most SNS users as it is also incorporated into other popular SNSs such as LinkedIn, Instagram, YouTube and Strava.

Phishers take advantage of current affairs and controversial news and events reported on social media, thus preying on users' interests and *curiosity*. Krombholz et al. (2015) note that "curiosity" is a technique overlooked by Cialdini (2007). However, curiosity has been equated with an openness to experience personality trait (McElroy et al., 2007). This technique has also shown to be effective as, in a study by Pfeiffer et al. (2014), participants reported being curious to see where the phishing link would lead. Moreover, a message could prey on users' religious beliefs, political views and special causes, or have similarities to those of the user. Figure 2.9 shows how the curiosity technique can be employed on Facebook messenger.

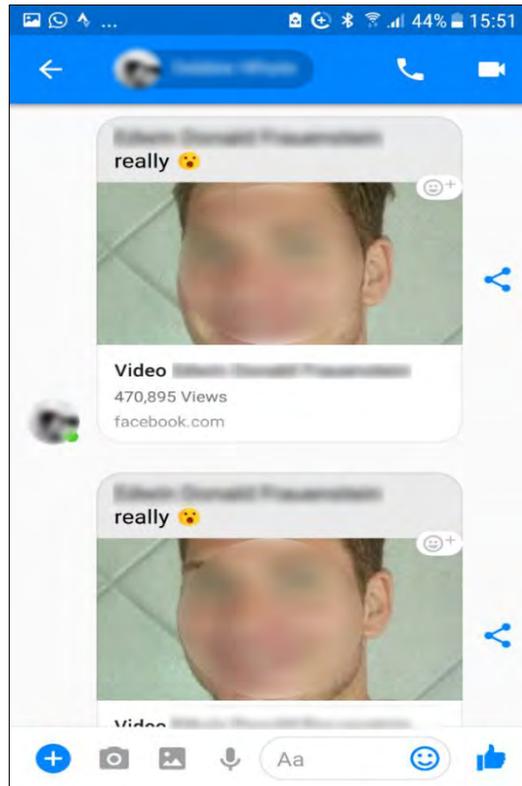


Figure 2.9: Curiosity principle applied in Facebook Messenger

Heartfield and Loukas (2015) classified this type of an attack as an instant message phishing. The effectiveness of this technique is enhanced by visual cues, as the message includes the statement “really” with a shocked emotion icon, as well as an exact image of the victim’s current profile picture. It also prompts the user’s attention and creates urgency, as it indicates that hundreds of thousands of users have already viewed the video and was sent twice in a row to the victim. Although not considered to be part of Cialdini’s persuasion taxonomy, this technique also uses “fear” in order to create urgency (Workman, 2008). Interestingly, user training interventions have made use of “fear appeals” as a means to counteract phishing attacks (Jansen & Van Schaik, 2018; Schuetz et al., 2016; Schuetz et al., 2020).



Figure 2.10: Curiosity principle applied in Facebook

Figure 2.10 is a fake post mimicking a video, which is often referred to as click-bait (as discussed in Section 2.4.7), and is used to draw the attention of Facebook users through the use of graphic images and controversial or enticing subject lines. Images used in fake video posts are usually disguised as a video, which often includes an image manipulated by photo editing software. While the user may be interested in clicking to watch the video only, malicious content may be downloaded or the user may be drawn into giving personal information in order to be able to watch the video. Drawing on protection motivation theory (PMT), another way in which the attacker could attempt to persuade users is to induce fear by means of a scenario in which the victim is convinced into believing they could in some way be personally affected.

Apart from the above-mentioned persuasion techniques, Petty and Cacioppo (1986) found that persuasive messages are more effective if the message is of relevance to the user and if it surprises them. Furthermore, the context in which persuasive techniques are executed can also play a substantial role in the success of an SE attack (Bullée et al., 2015). Moreover, if the persuasion principles are used in combination it may influence the way in which the user

responds. For example, Lawson et al. (2017) found that a combination of authority and scarcity persuasion principles was most likely to arouse suspicion in relation to phishing emails. As a result, it is difficult to identify which of the persuasion techniques users in general are more likely to fall victim to.

2.4.7 Other persuasion techniques and threats on social network sites

There are various other forms of cyberattacks conducted on SNSs that incorporate SE techniques. Apart from phishing, organisations are also faced by other forms of social media attacks such as clickjacking, financial abuse, identity theft, impersonation and physical crime (Adewole et al., 2017; Algarni et al., 2017). Facebook users are more at risk from SE attacks that can lead to more focused attacks, such as spear phishing and clickjacking (Algarni et al., 2017). Social media users are also exposed to other risks such as cyberbullying (Chan et al., 2021), sexting (Siegle, 2010), unprofessional behaviour and embarrassing photos (Garner & O’Sullivan, 2010), the spreading of hoax messages (Tacchini et al., 2017), publicly sharing location-based information (Kim, 2016), the spread of dangerous pranks and games (Branley & Covey, 2018), identity cloning (Fire et al., 2014) and identity theft (Reznik, 2013). Technical attacks may also be conducted on SNSs, Chi (2011) points out that phishing, inadequate authentication controls, cross-site request forgery (CSRF), cross-site scripting (XSS), information leakage, injection flaws, information integrity and inadequate anti-automation are some of the social media threats facing organisations. Posts on social media have the potential to damage the reputation of organisations, as user reactions could be influenced if they are in a highly emotional state such as a “hot state” (Wang et al., 2011).

Some of the abovementioned attacks include elements of persuasion. For example, clickbait is exaggerated, provocative or sensationalised headlines taking advantage of cues aimed at attracting or enticing the user to “click” on them (Agrawal, 2016). Clickbait is regarded as a form of tabloidization and has been implicated in the rapid spread of rumours and false information online (Chen et al., 2015). Clickbait may also present itself in the form of a game, an attractive quiz or survey which could request personal information from the user. Like the persuasion strategies discussed earlier, its strength is that it preys on the curiosity of the user by invoking emotion and triggering interest. Clickbait is a serious concern to the extent that in

August 2014, Facebook announced that it was taking technical measures to reduce the impact of this threat on its site (Visentin, 2014).

Like phishing emails, clickbait can entice users to click on malicious links, a process referred to as clickjacking. To elaborate, as webpages can have several overlays, clickjacking aims to tricks users into clicking invisible elements on a web page to perform unintended actions that can benefit the attacker (Kim & Kim, 2015). These elements appear as an object to the naked eye, such as a button, video or photo, but are in fact an overlay that the user did not intend to click on; for example a button calling on the user to endorse or “like” a product, follow a Twitter account, or send spam to your friends and networks. This includes enticing a user to download a harmful attachment which installs a malicious backdoor program that allows the attacker full access to the victim’s system, or clicking on an embedded hyperlink. Literature has been focused on addressing clickjacking on smartphone devices, in particular those with Android operating systems (Possemato et al., 2018; Wu et al., 2016). Vishwanath (2014) investigated farcing, which is considered a type of phishing attack whereby in the first phase, attackers use a phoney social media profile to befriend victims and, in the second phase, aim to acquire personal information directly from them through the fake profile. This is a form of information harvesting where valuable information can be gathered from the targets’ posts, tags or comments and also to the targets’ associated groups, events and pages.

As web browsers support “tabbed browsing”, this allows users to open multiple webpages, in the form of tabs within a single browser window. Tabnapping is a combination of the words “tab” and “kidnapping” and is a computer exploit and phishing attack, which persuades users to submit their login details and passwords to popular websites by impersonating those sites and convincing the user that the site is genuine (Chiew et al., 2018). This attack takes advantage of the users’ lack of attention more successfully when the user has opened several tabs and forgets the webpage they were using.

The activity of copying and pasting content from social media posts into a user’s profile, as inconspicuous as it may seem, also presents risks. According to Steinberg (2017), copying and pasting content offer several advantages to threats over sharing content. Hoaxes being “shared” are prone to being deleted – either by the victim when they eventually realise it is a hoax, or by it being reported to Facebook to be removed by the removal of a single post. However, copying and sharing a hoax post (which can be phishing or malware) makes it difficult to trace,

thus making it more difficult for Facebook to associate the new posts with the original hoax. Moreover, Facebook's algorithms to detect these specific posts stand a greater chance of missing a copy-and-paste version than they would an original post that is shared. Steinberg (2017) also adds that when you re-share a post from someone else, the post is subject to the original post's privacy settings as well as yours.

The above-mentioned techniques have been recognised as some of the common threats in the present SNS environment. As SE is just a technique, it is impossible to anticipate all types of threats and its attack vectors, as many different types of attacks can be launched using SE techniques.

This section described how various forms of persuasion can be employed on SNSs. The persuasion principles embedded in each stimuli presented in this section contain certain attributes or cues which could influence the user to trust or act on the message. In this regard, when users attempt to evaluate the message, they would typically resort to performing "heuristic" processing (Butavicius et al., 2015). As a result, this can make them susceptible to phishing. Heuristic processing is further discussed in Chapter 4. The next section discusses human factors that can influence behaviour, sometimes referred to as dispositional factors. Dispositional factors (internal factors) are individual characteristics that influence behaviour and actions in a person. McElroy et al. (2007) points out that dispositional factors include self-efficacy, locus of control, personality factors, and cognitive style and so on. The characteristics of each of these factors can also make users more vulnerable to SE techniques.

2.5 Habit

Gardner (2015) defines habitual behaviour as "any action, or sequence of actions, that is controlled by habit". Habit is a learnt behaviour that can be triggered by a particular context. Habits can begin as intentional and goal-directed behaviour (e.g. eating particular foods for weight loss) which could ultimately develop into a healthy routine. Habits are widely considered to predict and explain human behaviour in relatively stable situations (Aarts et al., 1998; Gardner, 2015). According to Verplanken and Aarts (1999), habits are "learned sequences of acts that have become automatic responses to specific cues, and are functional in obtaining certain goals". In the context of using information systems, habit refers to the "extent to which people tend to perform behaviors automatically because of learning" (Limayem et

al., 2007). In the information systems (IS) domain, prior literature on the influence of “habits” has focused on many areas, including password management (Florencio & Herley, 2007; Friendman, 2014; Stobert & Biddle, 2016), social media usage (LaRose, 2015; LaRose et al., 2003; LaRose et al., 2011; Mouakket, 2015; Thadani & Cheung, 2011; Turel & Serenko, 2012; Vishwanath, 2015b), information systems usage (Limayem et al., 2007), payments using e-commerce (Dahlberg & Oorni, 2007), e-commerce continuance (Liao et al., 2006), mobile internet usage (Venkatesh et al., 2012), mobile SNS usage (Yang et al., 2016) and information security policy compliance (Vance et al., 2012). Recently, habits have emerged as an area of interest in phishing research (Alqarni et al., 2016; Frauenstein & Flowerday, 2016; Vishwanath, 2015a, 2015b; Vishwanath et al., 2011; Vishwanath et al., 2018). While no pure theory exists for habit, the Triandis model of Interpersonal Behavior (TIB) is a theory that includes “habit” as a significant, and the potentially most important, influence on behaviour (Triandis, 1977). Robinson (2010) reports that various scholars have noted that making use of the TIB, in particular habit, increases the model’s predictive power over other behavioural models such as the widely popular TPB.

According to Vance et al. (2012), prior research has investigated habits from three perspectives: the moderating influence of habit on intention and IT usage, the direct effect of habit on IT usage, and the direct effect of habit on intentions to use IT. The more frequently a habit is performed, the more *automated* the choice process often will become (Jager, 2003; Triandis, 1980). As a result of this automaticity, habits allow people to efficiently allocate their limited cognitive capabilities, thus significantly saving on cognitive effort (Jager, 2003). However, this study posits that this advantage may put users at risk, as this may result in users overlooking potential phishing messages on SNSs. This is supported as Vishwanath (2015a), who found that users who habitually engage on Facebook are more at risk of falling victim to a social media phishing attack. According to Wood and R nger (2016), two defining features of habit automaticity are 1) activation by recurring context cues and 2) insensitivity to short-term changes in goals, including changes in the value of response outcomes and the response-outcome contingency. In the context of the present study, on SNSs these situational cues might for example be “liking”, “sharing” and clicking on malicious links that appear on the users’ social network timeline or newsfeed. According to Wood and R nger (2016), habits strengthen through associative and reward-learning mechanisms that capture the slow incremental nature

of habit formation. Similarly, Aarts et al. (1998) state that perceived gratification can form habits; this is discussed in further detail next as it forms part of the antecedents of habits.

2.5.1 Antecedents of habit

In the applied behaviour analysis discipline, antecedents are regarded as stimulus events, action(s), situations or circumstances that precede the behavioural response (Miltenberger, 2016). It is thus important to identify the antecedents of a particular behaviour as they can reveal the circumstances of how and why such a behaviour is stimulated (Miltenberger, 2016). Thadani and Cheung (2011) investigated habits in the IS literature and found three primary antecedents that contribute to habit formation, namely, *satisfaction*, *comprehensiveness of usage*, and *frequency of prior behaviour*. Similarly, Vishwanath (2015a) identified three antecedents of “Facebook” habits, namely, *consumption frequency*, *gratifications*, and *automaticity*. The TIB, alluded to earlier, also include the frequency of past behaviour as an antecedent of habit. According to Vance et al. (2012), habits are often determined with a measurement of past behaviour or behavioural frequency. However, Ajzen (1987) and Verplanken and Orbell (2003) argue that past behavioural frequency is not entirely an accurate predictor of future behaviour. To elaborate, Ajzen (2011) posits that it is not the frequency of behaviour that determines the strength of a habit, but the extent to which the behaviour has been automated and is being performed without cognitive elaboration.

Habit has mostly been viewed as a positive outcome to improve IS usage (Turel & Serenko, 2011). While IS usage can facilitate the development of a habit (Limayem et al., 2007), SNSs have the potential to result in negative consequences such as addiction (LaRose et al., 2011). Turel and Serenko (2011) investigated habits in a SNS context and found that adverse outcomes such as technology addiction and high engagement are augmented by habit. Similarly, Thadani and Cheung (2011) found that dependency on SNSs facilitates the formation of habits. According to Vishwanath (2015b), the Facebook platform encourages repeated interaction. Motivations around this repeated interaction may be as a result of the gratification stemming from their usage, which might potentially lead to addiction (Kuss & Griffiths, 2011).

The automaticity aspect of habits offers an advantage by reducing the cognitive effort needed to perform certain activities (Jager, 2003). In this regard, Turel and Serenko (2012) refer to people as “cognitive misers”. Vishwanath (2015a) uses an example to explain that habit can be viewed as a lack of awareness or attention and performing the habit could range from being completely conscious (i.e. aware) and intentional at times (checking email on a phone) to being unconscious and automatic at other times (checking email while driving).

A variety of cues may *trigger* habit performance, including aspects of the physical environment, other people, and preceding actions in a sequence (Wood & R nger, 2016). Such situational cues may thus trigger the performance of a habit automatically in that context (Verplanken & Aarts, 1999; Wood & Neal, 2007). Moreover, habits can be formed and strengthened through associative and reward-learning mechanisms (Wood & R nger, 2016). For example, the gratification that arises from frequent participation on SNSs can develop into a habit (LaRose et al., 2011). This habit becomes reinforced if the user experiences an enjoyable emotion when interacting on SNSs (Yang et al., 2016). This supports Turel and Serenko’s (2012) finding that on SNSs, perceived enjoyment is a key antecedent of habit. In this regard, some features or activities on SNSs could stimulate the gratification gained from receiving appreciation for sharing a post, compliments on photos and suchlike. The outcomes stemming from these types of behaviour may encourage a user to continue performing the behaviour, thereby forming and reinforcing the habit. In some cases excessive use of the habitual behaviour can develop into an addiction such as a Facebook addiction (Balcerowska et al., 2020) and fear of missing out (FOMO) (Kuss & Griffiths, 2017). Habits may develop in a specific situation, and thus be limited only to that situation (Verplanken & Aarts, 1999). In the context of this study on SNSs, situational cues may, for example, be boredom, low self-esteem, app notifications, or FOMO. As a result of the user’s repetitive and frequent interaction on SNSs, it is possible for habits to develop from seemingly innocuous behaviours. Examples of such behaviours on SNSs include:

- Seeking attention, recognition or rewards by posting messages, photos and videos, and checking in, and then regularly checking to see if members have responded or reacted to the post in the form of an emotion (e.g. likes and emoji’s), comments or receiving newly added followers/subscribers.
- Persistently “liking” or sharing posts on social media platforms.

- FOMO, leading to regularly checking to see what others have posted (i.e. updates, news), checking the messenger inbox, and checking out the notification icon.
- Seeking self-beneficial opportunities, for example goods for sale and discounted goods, entering competitions, companionship and employment opportunities.

Most of these aforementioned behaviours appear as self-promoting and superficial, and are associated with the narcissism trait – one of the Dark Triad personality traits (Mehdizadeh, 2010). These behaviours could also, to a certain extent, relate to users with low self-esteem who will thus be more engaged and active in promoting themselves on Facebook to fill this void (Mehdizadeh, 2010). Moreover, these behaviours also highlight the three primary antecedents of habit, namely satisfaction, usage, and frequency of prior behaviour (Limayem, et al., 2007). The reflexive or habitual liking of posts presents another concern – “like-farming” also known as “like-harvesting”. Like-farming is performed mostly on Facebook and is used by commercial entities and scammers to raise the popularity of a site with the intention of making it go viral around the world (Arntz, 2019). Similar to click-bait, like-farmers rely on users to click habitually or share attention-grabbing topics such as animal welfare, competitions and the like. Phishers can transform these popular posts so that they lead to spoofed websites and trick the user into divulging personal information. Users could also unknowing put themselves to privacy-related risks and at risk for other forms of crime through their habitual behaviour, including using the Facebook “check-in” feature or simply sharing location-based information relating to places they intend to go to (Kim, 2016). For example, in 2008, seven American youths, popularly known as the “Bling Ring”, used Twitter to track the information Hollywood celebrities were sharing when they went out to events, and used Google Earth to identify where the celebrities lived. This information was used to break into several celebrity homes and the thieves, who were later convicted, stole approximately \$3 million in cash and other belongings (Corner, 2012).

This section discussed habitual behaviour on SNSs that could lead to users becoming primed to behave in a certain way without much thought about performing the behaviour concerned, because the actual behaviour has to a large extent been automated. While this may have the beneficial effect of freeing people’s cognitive abilities, thus allowing them to direct their focus to other activities, it can present other risks as users may overlook suspicious messages owing to this habitual influence on their behaviour. As a result, habit could also

result in users processing message superficially (i.e. heuristic mode). This potential relationship is discussed further in Chapter 4. This study posits that habit (i.e. sharing and/or clicking of posts) may present a potential risk to users, as their lack of awareness may catch them off-guard and cause them to overlook phishing messages. Habit is therefore an important consideration in the current study.

2.6 Perceived risk

According to Rehman et al. (2020), there is currently no universally accepted definition of perceived risk. Perceived risk can be described as a dispositional factor that can help predict an individual's likelihood of accepting risk in an uncertain situation and context. Pavlou (2003) defined perceived risk as "the subjective belief that there is some probability of suffering a loss in pursuit of the desired outcome". Risk perceptions are domain-specific and can be either optimistic or pessimistic (Ferrer & Klein, 2015). People have a tendency to have biased insights, overestimating the likelihood of positive events and underestimating the likelihood of negative events – known as optimism bias (Sharot, 2011). Similar to "optimism bias", users may be aware of security threats and the techniques associated with phishing, but may be of the view that falling victim could not happen to them personally (Kim & Hancock, 2015). As such users may perceive risks as distant and as such believe that organisations and others are the main targets of attackers instead of them (De Bruijn & Janssen, 2017; West, 2008). This perception of "it won't happen to me" (Krasnova et al., 2009) can potentially put users more at risk of security threats as they are less prepared to deal with them. In addition, some users may be overconfident about their ability to detect phishing messages (Moody et al., 2017) and the perception of risks may vary depending on the product or the service (Featherman & Pavlou, 2003). For example, smartphone users tend to be complacent in their security behaviour and demonstrate high levels of trust towards their apps (Ophoff & Robinson, 2014). This also poses a risk, as SNSs, in particular Facebook, are mainly accessed via smartphone apps. These examples show that people differ in their perceptions of risk and their reactions, particularly if a threat involves a perceived shortage rather than a direct personal threat (Dowd & Seibel, 1990).

Perceived risk or perceived susceptibility is typically viewed as a product of two variables: the perceived likelihood of an event and perceived damage if the event takes place (Krasnova

et al., 2009). According to Cox (2012), perceived severity and perceived vulnerability are both components of the threat appraisal process. When making behavioural decisions, individuals will often decide based on their estimates of the risks associated with the various options (Parsons et al., 2011). This is supported by empirical studies that have suggested that when people perceive a threat as severe and likely, they undertake measures that they think will be effective in preventing that threat from causing them harm (Herath et al., 2014).

According to Jacoby and Kaplan (1972), there are six types of perceived risk, namely, financial, performance, social, physical, privacy, and time loss. For example, some users might not perceive any physical risk when engaging in online banking, but on the other hand, they may perceive financial and/or privacy risks if their personal information falls into the wrong hands. In the context of information security, earlier studies conducted by Huang et al. (2007) and Pattinson and Anderson (2005) identified several factors that influence users' perceptions of risk. In a survey consisting of 602 participants, Huang et al. (2007) found that perceptions of risk are influenced by knowledge, impact, severity, controllability, possibility and awareness. The perceived danger of threats was significantly higher when there was little knowledge of the risk, the potential impact of the risk was high, the potential severity of the risk was high, or there was a greater possibility of the risk occurring. Van Schaik et al. (2017) found that perceived risk was highest for identity theft, key logger, cyberbullying and SE. Pattinson and Anderson (2005) suggest that perceptions of security risks are generally influenced by factors such as the individual's disposition at the time, recent media reports, past experiences and prior knowledge of technical aspects. In a particular situation, an individual's perception of risk affects their confidence in their judgement and, as such, will lead to them adjusting their behaviour to be more cautious if their perception of the potential damage or danger increases (Davis & Tuttle, 2013; Workman, 2007). Similarly, Rogers' (1975) PMT postulates that when individuals perceive that they are more at risk of security threats, and when the threats are more severe, they are more likely to adopt a recommended response to the threat and adjust their behaviour according to the amount of risk they are willing to accept or tolerate. This "trade-off" is also known as risk homeostasis (Wilde, 1998).

The emergence of SNSs has uncovered new risk factors, with caution, online security risks and privacy being the most investigated (Rehman et al., 2020). A study by Cheung et al. (2015) found that users focus on the benefits as well as social influence when they decide to reveal personal information about themselves on SNSs, but pay less attention to the potential privacy

risks. Facebook users, in particular, have a greater sense of trust compared to users of other popular SNSs (Fogel & Nehmad, 2009; Kim & Hancock, 2015). This may be also attributed to the fact that SNSs often do not present any security warnings to users.

In a study by Choi, Yoo et al. (2017), it was found that heuristic-systematic processing and self-efficacy had an influence on the risk perceptions of SNS users. Vishwanath et al. (2018) used cyber-risk beliefs as a construct in their model and proposed that individuals are more likely to thoroughly assess (systematically process) emails when they perceive their cyber actions to be risky. In contrast, Vishwanath et al. (2018) found that individuals resort to heuristic processing when they perceive a lower need for confidence in their decisions because the outcomes of their actions are seen as less risky. A more detailed discussion on heuristic and systematic processing is provided in Section 4.3.

From the various types of risk discussed in this section, it is apparent that SNSs are platforms that can present users with a number of risks, including financial, social and security/privacy risks. With regard to the last mentioned, for example, prior literature has revealed that Facebook users will accept friend requests from strangers based on textual cues (Leow & Wang, 2019). On the other hand, users may be of the perception that they are able to exercise control in situations that present themselves as having risk. This will be discussed further in the next section.

2.7 Computer self-efficacy

Before computer self-efficacy can be understood it is necessary to first define the term “self-efficacy”. Self-efficacy refers to an individual’s belief or personal judgement in their capabilities to accomplish a task or goal in a given situation (Bandura, 1982). According to Cox (2012), self-efficacy is an important factor in determining a person’s use of information security tools. Perceived self-efficacy is concerned with people’s beliefs that they can exercise control over their motivation and behaviour and over their social environment (Bandura, 1990). Perceived behavioural control, which is a variable included in the TPB, originated from the construct of *self-efficacy* derived from social cognitive theory (SCT) (Bandura, 1977). As noted by Bulgurcu et al. (2010), self-efficacy is used instead of perceived behavioural control because the latter essentially measures the same latent construct as self-efficacy. People’s

beliefs about their capabilities have an influence on what they choose to do, how much effort they muster, and how long they will persevere if they are faced with challenges, along with the negative emotional experiences associated with those challenges (Bandura, 1990).

Computer self-efficacy refers to the user's judgement of their capability to use computers to achieve a particular purpose (Safa et al., 2015). Compeau and Higgins (1995) define computer self-efficacy as the “judgment of one’s capability to use a computer”. Computer knowledge and prior computer experience are considered to be influencing factors for computer self-efficacy (John, 2013). Wright and Marett (2010) classify computer self-efficacy as an experiential factor. Researchers have investigated the influence of computer-self efficacy on the adoption and use of SNSs (John, 2013; Wang et al., 2015), and Saleem et al. (2011) found that a particular gender and certain personality traits have an influence on computer self-efficacy in people. Hocevar et al. (2014) introduced the concept of social media self-efficacy (SMSE) and found the users in their experiment who were high in SMSE to be more at risk, as they tended to trust information posted by SNS users. Yao et al. (2007) found internet users with high computer self-efficacy to be confident in their abilities to handle online threats and secure their privacy. This was confirmed by studies that found that users with high computer self-efficacy are better at avoiding phishing attacks (Arachchilage & Love, 2014; Sun et al., 2016; Wright & Marett, 2010). However, this confidence can also present other risks, as high levels of computer self-efficacy can also lead to high levels of actual confidence about how to proceed when presented with system errors (Davis & Tuttle, 2013). This presents another concern as prior research has shown that users ignore phishing warnings, which might indicate that they are less motivated to pay attention to warnings, especially in environments they frequently engage in. Moreover, Vishwanath et al. (2011) found that computer self-efficacy was negatively related to users examining a phishing message closely. This indicates that individuals who consider themselves to be technologically sophisticated are just as likely to be phished as those who are not.

2.8 Social norms

The theory of planned behaviour (TPB) proposed by Ajzen and Fishbein (1980) is used to predict an individual's intention to engage in a particular behaviour at a specific time and place. The TPB, an extension of its predecessor, the theory of reasoned action (TRA), has been used

to predict intentions to perform behaviours of different kinds, including attitudes toward the behaviour, subjective norms and perceived behavioural control (Ajzen, 1991). “Subjective norms” refers to “the perceived social pressure to perform or not to perform the behavior” (Ajzen, 1991). Social norms, on the other hand, are typically defined as “rules and standards that are understood by members of a group, and that guide or constrain social behaviors without the force of law” (Cialdini & Trost, 1998). From these definitions it is apparent that social norms are a type of subjective norm (Ham et al., 2015; Marino et al., 2016). Social norms develop and evolve as a result of the interaction between individuals in social groups (Cialdini & Trost, 1998). Users may adjust their behaviour based on the perceived expectations of others (Dincelli & Goel, 2017). Accordingly, in the context of SNS, users may be pressured, or perceive themselves as being pressured, by their friends into performing certain behaviours such as sharing posts, responding to posts or liking posts. This explanation resembles the *reciprocity* principle of Cialdini (discussed in Section 2.4.3) which is related to normative commitment (Workman, 2008), as discussed in Section 2.4.4. For the purposes of marketing, some legitimate organisations on Facebook offer the public a chance to win a prize if they, for example, like their page, share the post with friends, copy and paste, or tag two or more friends. As a result, users become accustomed to following these “rules” and seeing this behaviour also performed by others. Such willingness to meet the expectations of significant others, increases the riskiness, especially if the posts are malicious. In addition, if the user has not made an effort to validate the authenticity of the message, it may lead to an increase in the spread of hoax messages.

2.9 Summary

This chapter addressed the first research sub-question of the study by determining the type of phishing threats social media users are exposed to. The chapter began by presenting the literature on SE and SE approaches. The literature review also pointed out that to generate accurate findings, the methods used to conduct phishing and/or SE-based experiments also bring ethical concerns and limitations and, as such, researchers have seen a need to provide better guidance. The literature has also revealed that phishing warnings have proven to be ineffective, which has led researchers to test and design better warning mechanisms to counter SE attacks that users will be more likely to respond to. However, on SNSs no warning indicators have been designed so far. Moreover, the literature has pointed out that there is no

guarantee that users will pay sufficient attention to or heed such warnings. The focus of the chapter was to demonstrate that the persuasion techniques typically employed in phishing emails can also be incorporated into posts (on timelines) and instant messages on SNSs. These persuasion strategies play an important role in influencing users into responding especially when employing the *authority* principle. On SNSs this technique can be very effective considering that users may be more likely to trust posts emanating from their friends and family, or institutions that they recognise and like. With regard to Cialdini's principles of persuasion, it was also found that there is no consideration for principles such as "curiosity" and "fear" which are often used in phishing and clickbait posts. It is acknowledged that not all users will fall victim to the persuasion techniques and, as such, this chapter included a discussion on dispositional factors (interpersonal factors) such as habit, perceived risk, computer self-efficacy and social norms, as these variables can potentially influence users' responses to phishing. The most important of these is habit, which can potentially override a user's rational thinking when coming across posts on SNSs. Recent literature on phishing has investigated the influence of personality traits on user behaviour, especially in the context of SNSs. Accordingly, as the phishing literature suggests, it is important to consider users' personality traits as another vulnerability factor, hence this is discussed in further detail in the next chapter.

3

PERSONALITY TRAITS AND THEIR INFLUENCE ON HUMAN BEHAVIOUR

3.1 Introduction

The previous chapter pointed out that, in comparison to the time, expertise and effort needed to use technical approaches to infiltrate information systems, the persuasion techniques employed in phishing are an effective means of exploiting human vulnerabilities to gain information from unsuspecting users. Moreover, the literature revealed that certain persuasion principles might influence particular types of user and not necessarily “all” users. Thus, instead of generalising phishing susceptibility to the type of phishing attack or persuasion technique users would most likely fall for, identifying and classifying users according to their personality traits is one way of better understanding the behavioural problem. This can be advantageous to organisations as they can design appropriate security strategies, such as training and awareness programmes, to address these “weak links” (Ki-Aries & Faily, 2017; Warkentin et al., 2012). This can only be achieved if it is based on a deeper understanding of “psychological profiles” (Warkentin et al., 2012). Moreover, the effectiveness of security awareness programmes may also in itself be influenced by personality traits, as Kajzer et al. (2014) determined that, depending on their personality traits, certain users are more or less receptive to information security awareness message themes. It is thus important to consider personality traits as performing an influential role in protecting users and organisations from phishing threats (Luo et al., 2011).

The chapter begins by introducing personality traits and giving a brief history of their measurement scales. This is followed by a description of the characteristics of each of the Big Five personality traits. Recent developments involving the use of personality traits in the information security domain are also presented. The chapter emphasises that personality traits

play an influential role in user behaviour, in particular social network users, when users are confronted with phishing. By doing so, this chapter addresses the second sub-question of the study.

3.2 Personality traits

SNSs are predominantly used as a tool for self-expression, with users frequently posting information about their daily activities, “selfies”, and their thoughts and opinions on a variety of topics. To an extent, the information that social network users share reveals certain things about them. Owing to the influence of social norms, some observable behaviours are common across different cultures and age groups, such as the use of emotion icons (or emoji’s) and internet slang (e.g. lol, omg, lmfao). However, while this might indicate that social media users tend to behave similarly, not all users behave in exactly the same way. For example, some users may post information frequently while others observe passively. Some users may be impulsive, clicking on links that attract their interest or attention, while others may do this when in an emotional state, under the influence of alcohol, or acting out on their frustrations (Wang et al., 2011). Some may be inclined to be helpful, accepting and trusting of others. Some Facebook users post messages with the intention of being perceived in a favourable light, but do not think about the unintended consequences of what they are posting and misjudge the culture and social norms of their social circles (Wang et al., 2011). These disparities in user behaviour have motivated the need to categorise behaviours more uniformly and has led to the development and adoption of personality trait scales. In view of this, it is not surprising that a large body of literature has investigated personality traits and their influence on social media usage (Amichai-Hamburger & Vinitzky, 2010; Correa et al., 2010; Mancinelli et al., 2019; Moore & McElroy, 2012; Ross et al., 2009; Ryan & Xenos, 2011; Wilson et al., 2010).

Personality traits describe individual differences in terms of characteristic thoughts, feelings and behaviours (Funder, 2001). Personalities are unique to each individual as they are predominantly determined by inheritance, social and environmental influence, and experience (McCrae & John, 1992). Common aphorisms such as “like father like son” or “the apple does not fall far from the tree” highlight the hereditary nature of traits that stem from family genetics (Jang et al., 1996). Personality characteristics are integral to the way humans think and behave, and therefore have an influence on whether or not an individual is likely, be it intentionally or

unintentionally, to become involved in malicious activities or risky behaviour (Nurse et al., 2014). Personality is considered a leading factor in understanding why people behave the way they do on the internet (Amichai-Hamburger, 2002). Personality traits are also influenced by gender differences which subsequently affect internet usage behaviour (Amichai-Hamburger & Ben-Artzi, 2000) and the likelihood of phishing deception (Darwish et al., 2012). A large body of literature has also investigated personality traits and their influence on social network use (Amichai-Hamburger & Vinitzky, 2010; Correa et al., 2010; Moore & McElroy, 2012; Ross et al., 2009; Ryan & Xenos, 2011; Wilson et al., 2010). Personality traits can also predict the security behaviour intentions of users towards protecting their computer devices (Gratian et al., 2018) and also has a significant effect on perceived trust and risk, and decision performance (Cho et al., 2016). Personality traits have also been found to be closely related to internet addiction (Kayış et al., 2016).

Research involving personality traits has been a topic of interest over several decades, with several rating instruments applied in many studies across various disciplines and contexts (Costa & McCrae, 1992a; John & Srivastava, 1999). Scholars, particularly in the psychology domain, continue to explore a variety of focus areas within personality trait research. For example, anxiety and anger, which are among the neuroticism personality traits, are positively associated with risky driving behaviour (Yang et al., 2013). In the information security domain, studies that involve personality traits have gained the interest of scholars, as certain traits are considered important predictors of human behaviour (Albladi & Weir, 2017; Gratian et al., 2018). Of the many types of personality scales scholars can adopt, the Big Five has been noted as the most widely accepted as it shows consistency across time, culture and age groups and is considered more structured, as the five traits do not overlap each other (Erder & Pureur, 2016). The Big Five model and a brief history of its scales is discussed next.

3.2.1 Brief history of the Five Factor Model

The five-factor model (FFM), consisting of the “Big Five” personality traits, is the most widely used and extensively researched model of personality (John & Srivastava, 1999; McCrae & Costa Jr, 1999). The five-factor model of personality is regarded as a leading theoretical model, and has gained popularity over recent years (Xu et al., 2020). It is comprised of the following empirically derived five factors or dimensions: openness, conscientiousness, extraversion,

agreeableness and neuroticism, which usually represented by the acronym OCEAN or CANOE. Known today as the “Big Five”, this model has resulted from decades of research with numerous improvements, refinements and iterations which have led to a wide array of personality scales. However, the wide array of scales has been criticised by John and Srivastava (1999) as having little guidance and no overall rationale. To elaborate, John and Srivastava (1999) argue that while these scales may use the same name, they often measure concepts that are not the same and moreover the scales that use different names often measure concepts that are quite similar.

There is a wide variety of existing personality scales related to the Big Five which researchers may adopt. According to Furnham (1996), the Myers-Briggs Type Indicator® (MBTI), developed by Myers et al. (1998), is the most popularly used personality instrument in the consultancy and training world. The MBTI consists of 93 questions and categorises the trait behaviour into four dichotomies, namely, Introversiion/Extraversiion, Sensing/Intuition, Thinking/Feeling, and Judging/Perception. From this, each person is said to have one preferred quality, producing 16 different discrete "types". However, the MBTI has received criticism in recent years stating that no significant conclusions can be based on this test (Erder & Pureur, 2016).

One of the earlier scales to measure the Big Five traits was the Neuroticism-Extraversiion-Openness Inventory Personality Inventory (NEO-PI). The NEO-PI scale included 180 items rated on a five-point scale; 48 items constituted each of the five domains and 18 items were used to assess agreeableness and conscientiousness (Costa & McCrae, 1985). The original version of the NEO-PI did not provide facet scores for agreeableness and conscientiousness. Subsequently, Costa and McCrae (1992a) revised their instrument and created the revised Neuroticism-Extraversiion-Openness Personality Inventory (NEO-PI-R). The NEO-PI-R consists of 240 items and measures the five traits and, in addition, includes six facets for each factor, with eight items corresponding to each facet being rated on a five-point scale ranging from 0 (strongly disagree) to 4 (strongly agree). The internal consistency of the NEO-PI-R was high (Costa & McCrae, 1992a). However, its disadvantage was that the scale was still lengthy and it required approximately 45 minutes for participants to complete the test (Gosling et al., 2003). McCrae and John (1992) devised the NEO Five Factor Inventory (NEO-FFI), which is a brief measure of the five personality dimensions and is composed of 60 statements with each of the five dimensions being assessed by 12 items. The term “Big Five” was coined by

Goldberg (1992), who assessed the five traits using a 100-item nine-point Likert scale. Later, Goldberg et al. (2006) developed a 50-item International Personality Item Pool (IPIP) scale, with ten items per domain. A shortened, 44-item Big Five Inventory (BFI) scale (used in the present study) was developed by John and Srivastava (1999). The original BFI was revised to form the BFI-2. According to Soto and John (2017), the BFI-2 introduced “a robust hierarchical structure, controls for individual differences in acquiescent responding, and provides greater bandwidth, fidelity, and predictive power than the original BFI”.

Owing to the large number of items that most personality scales measure, they require approximately ten to 15 minutes for participants to complete. As a result, researchers have recognised the need to further improve their measurement scales to more shortened versions, while still ensuring that their reliability and validity are maintained. For example, Saucier (1994) reduced the set of 100 items developed by Goldberg (1992) to a robust subset of only 40 items. Moreover, Gosling et al. (2003) developed a ten-item measure of the Big Five dimensions. This indicates that personality scales are continuously under refinement and thus it is expected that more will be developed.

3.2.2 Description of the Big Five personality traits

The characteristics of each of the traits of the FFM model of personality are described in Table 3.1 as follows:

Table 3.1: Summary of Big Five personality trait characteristics (adapted from Costa & McCrae, 1992b; John & Srivastava, 1999; Zhang, 2006)

Trait	Characteristics
Openness to experience	Open-minded, independent of judgement, inquisitive, intellectual, creative, seeks new experiences, open to risky behaviour, active imagination, non-judgemental, higher cognitive abilities, appreciation for art, nature and different ideas and beliefs
Conscientiousness	Honest, thorough, cautious, trustworthy, organised, hardworking, self-disciplined, responsible, strong-willed, goal-oriented, prudent, follows rules, standards and procedures
Extraversion	Friendly, enthusiastic, excitement, sociable, energetic, talkative, assertive, impulsive, dominant, needs stimulation
Agreeableness	Self-conscious, tolerant, compassionate, accepting, modest, polite, cooperative and trusting of others, respect for other people's beliefs and conventions
Neuroticism	Emotionally unstable, withdrawal, sensitive, volatile, impulsive, need for belongingness, experiences negative emotions (e.g. sadness, nervousness, anger, fear, pessimism, low-self-esteem, embarrassment, disgust and guilt)

Like most personality scales, the Big Five comes with its critics and contradictory findings. Studies have shown that the Big Five traits are universally held, regardless of the culture (McCrae & Terracciano, 2005). However, Rolland (2002) found that when examining the Big Five traits across 16 cultures that the *agreeableness* and *extraversion* traits were more sensitive to the cultural background of the individual, while *neuroticism*, *openness* and *conscientiousness* were fairly generalisable. Gender differences is another factor that has been found to influence personality traits between cultures (Costa Jr et al., 2001). Genetics has also been found to have an effect in determining personality, although research has not conclusively determined to what extent personality is genetically predetermined (Krueger et al., 2008). As

pointed out by Ardel (2000), Srivastava et al. (2003) and Terracciano et al. (2006), it was once believed that particular personality traits remain predominantly stable after the age of 30 and “set like plaster” and thus stop changing beyond this age (McCrae & Costa Jr, 1990). However, a more recent, ground-breaking, longitudinal study conducted by Damian et al. (2019) over a 50-year period, utilising the Big Five measures of John and Srivastava (1999), demonstrated that personality traits remain stable and malleable from the age of 16 to 66. Srivastava et al. (2003) discovered that certain personality traits, particularly the traits of *conscientiousness* and *agreeableness*, increased throughout early and middle adulthood at varying rates. Moreover, the *neuroticism* trait decreased among women and did not change among men (Srivastava et al., 2003).

The next section presents a literature review of the Big Five personality traits in the information security domain. As this study focuses on phishing, particular emphasis is placed on personality traits and their influence on susceptibility to phishing.

3.3 Background on personality traits relating to information security behaviour

This section presents a literature review of personality traits as they relate to the information security discipline. Following a literature investigation, personality traits were categorised into employee security behaviour, internet dependency, persuasion principles, gender differences, social network sites (SNSs) and phishing.

3.3.1 Influence of personality traits on employee security behaviour

In the workplace, it has been found that personality traits have an influence on the behaviour of employees. Shropshire et al. (2015) found the relationship between intentions to adopt software and actual use are moderated by the traits of *conscientiousness* and *agreeableness*. Similarly, users high in *agreeableness* are more likely to adopt security software and have greater concern for the security of their personal information (McCormac et al., 2017).

Studies conducted more than a decade ago show linkages between the Big Five personality traits and cybersecurity compliance behaviour (Shropshire et al., 2006). Participants who possess *openness*, *conscientiousness* and *agreeableness* are more likely to comply with

cybersecurity policies while, conversely, *extraverted* and *neurotic* participants are more likely to violate them (Warkentin et al., 2012). Similarly, McBride et al. (2012) noted that individuals who are more *extraverted* are more likely to violate cybersecurity policies in comparison to *neurotic* and *conscientious* individuals.

It was also found that the Big Five are predictors of counterproductive behaviours in the workplace (Salgado, 2002). *Conscientiousness* predicts deviant behaviours and turnover and *extraversion*, *openness*, *agreeableness* and emotional stability have been found to predict turnover, while none of the traits predicted absenteeism or accidents (Salgado, 2002). A study by Cellar et al. (2001) revealed that *conscientiousness* and *agreeableness* are the two most influential personality traits in predicting fewer workplace accidents. In the context of supply chain management, Erjavec et al. (2019) investigated the way personality traits influence decision-making. They found that decision-makers with lower levels of *extraversion* and *agreeableness* and higher levels of *conscientiousness* and *openness* make better decisions. On the other hand, *neuroticism* and *agreeableness* negatively affect confidence in making decisions.

3.3.2 Influence of personality traits on internet dependency

Internet dependency and internet compulsivity are terms used interchangeably in the literature to represent compulsive overuse of the internet, also known as internet addiction (Mitchell, 2000). Internet addiction is a significant predictor of risky cybersecurity behaviours (Hadlington, 2017). Various types of addictions exist within the framework of “internet addiction”, which can influence users’ behaviour as well as negatively affect their psychological well-being (Young, 1999). The types of addiction include SNS addiction (Karaiskos et al., 2010; Kuss & Griffiths, 2011), internet gaming addiction (Kuss & Griffiths, 2012), internet gambling addiction (Griffiths, 2003), and internet sex addiction (Griffiths, 2012). Cash et al. (2012) and Karaiskos et al. (2010) refer to these various categories of internet addiction as an internet addiction disorder (IAD). Cash et al. (2012) state that there is still a debate on whether IADs should be classified as a behavioural addiction, an impulse-control disorder or even an obsessive compulsive disorder (OCD). SNSs encourage a need for belongingness and it was found that gratification stemming from social enhancement needs increases the likelihood of social network OCD (James et al., 2017). Thadani and Cheung

(2011) discovered that dependency to SNSs is a significant antecedent of habit. Although internet dependency may be reinforced by the gratification resulting from the behaviour, it is not to be confused with “habit” which was discussed in the previous chapter as one of the dispositional factors influencing security behaviour.

Prior literature has found that certain Big Five personality traits have an influence on some of the aforementioned internet addictions. SNS users who seek a sense of belonging, related to *openness*, on SNSs appear to be at risk for developing an addiction to SNSs (Kuss & Griffiths, 2011). It was also found that younger people, possessing the *neuroticism* trait and FOMO, predicted social media use, while only FOMO predicted social media addiction (Blackwell et al., 2017). James et al. (2017) found that a sense of belonging on SNSs increases the likelihood of FOMO which is a negative emotion that drives OCD. In an Australian university student sample of 201 participants, Wilson et al. (2010) found that high *extraversion* and low *conscientiousness* significantly predicted both addictive tendencies and the time spent using an SNS. The researchers suggested that the relationship between *extraversion* and addictive tendencies could be explained by the fact that using SNSs satisfies the extraverts’ need to socialise with others on the platform.

3.3.3 Influence of persuasion strategies on personality traits

As discussed in the previous chapter, of the most popular persuasion strategies, Cialdini’s six persuasion strategies have been most adopted, particularly in phishing research. Owing to the influence of personality traits, not all persuasion strategies have the same effect on individuals (Alkış & Temizel, 2015; Gkika et al., 2016; Lawson et al., 2020; Oyibo & Vassileva, 2019; Uebelacker & Quiel, 2014). This has prompted an interest by various scholars in investigating the extent to which personality traits are influenced by persuasion strategies. Literature has revealed direct interactions between each personality trait and persuasion strategies (Lawson et al., 2020).

In a study using 381 university students, Alkış and Temizel (2015) found that the traits of *agreeableness*, *conscientiousness* and *openness* are the strongest predictors of Cialdini’s six principles of persuasion. *Agreeableness* is the most susceptible personality trait, while *openness* is the least susceptible trait to persuasion strategies. The *extraversion* and *neuroticism* traits were found to be more susceptible to strategies that incorporated the *scarcity*

principle, while all personality traits, except *openness*, were found to be susceptible to the *reciprocation* strategy (Alkış & Temizel, 2015).

With a sample size of 216 participants, Oyibo et al. (2017) found that individuals high in *conscientiousness* were more susceptible to the principles of *commitment* and *reciprocity*, but less susceptible to the *liking* principle. Their findings also revealed that participants high in *agreeableness* were more susceptible to the principles of *authority*, *commitment* and *liking*. Participants who scored lower in *openness* were more susceptible to the persuasion strategies of *authority*, *social proof* and *liking*, which individuals high in the *neuroticism* trait were more susceptible to the strategy of *social proof*. Oyibo et al. (2017) found that none of the personality traits predicted *scarcity*. More recently, Oyibo and Vassileva (2019) found that the traits of *neuroticism*, *openness* and *conscientiousness* are susceptible to social proof.

Gkika et al. (2016) found participants identified as high in *extraversion* appeared to be influenced by all six persuasion principles. However, contrary to their expectations, they found that participants low in the *conscientiousness* trait did not follow their original decision behaviour and were thus influenced by the *consistency* principle.

By means of a semi-structured expert interview, Cusack and Adedokun (2018) ascertained the expert participants' reasons for why they had fallen victim to an SE attack. Following this, the authors conducted a personality test to determine each of the experts' personality traits. One of the experts' noted that the attack had been successful because of a trusted relationship, love, or humour, thus related to the *extraversion* and *agreeableness* traits, which was consistent with the experts' personality test outcome. Another expert believed that it was the desperation of finding a job and trust in other people's feedback that made them a victim, thus indicating traits of *conscientiousness* and *agreeableness*. However, the participant's personality test result did not correlate with their reasons. The third participant believed it was the fear of being locked out of a personal account and not having access to money, as well as a lack of training and education, that allowed the SE attack to be successful. This reason indicates the participant was compliant as well as fearful of losing out, which are characteristics related to the traits of *agreeableness* and *neuroticism*. This participant's reasons did not entirely conform to their personality results. The last participant indicated that money, curiosity and trusting positive feedback from others had led to the successful SE attack. These traits are indications of *openness*, *extraversion* and *agreeableness* which correlated with the participant's personality

test result. As mentioned in Section 2.4.6, curiosity has been equated with facets of the *openness* trait (McElroy et al., 2007; Moody et al., 2017). Cusack and Adedokun (2018) concluded that users high in *agreeableness* and *extraversion* traits are likely to be more susceptible to SE attacks than others. Furthermore, Cusack and Adedokun (2018) are of the view that traits are also influenced by moderating variables such as the emotional state, the environment and motivations.

Based on their personality traits, some users may be more susceptible to specific persuasion techniques employed in phishing emails (Butavicius et al., 2015; Gkika et al., 2016; Lawson et al., 2017; Lawson et al., 2018; Oyibo et al., 2017; Uebelacker & Quiel, 2014). Lawson et al. (2017) and Lawson et al. (2018) investigated the influence of personality traits on four of Cialdini's (2007) persuasion principles and their effect on their participants' ability to distinguish between phishing emails and legitimate emails. Lawson et al. (2017) found that participants were more likely to correctly identify legitimate emails when they utilised the *liking* principle, and conversely with phishing emails. For both legitimate and phishing emails, the *authority* and *scarcity* principles used in combination were most likely to arouse suspicion and were correctly identified as phishing. High *extraversion* was confirmed as being highly predictive of susceptibility to phishing emails, while *conscientiousness* was found to be associated with increased detection of phishing attacks when utilising three of Cialdini's principles in combination.

Uebelacker and Quiel (2014) conducted a literature review and developed the Social Engineering Personality Framework (SEPF) which depicts the theoretical relations between each of Cialdini's persuasion strategies and the Big Five personality traits. For each personality trait, examples were provided of an attack method as well as a coping strategy. Although the framework is yet to be empirically evaluated, the authors propose that *conscientiousness* is influenced by the persuasion principles of *authority*, *commitment* and *reciprocity*. For *extraversion*, they propose this trait to be susceptible to the *liking*, *social proof* and *scarcity* principles, and *agreeableness* to be susceptible to *authority*, *reciprocity*, *liking* and *social proof*, while *openness* is susceptible only to *scarcity*. Uebelacker and Quiel (2014) are of the view that the *neuroticism* trait is least at risk, as individuals high in *neuroticism* are unlikely to succumb to any persuasion technique owing to their distrusting nature. In summary, Uebelacker and Quiel (2014) found that *agreeableness* increases and *neuroticism* decreases susceptibility and argue that *conscientiousness*, *extraversion* and *openness* show both

increased and decreased susceptibility to SE depending on context and sub-traits. Interestingly, in terms of these aforementioned traits, Costa Jr et al. (2001) found that there was no consistent difference in terms of gender and culture.

Wall et al. (2019) investigated the relationship between the Big Five, the Dark Triad and Type D personality profiles and susceptibility to persuasion. By means of latent profile analysis, the authors identified three distinct profiles which were labelled socially apt, fearful and malevolent. The malevolent profile was more susceptible to *scarcity* and least susceptible to *reciprocity* and *authority*. The socially apt profile appeared to be more inclined to be persuaded to do something if it was consistent with their beliefs. The fearful profile was more likely to report obeying those in *authority*.

3.3.4 Influence of personality traits and gender on phishing detection

It is not unexpected that the gender of an individual can also predict certain behaviours. For example, people stereotypically anticipate that men will be more assertive or aggressive than women. This was found in a study by Costa Jr et al. (2001) in which female participants who were reported to be high in *openness* to aesthetics and feelings were higher in other *extraversion* facets such as warmth. Men scored higher on some facets of *openness*, such as openness to ideas, and higher on some facets of *extraversion* such as excitement seeking and assertiveness. Moreover, Costa Jr et al.'s (2001) findings also hold true for older adults as a study by Chapman et al. (2007) revealed that gender differentiation remains stable over the lifespan.

Researchers have explored the influence of gender and personality traits on phishing susceptibility (Halevi et al., 2013; Parrish Jr et al., 2009; Pattinson et al., 2012; Sumner et al., 2011). In a study by Hong et al. (2013), participants were required to differentiate legitimate and malicious emails, deleting them if they were considered to be suspicious. In their experiment they used seven phishing emails, one spam and malware, and five legitimate emails. Although 89% of the participants reported that they were “confident” in their ability to identify malicious emails, 92% of them misclassified phishing emails. Furthermore, 52% of the participants misclassified more than half of the phishing emails, while 54% deleted at least one genuine email. Women were less likely to identify phishing emails than men. In terms of

personality traits, *extraversion* and *openness* were correlated with deleting legitimate emails. Participants who considered themselves as “less trusting, introverts, or less open to new experiences” were more likely to delete legitimate emails. Only 2% of participants correctly classified all emails, indicating that approximately 98% would have experienced adverse consequences. The results of Hong et al. (2013) further highlight the influence gender, trust, and personality have on phishing vulnerability.

Halevi et al. (2013) examined the relationship between the Big Five personality traits and email phishing responses, as well as how these traits affect users' privacy behaviour on Facebook. Their study revealed that 17% of the participants had been “phished” and found women to have a very high correlation to *neuroticism*, while for the men no correlation was found to any personality trait, although *neuroticism* and *openness* had an inverse correlation to *extraversion*.

As discussed in Section 2.7, computer self-efficacy plays an important role in one's acceptance and use of information technology. Furthermore, the more familiar people are with computers, the better they manage phishing (Pattinson et al., 2012). Taking gender into account, the traits of *neuroticism*, *extraversion* and *agreeableness* are significantly related to computer self-efficacy for women but not for men (Saleem et al., 2011).

3.3.5 Influence of personality traits on social media user behaviour

Privacy concerns and practices on SNSs have long remained an area of interest for scholars (Feng & Xie, 2014; Gross & Acquisti, 2005; Nyoni & Velempini, 2018; Wang et al., 2011). In terms of gender differences, Fogel and Nehmad (2009) for example found that men were less concerned about privacy issues and identity information disclosure on SNSs than their female counterparts. To a certain extent, it is possible to determine the personality traits of social network users by the information they publicly post. Using the posts (i.e. tweets) of Twitter users, Carducci et al. (2018) introduced a supervised learning approach to compute personality traits, thereby predicting the five personality types of Twitter users.

Recently, scholars have investigated the influence personality traits have on privacy behaviour (Li et al., 2019). In online communication, such as chat rooms, Amichai-Hamburger et al. (2002) found that individuals high in the *neuroticism* trait demonstrated a lack of privacy

as they were more willing to post accurate personal information on their profiles. By contrast, a study by Ross et al. (2009) found that *neuroticism* was not linked to the posting of personally identifying information such as email addresses or phone numbers, nor was it related to the use of communicative features on Facebook. Their findings supports those of Li et al. (2019) and Sumner et al. (2011), who found users who are concerned for their privacy are significantly positively correlated with *neuroticism*. Halevi et al. (2013) found that people who score high on the *openness* trait tend to both post more information on Facebook and have less strict privacy settings. Sumner et al. (2011) found that Facebook users' privacy concerns were significantly negatively associated with *extraversion* and *agreeableness*, with the latter finding being supported by Li et al. (2019).

Prior literature has shown that personality traits have a significant influence on user behaviour on SNSs such as Facebook. The personality traits of *extraversion* and *openness*, were found to be significant predictors of social network users being at risk to cyberbullying, while most personality traits were significant predictors of at least some form of risky SNS practices (Peluchette et al., 2015). More recently, after conducting a systematic literature review, Van der Schyff and Flowerday (2019) posited that individuals' personality traits, in conjunction with their awareness of social media surveillance, may influence their privacy behaviour in relation to their intention to use third-party Facebook apps.

Extraversion was found to be a predictor for the number of friends in the real world and for the number of Facebook contacts (Quercia et al., 2012). However, Ross et al. (2009) found that *extraversion* was not significantly related to number of Facebook friends, time spent online or use of the communicative Facebook features. On the contrary, Amichai-Hamburger and Vinitzky (2010) and Sumner et al. (2011) found that individuals high in the *extraversion* trait will show a higher number of friends. Amichai-Hamburger and Vinitzky's (2010) explanation for Ross et al.'s (2009) contradictory findings is that Ross et al.'s (2009) study relied on self-reports by participants whereas their own is based on more objective criteria.

Apart from the Big Five, some authors have shown interest in the Dark Triad of personality traits. The Dark Triad traits consists of psychopathy, machiavellianism and narcissism, and was found to have an influence on certain behaviours on Facebook (Lopes & Yu, 2017; Withers et al., 2017). Kuss and Griffiths (2011) found that young vulnerable people with narcissistic tendencies are particularly prone to engaging on SNSs in an addictive way (La

Barbera et al., 2009). Individuals who scored higher on the dark personality trait of narcissism and lower in self-esteem were related to greater online activity as well as some self-promotional content on Facebook (Mehdizadeh, 2010).

3.3.6 Influence of personality traits on user susceptibility to social engineering and phishing

Parrish Jr et al. (2009) proposed a conceptual framework, referred to as the Phishing Susceptibility Framework (PSF), which utilises the Big Five personality traits. This framework consists of four main groups of factors: personal, experiential, personality profile, and phishing susceptibility. The personal factors include gender, culture and age. According to Parrish Jr et al. (2009), experiential factors such as general experiences, technological experiences and professional experiences shape an individual's personality because of past events or experiences. Their framework posits that both personal and experiential factors will have a direct effect on the Big Five personality model, thus leading to susceptibility to phishing.

Alseadoon et al. (2015) found that *openness*, *extraversion* and *agreeableness* traits predict the likelihood of a user responding to phishing emails. The authors found that those high in *extraversion* were twice as likely as others to respond to phishing emails, while those higher in *agreeableness* scores were four times more likely than others. However, Pattinson et al. (2012) presented opposing findings, as they found that participants who possessed the *openness* and *extraversion* traits were better at managing phishing emails. Pattinson et al. (2012) also discovered that for managing legitimate emails, none of the personality traits, except for *agreeableness*, had a significant negative effect, thus indicating that the more “agreeable” an individual is, the worse they are at managing legitimate emails.

Mayhorn et al. (2015) found that both the dispositional and behavioural factors of users influence phishing detection. Although the authors did not explicitly use the Big Five traits, the participants' detection task was affected by personality factors that share some of the characteristics of the Big Five traits such as reservation, the ability to keep emotions under control, distrust, a belief that others are essentially evil, losing money without being reimbursed, and trusting that one may receive a legitimate request to confirm account information via email. According to Mayhorn et al. (2015), their results suggest that personality characteristics that support reserved behaviour, low impulsivity and distrust decreased phishing susceptibility in an email-based decision-making task.

3.4 Summary

The aim of this chapter was to show that personality traits are an important predictor of user behaviour which can put users at risk of phishing attacks. The chapter introduced personality traits and described in detail the characteristics of the Big Five traits, as well as giving a brief history of associated measurement scales. A literature review of personality traits conducted in the information security domain was also presented. Subsequently, the literature review categorised the influence of personality traits into the following main areas of information security behaviour: employee security behaviour; internet dependency; responses to persuasion principles; influence of gender differences on phishing detection; user behaviour on SNSs; and SE and phishing. The literature review on personality traits revealed overlaps in some of the aforementioned areas. For example, it is evident that most personality trait research conducted in the area of phishing susceptibility has considered the influence of gender. Furthermore, gender differences have also been explored in the context of SNSs, particularly on privacy issues related to the sharing of personal information. To a large extent, the literature has predominantly focused on examining the influence of personality traits in response to various persuasion strategies, as well as the influence of personality traits on privacy behaviour on SNSs. Both areas revealed contradictory findings by several scholars investigating this phenomenon. This has sparked interest by scholars in further investigating and replicating the studies. To a certain extent, it was found that these studies correlated with findings that were previously opposed or proposed other variables that influence the outcome. The review also identified a study by Cusack and Adedokun (2018) in which most of the participants who reported their reasons for falling victim to a phishing attack did not correlate with their actual personality trait test result. The above-mentioned findings pose concerns for the validity of certain personality tests and further highlight the need to conduct more studies on personality traits by considering dispositional factors in various environments, as well as exploring culture as an influence on behaviour. From what could be determined, there are currently no prior studies that examine the influence of personality traits on habits and moreover on information processing. To conclude, this chapter addressed the second research sub-question of the study in an attempt to answer to what extent personality traits influence phishing susceptibility. The next chapter introduces information processing which contributes as the theoretical foundation of the current study and also explains the nature of the problem under investigation.

4

INFORMATION PROCESSING AND ITS INFLUENCE ON USERS' ABILITY TO DETECT PHISHING

4.1 Introduction

Chapter 2 established that SE techniques, in particular the use of persuasion strategies, are an effective means of deceiving people into performing certain actions that will benefit the attacker in some way. Furthermore, dispositional factors such as habits, computer self-efficacy, perceived risk and social norms are factors that each bring on their own inherent set of vulnerabilities which subsequently influence people into behaving in a certain manner, thus putting them at risk of phishing attacks. The previous chapter introduced personality traits and characterised them into five broad personality dimensions, also known as the Big Five. Each of the traits are aligned with descriptions that characterise the behaviour typically associated with that trait. All of the aforementioned human factors can influence the way users will respond to the persuasive strategies typically incorporated in phishing messages. Conway et al. (2017) state that education is unlikely to be sufficient to curb risky behaviour if the user does not engage their implicit knowledge of the subject matter in order to evaluate threats. Similarly, Vishwanath et al. (2011) states that domain-specific knowledge acquired through education, awareness or experience has a limited effect on preventing phishing susceptibility because applying knowledge requires elaboration.

Appropriately, this chapter introduces information processing, utilising the Heuristic-Systematic Model (HSM) of information processing as an underlying factor to phishing susceptibility. As such this chapter aims to address sub-question six of the study. It also serves as the theoretical foundation for this study. Using the HSM, this chapter investigates why some

users give insufficient attention to detail when presented with phishing messages on SNSs. The chapter begins by presenting popular theoretical frameworks and models applied in the information security domain to explain or predict human behaviour. It then discusses the HSM as a contributing factor to phishing deception.

4.2 Psychological frameworks used in phishing research

The theoretical framework is the foundation on which knowledge is constructed and can be equated to a “blueprint” for the entire thesis inquiry (Grant & Osanloo, 2015). Theoretical frameworks introduce and describe appropriate theory that explains why the research problem under study exists (Grant & Osanloo, 2015). Scholars in the information security domain have adopted a wide variety of psychological theories and models to help understand, explain or predict human behaviour under different situations and in different contexts (Cameron, 2009). In addition, some scholars have incorporated more variables into existing theories in order to examine their effects and to advance theory. For example, Jansen and Van Schaik (2019) added two components to protection motivation theory (PMT) to provide a more comprehensive view on the effects of fear appeals and Venkatesh and Davis (2000) extended the Technology Acceptance Model (TAM) by explaining perceived usefulness and usage intentions with consideration to social influence and cognitive instrumental processes. Moreover, some scholars have attempted to integrate two or more theories to examine their effects. For example, Ifinedo (2012) investigated information systems security policy (ISSP) compliance by drawing upon both the TPB and the PMT. Meanwhile, Lee (2009) proposed a theoretical model, using TAM and TPB, to explain customers’ intentions to use online banking.

While the TAM is the most frequently used theory in the IS discipline (Lim et al., 2009), information security practitioners with a particular focus on human factors have largely been inspired by psychological theories and models emanating mainly from the health psychology discipline. In this area, theories that are widely popular include the theory of reasoned action (TRA), the theory of planned behaviour (TPB), and protection motivation theory (PMT). Burns et al. (2013) state that these aforementioned theories are referred to as “continuum models” and that their major drawback is that such models assume that when people intend to behave in a certain way, they do actually perform that behaviour. This problem is referred to as the “intention–behaviour” gap (Sheeran, 2002).

The PMT, which was developed by Rogers (1975), emerged from the health psychology discipline and has been utilised as a theoretical framework in many behavioural studies in the information security domain (Briggs et al., 2017; Crossler & Bélanger, 2014; Jansen & van Schaik, 2019; Schuetz et al., 2016; Schuetz et al., 2020; Van Bavel et al., 2019; Vance et al., 2012; Williams et al., 2018). PMT suggests that when users encounter a communication that encourages fear, they undergo a threat appraisal process in which they are motivated to carry out a protective action (Williams et al., 2018).

Another theory adopted by scholars in phishing research is signal detection theory (SDT) (Lawson et al., 2020). SDT is used when psychologists intend measuring how users make decisions under conditions of uncertainty. SDT assumes that the decision-maker is not a passive receiver of information, but an active decision-maker who makes difficult perceptual judgements under conditions of uncertainty. These judgements range from a hit, miss, false alarm to a correct rejection. These judgements are influenced by two kinds of “noise factors” that contribute to the uncertainty: internal noise and external noise. Canfield et al. (2016) applied SDT by using an online scenario task to compare users’ detection of phishing and their subsequent behaviour following detection. Sheng et al. (2007) designed an educational game to help improve users’ ability to detect phishing websites. Using SDT, Sheng et al. (2007) revealed that while existing online training materials increased awareness about phishing, their game also made users more knowledgeable about techniques they could use to identify phishing websites. Inspired by Sheng et al. (2007), Mayhorn and Nyeste's (2012) study revealed an association between anti-phishing training techniques and individual differences with regard to phishing susceptibility. The benefit of users’ training was assessed by determining how susceptible they were to phishing at two different stages: during the first week of training, and one week after the training. Kumaraguru et al. (2010) used SDT to quantify users’ ability to discern between phishing websites and legitimate websites. Their study revealed that users were able to detect phishing more accurately after playing the game and they were able retain their knowledge of the game for at least one week. They used two measures: sensitivity and criterion. Sensitivity to be the ability to distinguish phishing websites from legitimate websites and criterion is defined as the tendency of users towards caution when making a decision. Butavicius et al. (2015) used SDT to assess users’ performance on both genuine and fraudulent emails using two measures: discrimination and bias. The discrimination measured determined how well users could distinguish between genuine and

fraudulent emails and the bias measured users' tendency to identify an email as either genuine or fraudulent.

As phishers continuously aim to improve the authenticity of spoofed websites, the visual discrepancies between spoofed websites and their original counterparts are often difficult for users to detect, despite training. In this regard, Moreno-Fernández et al. (2017) focused on typography, as phishers appear to be having difficulty in mimicking exactly the same “font” as in the original website. Moreno-Fernández et al. (2017) go on to argue that because participants are primed to think about security, this increases their alertness to the possibility of being tricked, thus encouraging more cautious responses. SDT was used to evaluate how training may affect discriminative abilities (i.e., sensitivity to perceptual differences that define the original and fake websites) independently of response strategy.

The importance of utilising theories to understand a phenomenon should not be underestimated. Pertaining to phishing, Luo et al. (2013) state that there has been “limited theory-grounded research”. The current study posits that susceptibility to phishing extends beyond attitude to a lack of prior knowledge and awareness. A theory that has recently emerged to help explain the factors causing users to overlook suspicious phishing messages is the Heuristic-Systematic Model (HSM) of information processing. In the current study, this model was adopted as the appropriate theoretical basis for this study as it not only identifies the influential factors that make users susceptible to phishing, but also explains the mechanisms in terms of which these factors function (Xu & Zhang, 2012). As such, it also creates opportunities to design counter-measures that address these factors. Moreover, as aligned with the justifications made by Davis and Tuttle (2013), the HSM is designed to be applicable to a wider range of validity-seeking contexts and information processing activities than are other dual-process theories. Various studies have adopted it to understand information processing in different contexts and situations: online communities, videoconferencing and collaboration. HSM is also applicable to communication contexts in which individuals are informed about a situation in which they must reach a decision or exercise a judgement. The HSM is discussed in more detail in the next section.

4.3 The Heuristic-Systematic Model

Today, the distinction between professional and social communication is often blurred, as both streams of information can originate from various channels such as email, social media, communication apps and instant messenger apps. Situational events, for example the Covid-19 period, may present users with a large volume of messages stemming from the various platforms, thereby increasing the cognitive effort needed to identify and respond to messages. Social media users are, in particular, prone to information overload due to the constant attention that needs to be given to the large volume of information on these sites (Lee et al., 2016). For example, the structure of the Facebook news feed/timeline could potentially cause information overload (Thomas-Jones, 2010). In this regard, depending on what the user has subscribed to (i.e. follow), a news feed/timeline can contain a mixture of content stemming from friends, pages, groups and unsolicited advertisements. Moreover, a user could simultaneously be dividing their attention between SNSs and computer-related or online tasks such as emailing, viewing websites or working on open documents, thereby increasing the demands on their attention. As such, they may not be motivated to consider the security aspects associated with that threat (Moreno-Fernández et al., 2017). Consequently, to save time and effort, users may resort to other forms of “cognitive shortcuts” when attempting to make decisions about a message. This poses another risk, as cognitive load is an important indicator in recognising deception (Jian et al., 2019) and presents an opportunity for phishers to take advantage of this lapse in attention to detail.

Buller and Burgoon (1996) state that identifying the verbal and non-verbal leakage cues in social interactions is key to detecting deception. However, in an online environment, email-based phishing and social network phishing involve no direct physical interaction. As a result, in order to judge the authenticity of messages, users have to resort to other cues such as the content, language and design (Pfeiffer et al., 2014). However, while researchers note that some of the aforementioned characteristics of phishing messages should be obvious for users to identify, prior literature has nevertheless shown that not all users will pay attention to these cues or comply with the security indicators, even those who have computer experience. This is because users do not approach phishing emails with similar cognitive processes or capabilities (Bayl-Smith et al., 2020). To elaborate, a study by Lin et al. (2011) found that certain people paid attention to different elements of the browsing interface to judge a

website's legitimacy. Some based their decisions on (a) institutional brand; (b) content as presented in the main pane of the browser; (c) input information requested, and (d) information in the address bar and other security indicators. However, others when evaluating information online, will use heuristics and cues (Metzger & Flanagin, 2013). Consequently, Sterrett et al. (2019) suggest that people are likely to use two cues, namely: (1) who shared the information and (2) the original reporting source of the story.

Given these concerns, researchers have turned to the cognitive processing paradigm, where the lack of systematic processing is viewed as the sole reason why individuals do not notice the aforementioned deceptive clues (Grazioli, 2004; Vishwanath et al., 2011; Vishwanath et al., 2018). A dual-process theory associated with this paradigm is the Heuristic-Systematic Model (HSM) of information processing (Chen et al., 1999). The HSM and the Elaboration Likelihood Model (ELM) are very similar models and are known as dual process models. According to Crano and Prislin (2006), dual-process models are the most influential persuasion paradigms. The key difference between the two models is that HSM explicitly recognises dual processing (i.e. parallel or jointly), while ELM suggests information processing occurs on a continuum metaphorically referred to as an "elaboration continuum" ranging from high to low. The HSM originated from persuasion research in social psychology (Eagly & Chaiken, 1993) and has motivated scholars to attempt to predict human behaviour when presented with phishing (Bayl-Smith et al., 2020; Frauenstein & Flowerday, 2020; Goel et al., 2017; Harrison et al., 2015; Harrison et al., 2016; Hassandoust et al., 2020; Shan et al., 2016; Valecha et al., 2015; Vishwanath, 2017; Xu & Zhang, 2012; Zhang et al., 2012). In persuasive contexts where users have to perform an evaluation to arrive at a judgement, the HSM proposes two distinct modes of information processing, either heuristic or systematic (Bohner et al., 1995; Davis & Tuttle, 2013; Trumbo, 2006). These two modes of processing are discussed next.

4.3.1 Systematic processing

At the one end of the information-processing continuum is *systematic processing*, which requires considerable effort and motivation as it is analytically orientated. In this mode, users will carefully scrutinise, compare and relate arguments (Trumbo, 2006). Users may also seek to perform further investigation to validate the authenticity of a message (Luo et al., 2013). As mentioned by Griffin et al. (2002), systematic processing not only depends on one's capacity

to think critically but also other factors such as one's existing knowledge, self-efficacy in obtaining relevant information and the perceived usefulness and credibility of available information. Ideally, systematic processing would be the preferred method of choice when users are presented with phishing on SNSs; however, this type of processing requires more effort, time and cognitive resources to make a judgement. Moreover, users may be engaged in other activities at that moment, such as using other software applications, which might distract them from giving full attention to the persuasive message.

4.3.2 Heuristic processing

In contrast, at the other end of the continuum, *heuristic processing* is an efficient, shallower form of processing where individuals utilise limited cognitive resources, relying on superficial cues and simple decision rules, often termed “rules of thumb”, on which to base their judgement of the validity of a message (Trumbo, 2006). Eagly and Chaiken (1993) explain heuristic processing as “a limited mode of information processing that requires less cognitive effort and fewer cognitive resources” (p. 327). Heuristic processing is also performed when people lack motivation or cognitive resources to further evaluate a message (Bohner et al., 1995). As mentioned, phishers resort to persuasion techniques in order to encourage the user to perform a validity assessment quickly and without deliberation (Luo et al., 2013). In this regard, perceived credibility, likeability, or the attractiveness of the message source encourage heuristic processing (Workman, 2008). As this processing occurs at a shallow or peripheral level, the receiver typically forms judgements based on certain factors or indicators such as trustworthiness, appeal and the length of the message (Cameron, 2009) – all of which are characteristics commonly exploited by phishers. As a result, certain features contained within a message, such as an incorrect company name, spelling etc., are more likely to be overlooked (Williams et al., 2018). Furthermore, social media users could be at risk of information overload as they may receive large numbers of messages from their friends or those they follow (Valecha et al., 2015). As a result, to reduce information overload, social media users would typically resort to using a heuristic approach (Lin et al., 2016; Vishwanath et al., 2011).

4.3.3 Influences and barriers to information processing

The HSM includes a component termed “sufficiency threshold” which can be described as the “desired judgmental confidence” that people wish to reach when making decisions under a

given circumstance (Eagly & Chaiken, 1993). People will continue processing the message as much as possible until the sufficiency threshold is attained (Luo et al., 2013; Xu & Zhang, 2012). When compelled to make a decision quickly, people may lower their sufficiency threshold, thereby resorting to heuristic processing (Bohner et al., 1995). Furthermore, they will adjust the sufficiency threshold and their decision-making effort according to their perceptions on importance and risks (Eagly & Chaiken, 1993). Thus, not all decisions require systematic processing to achieve the sufficiency threshold and the same can be said for heuristic processing (Luo et al., 2013). Phishing messages almost always contain some attributes that, if systematically processed, can be identified correctly (Luo et al., 2013). However, it is possible that if a cue in the message is unnoticed by the user, or if the user is not aware of the implications of the cue, they will not process the content of the message heuristically (Xu & Zhang, 2012). It is also possible that systematic and heuristic processing can, under specific conditions, operate simultaneously, either independently or interactively (Bohner et al., 1995; Gardikiotis & Crano, 2015). Luo et al. (2013) suggest that phishing may be most effective if the message stimulates both the heuristic and systematic processing modes.

According to Harris and Yates (2015), users evaluate phishing based on two main criteria: the *visual quality* of the message and the *quality of the message argument*, the latter of which requires more motivation from the user to make a decision. Visual quality is concerned with aspects related to source address, company logos, grammar, context and the instruction given in the message (Wang et al., 2012). These are aspects that phishers incorporate in the design of phishing emails. Figure 4.1 briefly illustrates these visual aspects in an email that users will identify in order to perform a quick evaluation of a message.



Figure 4.1: Features in an email typically identified by users

As depicted in Figure 4.1, when users are presented with emails they will either perform heuristic or systematic processing. The features in the email as well as the user's own behavioural influences determine the choice of mode. If the user performs a rapid assessment and uses snap judgements regarding the features, as highlighted by the blue arrows, this approach is indicative of heuristic processing. Persuasion principles incorporated into the message can also be used to further encourage heuristic processing. To elaborate, Goel et al. (2017) state that a contextualised message that threatens the loss of something valuable increases the likelihood of a user responding quickly without considering the potential consequences of the action. Wall and Warkentin (2019) found that by including a fear appeal in a message, the user's perceptions on the quality of the argument influenced their attitudes and behaviour to comply with the appeal specified in the message. In this regard, Wall and Warkentin (2019) found that perceived argument quality influenced response efficacy and behavioural intention to comply with information security policies. Hassandoust et al. (2020) state that "contextualised" phishing attacks make individuals more likely to engage in heuristic processing and repress their systematic processing.

While email clients have built-in filters which can separate spam from legitimate email, this is not the case on SNSs. SNSs such as Facebook present an overwhelming amount of textual and visual stimuli spontaneously to users (in the form of a timeline), thus requiring more effort by the user to distinguish between what is important and what is not. Furthermore, SNSs include both asynchronous (i.e., personal messages sent within the SNS) and synchronous modes (i.e., embedded chat functions within the SNS) of communication (Kuss & Griffiths, 2011), which can distract users. It is thus not unexpected that users could use a heuristic approach on such platforms (Vishwanath, 2017). The hyperlinked structure of the WWW also makes it challenging for users to evaluate various sources as they move between various websites (Metzger & Flanagin, 2013). From the perspective of HSM, an email-based phishing attack can be effective if:

- Characteristics or cues in the message (i.e. pictures, logos, well-known source) overpower the victim's need to further scrutinise the message, thereby promoting heuristic processing.
- The victim is convinced to respond quickly thus promoting heuristic processing. This can be accomplished by using persuasion principles as discussed in Section 2.4, for example employing the “scarcity” principle by persuading the victim that they have won a limited prize offer or using a consequence to instil “fear”. (Refer to Chapter 2 for stimulus examples).
- The sufficiency threshold is reduced so that targeted victims will not engage in systematic processing (Luo et al., 2013).
- The overall message is sufficiently convincing that despite the victim closely examining the message, they still cannot determine that it is phishing (Luo et al., 2013), hence it can withstand systematic processing.

On SNSs, it is difficult to identify these features, mainly because some of these characteristics are not present in the posts on SNSs.

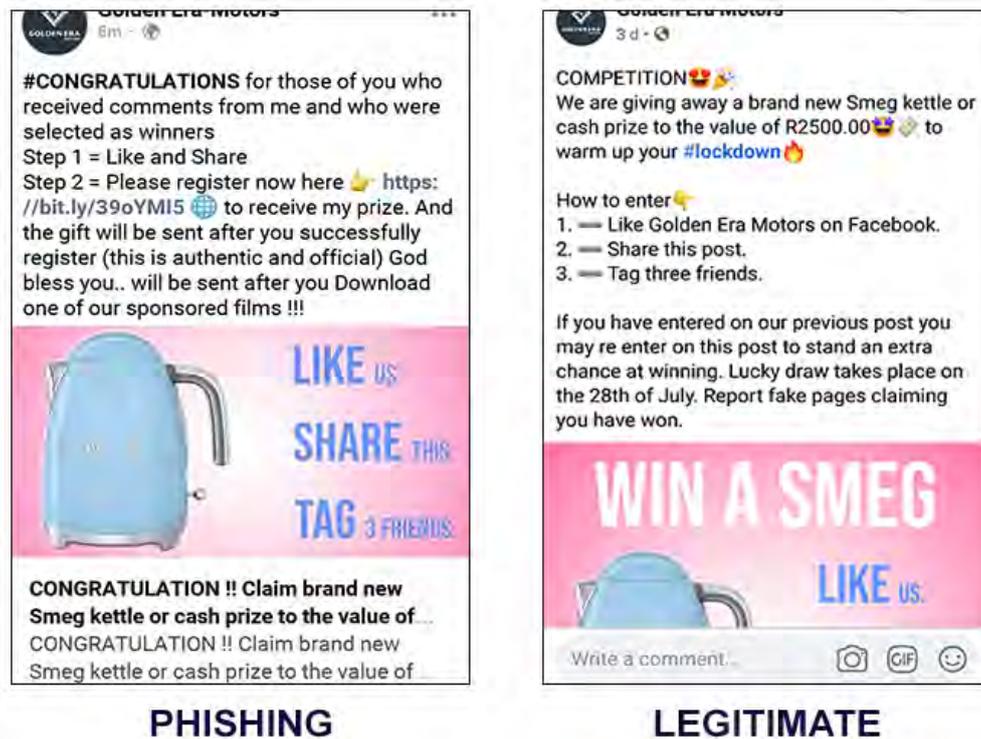


Figure 4.2: Comparison of a Facebook phishing post versus legitimate post

In Figure 4.2, resorting to heuristic processing reveals that the two posts appear very similar to each other with the most attention drawn to the image of the kettle contained in the post. It is only by performing a *systematic* evaluation, by comparing the two posts alongside each other, that one can deduce the similarities and differences. This close inspection of the two posts side by side (as depicted in Figure 4.2) requires effort and is time-consuming, and would not typically occur on SNSs. However, in this instance doing so reveals interesting similarities and highlights the challenge for users to distinguish phishing from legitimate posts on SNSs. In terms of similarities, both used the following:

- Same source logo and business name being impersonated. The difference is that the company name in the phishing post includes a hyphen.
- Same image of the kettle and brand with theme colour.
- Similar instructions to carry out in order to enter to win the prize.
- Attempt to convince the user that it is authentic. The legitimate post requests users to report fake pages while the phishing post states that it is authentic and official.

In terms of distinguishing the phishing post (on the left) from the legitimate post (on the right) in Figure 4.2, systematic processing is required, as well as possibly seeking further validation outside the post (i.e. information-seeking), possibly by visiting the organisation's website. Some of the characteristics used to identify email-based phishing are also common to phishing executed on SNSs:

- *Persuasion principles.* The authority principle was utilised by impersonating a motor dealership company (Golden Era Motors) and using a luxury brand (SMEG) as the main prize to build trust. Scarcity was used as the post indicates some winners have already been directly notified as winners by the source. Conformity or social proof is also being used as statements such as “God bless you” preys on Christian beliefs leveraging on good intentions.
- *Link.* The user is required to click on a link which redirects the user to another website to register. It also appears that they are required to download a film. The link has been shortened (bitl.ly) which masks the full website address from the user.
- *Spelling and grammatical errors.* The phishing post contains a number of errors (e.g. “Congratulation”).
- *Timing.* The phishing post was created a few days after the legitimate post.

HSM asserts that decision-makers prefer low-effort processing (Bohner et al., 1995; Davis & Tuttle, 2013). In order to identify the above-mentioned characteristics the user also needs to be motivated to apply systematic processing (Bohner et al., 1995; Chen et al., 1999). According to Bohner et al. (1995), the HSM emphasises three broad motivational forces that have an impact on the use of heuristic and systematic strategies, namely, accuracy, defence and impression motivation. These motivations can co-occur or adapt under specified conditions (Bohner et al., 1995). Referring to Figure 4.2, the user will thus be required to base their decisions on the credibility of the sender, the urgency and justification of the action requested in the message, as well as a consideration of the type of information being requested and the penalty (if any) for not complying (Wang et al., 2012). Furthermore, the user could attempt to seek further validation of the request by either searching for information online or contacting the authority directly. However, this may also be dependent on the sufficiency threshold alluded to earlier. Nevertheless, Bohner et al. (1995) caution that a motivated effort does not necessitate a search for “truth”. Thus, it is important to determine the influential factors that

might cause users not to follow a systematic approach. Luo et al. (2013) identified these factors as:

- perceived importance of the decision outcome
- perceived risks
- time and other pressures
- skill level
- distractions.

Attackers can exploit some the above-mentioned factors by utilising the persuasion strategies discussed in Section 2.4. For example, the time pressure factor (i.e. urgency) could be enhanced by employing the “scarcity” principle in the message.

Given the above factors, the HSM provides an ideal theoretical framework for investigating phishing susceptibility (Luo et al., 2013; Xu & Zhang, 2012). The next section presents a literature review of studies that utilised the Heuristic–Systematic Model of information processing as a theoretical framework, with a focus on phishing.

4.4 Background on the Heuristic-Systematic Model in the context of phishing

People typically judge the trustworthiness of others by their characteristics such as facial features (Yan et al., 2015), voice articulation (Imhof, 2010), background checks and the like. This evaluation is similar to uncertainty reduction theory that posits when people intend to establish a meaningful relationship with one another, in order to reduce any uncertainty they tend to seek more information about that individual. This can be subsequently used to predict the other’s behaviour (Berger & Calabrese, 1975). According to Harrison et al. (2016), individuals who are generally distrusting of others and are of the belief that they can discern trustworthy people in face-to-face settings are referred to as having the trait of generalised communicative suspicion (GCS). This trait is typically found in experts, such as police officers, and is used to determine deceptive behaviour (Masip et al., 2016). However, in the online world, users do not get an opportunity to engage directly with other people, and thus have to resort to other means when making a judgement. In relation to phishing, users have to consider various characteristics when forming a judgement which requires more effort from the user to

reach a decision. Incorporating persuasion strategies into the message, as discussed in Chapter 2, increases the likelihood of the user responding quickly or without much thought, thus favouring the attacker. In this regard, a user would most likely take the quality and trustworthiness of messages on face value, thus suggesting a heuristic approach. As cited by Workman (2008), Cialdini's (2007) six persuasion principles are associated with peripheral route persuasion – thus heuristic. Therefore, resisting persuasion techniques requires more cognitive resources than accepting them (Sagarin & Cialdini, 2004).

Concerns raised on the visual characteristics users identify as phishing can be dated back as early as Furnell's (2007) study. Using an online survey, Furnell (2007) presented participants with 20 messages. They were then given the task of judging the authenticity of each message. Feedback from the participants gave further insights into the aspects that influenced their choices. According to Furnell (2007), most of the responses could be classified as visual factors (i.e. logos, symbols such as copyright and trademarks, font styles), technical indications (i.e. URL in messages, using “https”), and language and content characteristics within the messages (language errors, presence/absence of recipient details, style of the message). Furnell (2007) noted that despite these useful insights, participants often arrived at “incorrect conclusions”. While the focus of Furnell's study was not on information processing, the study offered insights in terms of identifying what users look out for when evaluating messages. In this instance, users applied a heuristic mode by focusing more on visual characteristics than on the quality of the message argument.

As mentioned, users resort frequently to heuristics to save time and effort when judging information. For example, Dogruel et al. (2015) noted that users resorted to heuristics when downloading smartphone apps from the online app store, thus overlooking potentially important information related to them. When users are presented with email messages, they consider both the content and visual, technical and linguistic aspects when determining their trustworthiness (Pfeiffer et al., 2014). The study by Pfeiffer et al. (2014) also revealed that when users are determining the trustworthiness of a phishing email, their prior experience and education plays a role, as participants who felt suspicious reported “A bank will never ask you for account details via e-mail” and others suggested that email attachments do not typically appear in HTML format.

Griffin et al. (2002) introduced the risk information seeking and processing model (RISP) which synthesised components from HSM and TPB. The RISP is used to predict the extent to which a person pursues risk information and the extent to which he or she will spend time and effort analysing the risk. When participants were presented with risk information related to environmental hazards, Griffin et al. (2002) found that systematic processing was positively related to evaluation strength, attitude strength and the number of strongly held behavioural beliefs.

Harrison et al. (2016) introduced a theoretical model, which extended the HSM, combining three areas of research, that is, interpersonal lie detection, risk communication and phishing. Their model included GCS as having an influence on information insufficiency which in turn influences information processing. Harrison et al. (2016) found that high GCS increases uncertainty, leading to a desire for more information before making a judgement. As a result, this approach promotes systematic processing thereby reducing phishing susceptibility. Drawing on the HSM and the risk communication model, Valecha et al. (2015) investigated the extent to which Twitter users are willing to share phishing information contributed by other users, in relation to the phishing risk characteristics of coping, known and dread.

Harrison et al. (2015) conducted an experiment with the aim of determining how users would react to varying levels of richness and presence cues in phishing emails based on the HSM. They accomplished this by comparing two groups' reactions to a rich phishing email versus a lean-text phishing email, both of which contained the same message and hook. It was found that the rich email, containing the university logo, other logos, hyperlinks and so on, was processed by users heuristically instead of systematically, thus significantly increasing the risk of them falling victim to phishing. Harrison et al. (2015) suggest that the rich information in phishing emails could trigger feelings of social presence thereby reducing the users' considerations of mediation but also indirectly increasing the persuasiveness of the email.

Davis and Tuttle (2013) developed a theoretical model designed to uncover the motivational mechanisms needed to provide effective information processing with a focus on IS exceptions. They categorised their model into four main components that have a direct influence on systematic processing, namely, exception factors, task factors, user factors and information sufficiency. While the study by Davis and Tuttle (2013) did not investigate information processing in the context of phishing, their findings offered insights to help explain why users

might choose to ignore browser warnings when they encountered phishing. For example, Amer and Maris (2007), as cited by Davis and Tuttle (2013), found habitual behaviour influential for information processing as users chose to ignore IS exceptions by clicking “OK” in the hope that the situation would disappear.

Vishwanath et al. (2011) utilised ELM as a theoretical framework and tested the influence that four contextual variables, level of involvement, email load, domain specific knowledge, and computer self-efficacy, have on phishing susceptibility. Vishwanath et al.'s (2011) findings suggest that habitual patterns of media use combined with high levels of email load increased the risk of falling victim to phishing. They also found that users allocated more attention to urgency cues and email subject lines, thus stimulating systematic processing, and were therefore more likely to respond to a phishing email. As a result, this influenced users not to resort to heuristic processing as they overlooked obvious characteristics that could effectively be used to identify phishing emails such as the email source, grammar and spelling.

The study by Vishwanath et al. (2011) highlights the influence of persuasion techniques in effectively diverting users from processing messages heuristically. This creates a dilemma as, ideally, it is preferred that users be encouraged to process information systematically. However, in this case this had an adverse effect and still resulted in the user becoming a victim. This strategy links to the aforementioned strategies needed to make phishing attacks effective by creating a situation to get the user to respond quickly through the use of urgency cues.

Vishwanath et al. (2018) developed a theoretical model, referred to as the suspicion, cognition and automaticity model (SCAM). The model captures the conscious, cognitive factors, the preconscious influence of cyber-risk beliefs, and the automatic, non-conscious actions that lead to phishing susceptibility. Their model extended the HSM by including the constructs of cyber-risk beliefs, deficient self-regulation, media habits and suspicion. Using two experimental studies, Vishwanath et al. (2018) tested participant responses to a phishing link attack and an attachment attack. The findings of their study can be summarised as follows:

- heuristic processing decreases suspicion
- systematic processing leads to an increase in suspicion
- cyber-risk beliefs decrease heuristic processing
- cyber-risk beliefs increase systematic processing in a link attack

- cyber-risk beliefs influence suspicion
- deficient self-regulation influences email habits
- email habits negatively influence suspicion.

The study conducted by Vishwanath et al. (2018) includes the constructs of both *habit* and *information processing* in a single study to investigate susceptibility to phishing. However, they did not examine the direct influence of habit on information processing. As such, owing to the limited number of studies in this area, opportunities exist to further advance this model and explore how other variables could have an impact on habit and its outcome on information processing.

4.5 Summary

This chapter introduced popular psychological frameworks used in information security research. It then introduced information processing and revealed that the HSM provides an appropriate theoretical framework for this study in order to understand the effect information processing has on user decision-making, particularly when users are presented with messages related to phishing. The mode chosen by users to process messages has an influence on the outcome of a phishing attack. As users may not expect phishing attacks to be conducted on SNSs, they may become complacent and thus more prone to processing information heuristically, especially given that most content posted by users on SNSs is graphical (Vishwanath, 2017). The latter poses an increased risk for smartphone users engaging on SNSs (Vishwanath, 2016). The characteristics of emails and Facebook posts that a user will typically identify when performing an evaluation of the message were illustrated. It was revealed that when applying systematic processing to a phishing post on Facebook, more time and effort is required to distinguish it from a legitimate Facebook post. Overall, the literature has shown that studies that utilise the HSM have predominantly focused on two main factors: 1) identifying and examining the influence that certain cues have on users' ability to process information (e.g. email source, grammar, logos) and 2) the extent to which behavioural or motivational factors influence information processing of these cues (e.g. prior knowledge, perceived risk, computer self-efficacy). For example, users that have an increased perception of the risks of phishing are considered more likely to engage in systematic processing of the information described in emails (Williams et al., 2018). This chapter concludes the literature

chapters which contributed to establishing a theoretical foundation for the study. Overall, the theoretical foundation forms the basis for an attempt to investigate the extent to which dispositional factors such as personality traits, habits and information processing influence susceptibility to phishing attacks on SNSs, with the objective of developing a model that can help identify users at risk to phishing. In addition, the literature chapters also investigated social norms, computer self-efficacy and perceived risk and their influence on phishing susceptibility. The next chapter brings together the aforementioned factors and introduces the theoretical model and associated hypotheses.

5

HYPOTHESIS DEVELOPMENT AND MODEL

5.1 Introduction

This chapter addresses the rationale for the main research questions and hypotheses which were formulated to meet the study's research objective. This chapter introduces the theoretical model of the study as derived from the literature investigation. It presents the hypotheses with associated justification for such in favour for the research model of the study. By doing so it answers the research questions of the study and highlights the theoretical contribution the proposed model makes to theory.

5.2 Theory development and hypotheses

The search strategy assisted in identifying the behavioural factors related to phishing susceptibility but, more importantly, also identified gaps which helped define the research question and subsequently led to the development of the model constructs. Understanding how each of these constructs relate to each other helped to achieve the objective of the study. As such, the literature chapters formed the basis on which the model constructs and their relationships were conceived. The terms "hypothesis" and "propositions" are used interchangeably in the literature, although Zikmund et al. (2013) differentiate propositions and hypotheses. According to Zikmund et al. (2013, p. 41), propositions are an assertion of the logical association between certain concepts. On the other hand, a hypothesis is a formal statement of an unproven proposition that can be empirically tested. Thus propositions are considered at the abstract level, while variables and hypotheses are at the empirical level (Zikmund et al. 2013, p. 40). The next subsection presents the hypotheses for this study with a brief justification for each.

5.2.1 Influence of the Big Five personality traits on habit

The personality trait literature on phishing reveals a persistent trend of scholars justifying their contradictory results when explaining their findings. This indicates that personality traits may be influenced by variables such as the culture, specific situations, environments and other behavioural factors (Mischel, 2009; Shappie et al., 2020). Moreover, Bandura (1999) states that personality traits are essentially clusters of habitual behaviours, which may also indicate possible linkages between habits and personality traits. The self-representation of social media users can be linked to the Big Five traits, with an underlying social purpose that might predispose them to satisfying certain needs (Mancinelli et al., 2019), as characterised by Maslow's hierarchy of needs which consist of physiological, safety, love and belonging, esteem, and self-actualisation. Section 2.5.1 established that gratification stemming from repeated behaviour may satisfy a particular need, thereby fostering a habit. Thus, it could be argued that depending on the personality trait, augmented by the differing needs and motives of individuals, the need for satisfaction could be the force driving the habitual behaviour. This is plausible, as a study by Amichai-Hamburger and Ben-Artzi (2003) has shown that the relationships between personality traits and internet use, and moreover loneliness, are important indicators of psychological well-being. A study by Intiful et al. (2019) showed a relationship between Big Five personality traits and dietary habits. They found significant associations between *extraversion*, *agreeableness* and *openness* to neophagia, which is the acceptance of new and unusual foods typically from other cultures. In addition, *extraversion* was found to be significantly associated with food interest. These examples highlight that certain characteristics embedded in each trait can also influence behaviour in other contexts.

This study considers the behaviour of clicking or sharing a social media post as requested by a friend to be a habit and, as such, an undesirable behaviour that could potentially put users at risk of phishing attacks; for example a hoax post shared by a friend stating "I am Elon Musk, if you click and share this link, I will give you \$10,000.00". This can be likened to a phishing email in which the victim is enticed to click on a link in order to confirm or verify something. Thus, in this study it is proposed that a habit, manifesting from characteristics inherent in certain personality traits, makes SNS users vulnerable to phishing.

Lawson et al. (2018) found that incorporating a combination of persuasion techniques in a phishing email made those with the extraversion trait more susceptible. More recently, a study by Lawson et al. (2020) confirmed the extraversion trait to be predictive of increased susceptibility to phishing attacks. In the context of SNSs, extraverted Facebook users have been found likely to participate in risky behaviours (Amichai-Hamburger & Vinitzky, 2010), tend to spend more time using Facebook (Acopio & Bance, 2016; Blackwell et al., 2017; Correa et al., 2010; Mancinelli et al., 2019; Ryan & Xenos, 2011), have more Facebook friends (Acopio & Bance, 2016; Amichai-Hamburger & Vinitzky, 2010; Sumner et al., 2011; Wehrli, 2008) and could potentially comment on or “like” other people’s pictures or selfies (Choi, Sung et al., 2017). This trait has also an influence on their sharing preferences (Gou et al., 2014). In contrast to these findings, Ross et al. (2009) found the fact that a person was an extravert had no effect on the number of Facebook friends, time spent online or the use of Facebook features. From the perspective of social norms and phishing susceptibility, it is expected that extraverts will be more likely act out of habit as they are more engaged on SNSs and thus fulfil the need to receive gratification from the views and opinions of others on SNSs. This frequent engagement of and interaction by these enthusiastic users might reinforce the behaviour of clicking and sharing links originating from their friends. As such, this study proposes:

Hypothesis 1a. *The extraversion trait is positively related to habit on social networking sites.*

Modic and Lea (2012) found highly agreeable people are more susceptible to phishing because they are more inclined to trust in uncertain situations. Similar results by Cho et al. (2016) found agreeableness to have a significant effect on perceived trust and risk in terms of phishing vulnerability. Cusack and Adedokun (2018) also reported that people who display the agreeableness trait are susceptible to SE techniques. In the context of Maslow’s needs, recent personality trait research has seen belongingness emerge as the strongest need (Montag et al., 2020). According to Mancinelli et al. (2019), the agreeableness trait is associated with belongingness, and people with this trait are more prone to accept friend requests on Facebook (Sumner et al., 2011). Van der Schyff, Flowerday, Kruger, et al. (2020) found people with this trait use Facebook extensively. Given prior literature on the vulnerability of the agreeableness trait to phishing susceptibility and that belongingness might encourage agreeable users to

engage more with the online community, it is expected that users with this trait may develop a habit of clicking and sharing links, influenced by their compliant behaviour to “fit in”. Given the latter, this study proposes:

Hypothesis 1b. *The agreeableness trait is positively related to habit on social networking sites.*

As individuals with the conscientiousness trait are more cautious (e.g. noticing an absence of typographical and grammatical errors) and generally more risk averse, they may be less likely to perform risky behaviours (Parrish Jr et al., 2009). Pattinson et al. (2012) found less impulsive users were better at managing phishing emails, as they were likely to spend more time deliberating before making a decision on whether to open them or not. In the context of SNSs, Wehrli (2008) found that highly conscientious people tend to refrain from participation on SNSs. Similarly, Sumner et al. (2011) found people with this trait were less likely to join Facebook groups. Gou et al. (2014) found Twitter users with the conscientiousness trait to be less likely to share their preferences regarding their values and needs. Given the thoughtfulness and cautious nature of the conscientious user, it is expected that such users will be less prone to engage in clicking and sharing links at a habitual level. Accordingly, this study therefore proposes:

Hypothesis 1c. *The conscientiousness trait is negatively related to habit on social networking sites.*

The opposite of emotional stability is neuroticism. Prior literature has it that the neuroticism trait in individuals may decrease a user’s willingness to trust a system (Hwang & Kim, 2007), increase computer anxiety (Korukonda, 2007), promote a strong desire to avoid using the internet (Joiner et al., 2007) and increase resistance to adopting new technologies such as smartphones (Özbek et al., 2014) and instant messenger apps (Swickert et al., 2002). In the context of phishing, Halevi et al. (2013) found neuroticism to be the trait most at risk of responding to phishing emails, with gender-based differences in the responses. Owing to the distrusting nature and lack of technology adoption associated with the neuroticism trait, it is expected that such a user to be less likely to engage frequently on SNSs and thus unlikely to develop a habit of sharing and clicking links. It is thus proposed:

Hypothesis 1d. *The neuroticism trait is negatively related to habit on social networking sites.*

The curiosity, open-mindedness and explorative nature of individuals possessing the openness trait can raise behavioural concerns. For example, Johnston et al. (2016) found that individuals with this trait are likely to violate information security policies. Alseadoon et al. (2015) found that openness is closely related to high phishing susceptibility. In contrast, Uebelacker and Quiel (2014) proposed that due to prior literature indicating openness to be associated with computer proficiency and experience, this trait will be less vulnerable to SE attacks. In an SNS context, individuals possessing the openness trait were found to both post more information on Facebook and have less strict privacy settings (Halevi et al., 2013). Studies have also shown that Facebook users are friends with others who share similar values or traits, particularly with those who possess the openness trait (Lönqvist & Itkonen, 2016). This is a similar result to that of Amichai-Hamburger and Vinitzky (2010), who found users with high openness to be likely to spend more time with, and have more, friends. This is not surprising as the nature of SNSs is to encourage open interactions. Given the description of users with the openness trait being very active in a social network context, they might be more willing to share posts and click on links which could develop into a habit. As such, this study proposes:

Hypothesis 1e. *The openness trait is positively related to habit on social networking sites.*

5.2.2 Influence of habit on information processing

Gardner (2015) depicts habit on an impulsive pathway, such that the perception of cues activates low-level context-behaviour associations. As noted earlier, Turel and Serenko (2012) describe people who perform habitual behaviours as “cognitive misers”, as they tend to be economical with the amount of cognitive effort they allocate to tasks. This description closely resembles the characteristics of heuristic processing. Vishwanath (2015a) points out the conundrum that likening habits to mental scripts implies that “email” habits and heuristic processing may be part of the same process as they both lead to phishing susceptibility. Owing to the *automaticity* aspect associated with habits, which allows users to reduce their cognitive effort when performing tasks, it is anticipated that habits will have a positive relationship with

heuristic processing and an opposing relationship with systematic processing. As an individual who uses a systematic mode of processing applies deep thinking and critically evaluates information, it is expected that habit will not affect this form of cognitive processing. A paucity of research in this area creates an opportunity to examine the effect habits have on heuristic and systematic processing. The study therefore proposes the following:

Hypothesis 2a. *Habit is negatively related to the systematic processing of phishing messages on social networking sites.*

Hypothesis 2b. *Habit is positively related to the heuristic processing of phishing messages on social networking sites.*

5.2.3 Influence of habit on phishing susceptibility

Following an investigation of the effect habits have on both modes of information processing (as hypothesised in H2a and H2b), by examining the direct relationship that habits have to phishing susceptibility it can be determined whether habits are an independent or parallel process to information processing, specifically heuristic processing. According to Robbins and Costa (2017), the concept of habits refers not so much to “how the behaviour is performed” but to “which stimuli prompt the behaviour”, both of which involve aspects of automaticity. As studies by Vishwanath (2015a, 2015b, 2017) have shown, users acting out of habit are at risk of phishing. Accordingly, the study proposes:

Hypothesis 3. *Habit is positively related to phishing susceptibility.*

5.2.4 Influence of information processing on phishing susceptibility

As mentioned in Chapter 1, the effectiveness of SETA programmes in helping users to identify phishing attacks is limited owing to other behavioural factors. As this study considers information processing as a variable that can influence susceptibility to phishing on SNSs, the aesthetics and functions offered on SNSs may have an influence on user behaviour, in particular how they perceive and interpret information. While a lack of knowledge is considered a factor that influences phishing susceptibility, an individual’s capacity for

interpreting information and their motivation play a role in how users will react to risk (Park, 2018). Moreover, a study conducted by Vishwanath et al. (2011) found that participants with domain-specific knowledge did not conclusively yield the expected results in terms of processing information in depth. Within the context of phishing on SNSs, we posit that the snap judgements that users often make, based on superficial cues in posts, will lead them to overlook some of the cues that could raise suspicion. It is proposed that:

Hypothesis 4. *Systematic processing is negatively related to phishing susceptibility.*

Hypothesis 5. *Heuristic processing is positively related to phishing susceptibility.*

5.2.5 Influence of social norms on phishing susceptibility

While the literature has used various terms for subjective norm constructs, in this study social norms is used, these constructs share the common notion that individual behaviour is influenced by the perceptions of what people think others expect from them. This motivates the user to behave in a way that presents them as having what they perceive to be acceptable behaviours. Indeed, the user might possibly gain gratification, in the form of likes, as a form of approval for their behaviour. Workman (2008) found individuals who were higher in normative commitment would succumb to SE techniques. This relates to Cialdini's (2007) *reciprocity* persuasion technique which stems from the sense of obligation that people may feel when they are given something in exchange for something similar in return. Compared to email-based phishing where an individual deals with one particular person (email) at a time, users of SNSs may be influenced by a community of members to behave in a particular way. As such, the study hypothesises:

Hypothesis 6. *Social norms are positively related to phishing susceptibility.*

5.2.6 Influence of computer self-efficacy on phishing susceptibility

The more time an individual spends on the internet, the more likely they are to acquire experience or information about threats and therefore be better equipped to identify them (Moody et al., 2017). Pattinson et al. (2012) found that users who are familiar with computers

are better at detecting phishing. In addition, internet users with high computer self-efficacy tend to be confident in their abilities to handle online threats and ensure their privacy (Yao et al., 2007) and are better at avoiding phishing attacks (Sun et al., 2016; Wright & Marett, 2010). If users do not have the motivation or perceive themselves as lacking the capabilities associated with computer usage, this could prevent them from performing tasks related to detecting phishing (Wang et al., 2017). Examples of such tasks include recognising file extensions of attachments in emails or checking the security indicators of a website to determine whether it is safe or not. As pointed out by Cox (2012), people will avoid an action if they do not believe they have the ability to complete the action and achieve their desired results. It is thus necessary to investigate the influence computer self-efficacy has on users when presented with phishing, especially if their confidence in using SNSs might overshadow their need to practise caution. As such, this study proposes:

Hypothesis 7. *Computer self-efficacy is negatively related to phishing susceptibility.*

5.2.7 Influence of perceived risk on phishing susceptibility

In an online environment, users may perceive risks as distant and accordingly believe that organisations and others are the main targets of attackers rather than them (De Bruijn & Janssen, 2017; West, 2008). This perception of “it won’t happen to me” (Krasnova et al., 2009) can potentially put users more at risk of phishing as they are less prepared to deal with it. Moreover, Facebook engenders a greater sense of trust among its users compared to other popular SNSs (Fogel & Nehmad, 2009; Kim & Hancock, 2015). Social media users do not expect to be faced with phishing attacks (Volkman, 2019b). Thus, if they feel that in general the SNS environment poses a low risk to them, they are likely to continue using SNSs frequently (Herath & D'Arcy, 2015). Sheng et al. (2010) found that risk-averse users were less likely to fall for phishing. Wright and Marett (2010) found that individuals who are suspicious of others and display general distrust toward people are less susceptible to phishing. As individuals have different propensities for risk, it is important to consider perceived risk as yet another contributing factor that influences human behaviour and phishing susceptibility on SNSs (Algarni et al., 2013a). The study therefore hypothesises that:

Hypothesis 8. *Perceived risk is positively related to phishing susceptibility.*

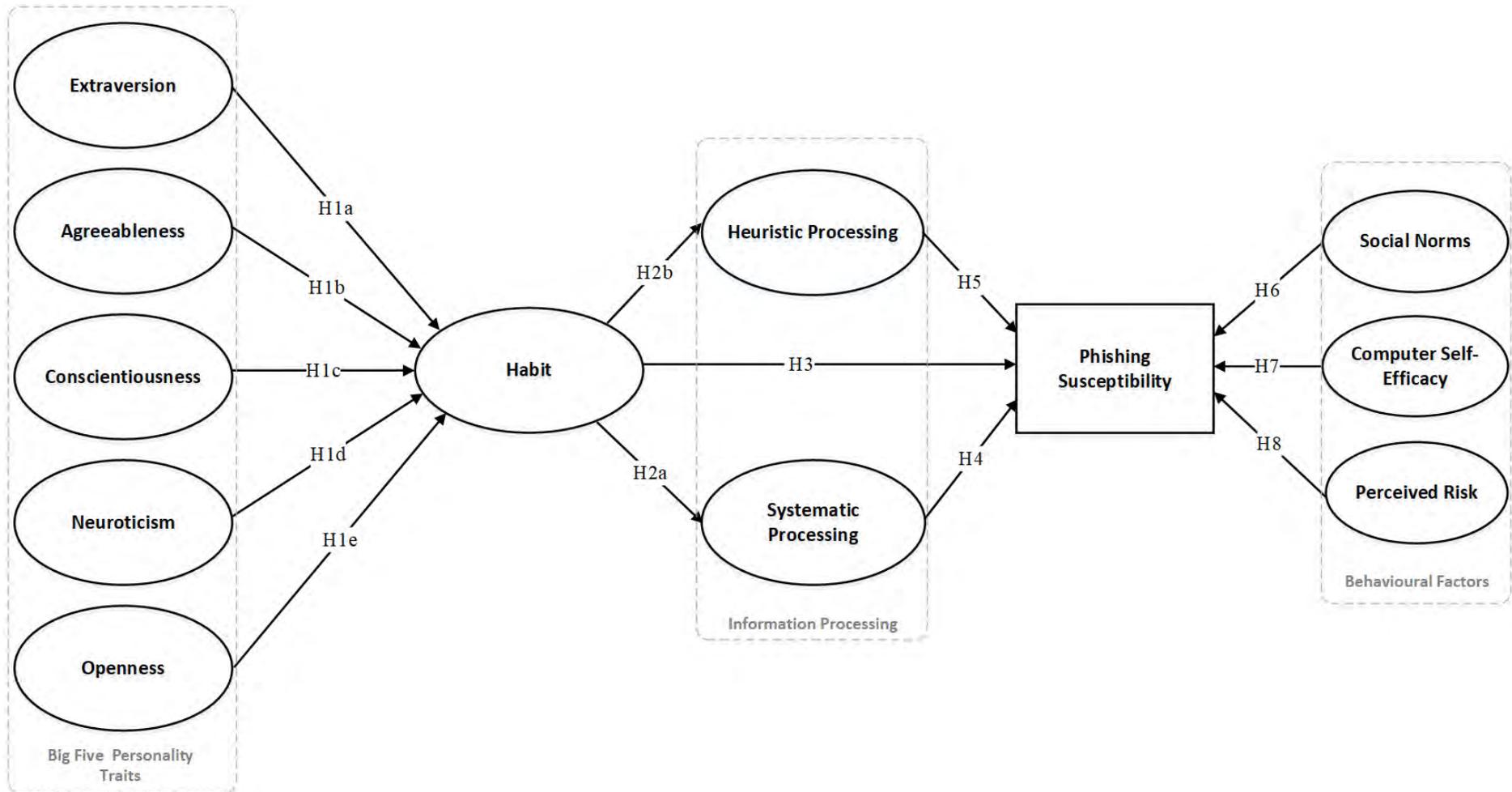


Figure 5.1: Proposed model

To conclude, the proposed model in Figure 5.1, posits that personality traits each have particular characteristics that can render individuals prone to developing habits on SNSs. These habits in turn can cause a user to pay insufficient attention or to overlook key aspects associated with distinguishing phishing messages. The model also proposes that other behavioural factors related to social norms, computer self-efficacy and perceived risk can also influence susceptibility to phishing.

5.3 Demography

This study also measured aspects related to demographics. These aspects included respondents' gender, age, course of study, duration (hours) spent on SNSs, motivation/reasons for visiting SNSs and addiction to SNSs. More information on this can be found in Section 6.9.8.

5.4 Summary

This chapter presented the hypothesised associations between the constructs on the basis of prior literature. The main relationships proposed in this study are as follows: It is hypothesised that certain personality traits, owing to certain motivations stemming from these traits, will have a direct influence on habit. Subsequently, a habit will have either a positive or a negative influence on information processing. Moreover, literature has provided evidence that the mode in which users process the stimuli (often referred to as cognitive processing) determines the likelihood of phishing susceptibility, with heuristic processing implicated as the cause. In this regard, the study utilises the HSM which served as the underlying theoretical foundation for investigating the choice of mode. The study also proposes that social norms, computer self-efficacy and perceived risk will have an influence on phishing susceptibility. Collectively, the associations and discussions contribute towards the development of the proposed theoretical model for the study. Each of the hypotheses are evaluated in Chapter 7. The next chapter describes the procedures and approaches used to ensure that the research process was carried out scientifically from both a qualitative (secondary data collection) and an empirical (primary data collection and analyses) perspective. Particular attention is paid to how the data analysis techniques and procedures were conducted to ensure that the constructs are valid and reliable.

6

RESEARCH DESIGN AND METHODOLOGY

6.1 Introduction

In the previous chapters, the literature established the relevance of the research problem and the theoretical foundation. The previous chapter presented the theoretical model and its associated hypotheses. The objective of this chapter is to describe the procedures and approaches used to ensure that the research process was carried out scientifically throughout the duration of the study. As such, the chapter highlights the philosophical stance underlying the study and discusses the research paradigm, research method and research strategy, as well as the associated data collection and analysis techniques. Furthermore, this chapter substantiates the choice of the specific research methodology and also how these methods made it possible to find solutions to address the research problem stated earlier in the study. Statistical procedures used to ensure that the variables in this study are considered valid and reliable are also discussed.

6.2 Research process

Research is conducted for a variety of reasons, including to understand, describe, predict, or control an educational or psychological phenomenon. Research is a process of *systematic inquiry* that is designed to collect, analyse, interpret and use data (Mertens, 2010, p. 2). Kothari (2008) defines research as “a scientific and *systematic* search for pertinent information on a specific topic”, while Saunders et al. (2016, p. 5) define research as “a process that people undertake in a *systematic* way in order to find out things, thereby increasing their knowledge”. Data is collected and interpreted *systematically* with the clear purpose of discovering

something. In light of the above-mentioned definitions of the research process, it is evident that research can be described as a “process” and that this process must be conducted in a scientific and “systematic” way. The research process consists of a series of actions or steps which are necessary to carry out the research effectively (Kothari, 2008). The research process can be thought of as a lifecycle with multiple stages that a researcher will encounter. Collis and Hussey (2014, p. 9), lists the main stages of the research process as follows: the research topic, literature review, defining the research problem, the research design, data gathering, data analysis, and interpretation and writing the thesis. According to Kothari (2008), the research process consists of a series of activities which, while not mutually exclusive, do not necessarily follow a sequential order, and may overlap. The following activities of the research process is a procedural guideline put forward by Kothari (2008):

1. Formulating the research problem
2. Extensive literature survey
3. Developing the hypotheses
4. Preparing the research design
5. Determining the sample design
6. Collecting the data
7. Execution of the project
8. Analysis of the data
9. Hypotheses testing
10. Generalisations and interpretation
11. Preparation of the report or presentation of the results.

In reference to the above-mentioned activities, up to this point in the thesis, the formulation of the research problem, conducting an extensive literature review and developing the hypotheses has been completed. As this study aims to produce a theoretical model, it is important to ensure that the research process is aligned with such. De Bruin et al. (2005) propose an iterative sequential methodology, outlining the generic phases to be followed when developing a model. Mashapa (2013) adopted a “generic procedure” for developing a model by combining phases based on the work of De Bruin et al. (2005), Becker et al. (2010) and Hevner et al. (2004). This study supports this approach by Mashapa (2013) and adopts a similar process in model formulation, in the form of iterative phases, as follows:

1. Problem definition
2. Requirements identification
3. Determining strategy for model development
4. Building the conceptual model
5. Evaluating the model
6. Presenting the model
7. Model application and maintenance

These phases of *model formulation* are very similar to the aforementioned research process described by Kothari (2008). Such a systematic research process is of great importance in order to achieve the study's objective in a sound manner.

6.3 Research design

Saunders et al. (2016, p. 163) emphasise the importance of the research design, stating that it is the general plan of how one will go about answering one's research questions. The research design includes a clear objective derived from the research questions, and describes the sources from which the data is collected, as well as the manner in which it is collected and analysed. Similarly, McGaghie et al. (2001) state that the research design assists the researcher to focus on the research questions and plan an orderly approach to the collection, analysis and interpretation of data that addresses the question. It must be ensured that these processes are carried out systematically using scientific procedures. In this regard, Abbott and McKinney (2013, p. 21) state that a scientific method gives researchers a systematic way to understand what they observe and ensures as much objectivity as possible in how they think about and observe the world. In terms of applying a scientific method, Abbott and McKinney (2013) use the analogy of a wheel, as in a "wheel of science", where the spokes that make up the wheel consist of the theory, the hypotheses, observation and empirical generalisation. Once the wheel is complete, it is observed and evaluated to determine whether the chosen theory correctly predicted what was found. Figure 6.1 illustrates the popular "research onion" (Saunders et al., 2016, p. 124), which is often used to explain research methodology when considering how it aligns with philosophies, approach to theory development, strategies and data collection techniques. In this regard, the items marked in red in Figure 6.1 indicate the relevant areas selected for this study.

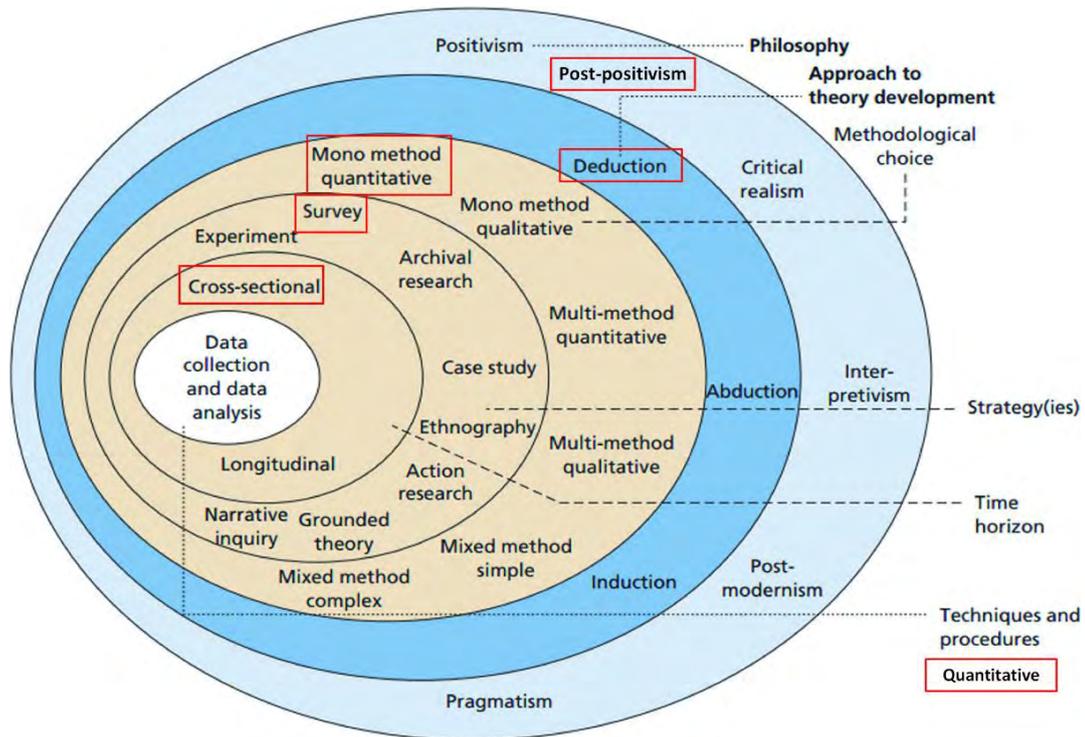


Figure 6.1: The research onion (adapted from Saunders et al., 2016, p. 124)

The research onion framework also provided guidance for this study on how to structure the layout and discussion for this chapter. Collis and Hussey (2014, p. 10) recommend that the first step in research design is to identify one’s research paradigm. For this reason, it is appropriate that the next section discusses the various types of philosophical paradigms and, importantly, the research paradigm chosen for this study.

6.4 Philosophical paradigms

The *research philosophy* refers to a system of beliefs and assumptions about the development of knowledge (Saunders et al., 2016, p. 124). Creswell (2007, p. 15) states that during the research process, the researcher carries their own set of worldviews, beliefs and paradigms to a research project. As such, when using a specific research method, certain assumptions about philosophical aspects are made (Clark, 1998). Moreover, Saunders et al. (2016, p. 124) states that the researcher may not consciously be aware of the types of assumptions that are being made pertaining to knowledge (epistemological assumptions), the realities faced in the research (ontological assumptions) and the extent in which one’s values influence the research

process (axiological assumptions). According to Pansiri (2009), in order to select the appropriate methods one must have a broad understanding of different paradigms and their application to research. Following this description, paradigms are discussed next.

The term “paradigm” was first used in 1962 by Thomas Kuhn and originated from the Greek word *paradeigma* which means “pattern”. According to Kuhn (1970), the term “paradigm” refers to a research culture with a set of beliefs, values and assumptions that a community of researchers has in common regarding the nature and conduct of research. Collis and Hussey (2014, p. 43) state that a *research paradigm* is a philosophical framework that guides the way in which scientific research should be conducted and that the choice of paradigm is influenced by the researcher’s philosophies and their assumptions about the world and the nature of knowledge. For centuries, positivism existed as the only research paradigm from which knowledge was constructed – known today as the natural sciences (Collis & Hussey, 2014, p. 43). The emergence of the social sciences led to the development of new paradigms. Saunders et al. (2016, p. 151) describe five major philosophical paradigms, namely, positivism, critical realism, interpretivism, postmodernism and pragmatism. Of these, Collis and Hussey (2014, p. 54) state that positivism and interpretivism, as depicted in Figure 6.2, represent the two extremes on the continuum of paradigms, from which the researcher must identify the paradigm most suited to the study from an ontological and epistemological position.

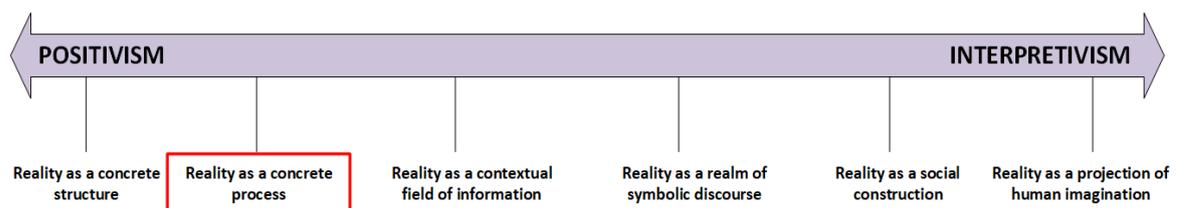


Figure 6.2: Typology of ontological assumptions on a continuum of paradigms (adapted from Morgan & Smircich, 1980, p. 492)

Pertaining to these two extremes, positivism is typically associated with quantitative approaches and interpretivism with qualitative. On the other hand, Trochim (2006) categorised paradigms into positivism and post-positivism – the latter, Trochim (2006) remarked, was a shift into a new era of philosophical thinking. In order to determine which paradigm is best positioned for this study, it is appropriate to discuss each of these paradigms next.

6.4.1 Positivism

According to Collis and Hussey (2014, p. 44), positivism is underpinned by the belief that reality is independent of us, and the goal is to identify theories based on empirical research conducted via scientific methods, in particular observations and experiments. Knowledge is derived from “positive information” because it can be scientifically verified as a means of justification or proof. In doing so, this would advantageously yield pure data and facts not influenced by human interpretation or subjective bias, thus suiting the positivists (Saunders et al., 2016). Saunders et al. (2016, p. 137) elaborates further that positivists attempt to remain neutral and detached from the research and the data in order to avoid influencing their findings. A positivist researcher may use existing theory to aid in the development of a theoretical framework and hypotheses for the study (Collis & Hussey, 2014, p. 77). These hypotheses would then be tested and accepted, completely or partially, thus leading to the further development of theory which subsequently may be tested again in the future (Saunders et al., 2016). From this description, it appears that positivism can be appropriately aligned with deductive reasoning (see Section 6.6).

Theories provide the basis for explanation, permit the anticipation of phenomena, predict their occurrence and therefore allow them to be controlled (Collis & Hussey, 2014 p. 44). Explanation consists of establishing causal relationships between the variables by means of causal laws and linking them to a deductive or integrated theory. Since it is assumed that the phenomena being investigated can be measured, it is thus associated with quantitative methods of analysis. This is supported by Saunders et al. (2016) who state that positivism is generally associated with quantitative methods of analysis, especially when it involves the use of predetermined and highly structured data collection techniques.

6.4.2 Post-positivism

While positivists use a quantitative approach to investigate a phenomenon, post-positivists attempt to explore and describe a phenomenon in depth. Opposition to positivist epistemologies stems from anthropology, ethnography, feminism, post-structuralism and critical psychology (Ryan, 2006). In support of post-positivism, Clark (1998) strongly suggests that positivism is an outmoded and rejected philosophy which should cease to significantly

shape inquiry. Furthermore, Patomäki and Wight (2000) state that positivism is both epistemologically and ontologically flawed and, in addition, is partly responsible for many of the social ills and political catastrophes of the modern world. Loughlin (2012) compares the two approaches by describing positivism as methodologically dogmatic and post-positivism as a “Socratic method”.

Post-positivism has “a certain pluralism”, balancing both positivist and interpretivist approaches (Panhwar et al., 2017), and is influenced by the philosophy of critical realism (see Section 6.4.4). This is also posited by Levers (2013), who states that a post-positivist paradigm is “conceptualized as having an objectivist epistemology and critical realist ontology”. According to Trochim (2006), a post-positivist is a critical realist. This author recognises that all observation is imperfect and can potentially contain error and, furthermore, theory can be modified. Post-positivism accepts that truth and universal laws exist but discovery of these truths is near impossible (Levers, 2013). Thus, a critical realist judges or questions our ability to confidently know reality (Trochim, 2006). Fox (2008) proposes the main features of post-positivism as follows:

- Post-positivism acknowledges that the objects of study are themselves engaged in an ongoing mission of constructing the social world, and therefore their interpretation of experiences must in itself become part of the subject matter of a social science and accept the context-specificity of knowledge.
- It recognises that human beings are the subjects of study and, as such, act as interpreters of the world.
- It requires social scientists to be “reflexive” about their interpretative work, to be objective but simultaneously to accept its ultimate impossibility.

As mentioned, post-positivists are typically of the view that measurements can be imperfect, and thus recognise the need for multiple measures and observations, and also that such may possess different types of errors. Similarly, Creswell (2007, p. 20) points out that post-positivist researchers will likely believe in multiple perspectives from participants rather than a single reality, and adopt rigorous methods of qualitative data collection and analysis. They will use different techniques for data analysis to ensure rigour, use statistical analysis programs, encourage the use of valid approaches, and write their qualitative studies in the form of scientific reports.

From this discussion on post-positivism, it is possible that post-positivist researchers may use both quantitative and qualitative approaches. Loughlin (2012) captures this well: “there is no such thing as ‘a post-positivist approach’, only post-positivist approaches.”

6.4.3 Interpretivism

As a result of the inadequacies in positivism in regard to the needs of social scientists, interpretivism emerged (Collis & Hussey, 2014). Interpretivism is underpinned by the belief that social reality is subjective because it is influenced by our perceptions (Collis & Hussey, 2014, p. 45). To elaborate, the researcher is involved with the phenomena being investigated and as a result, the researcher’s thoughts and opinions cannot be separated from those phenomena. In contrast to positivists, who measure the frequency of a phenomenon, interpretivists adopt methods that seek to describe, translate and find meaning. As a result, conclusions are general as they are not derived from any statistical analysis. Interpretivism emphasises that humans are different as they originate from different cultural backgrounds, under different circumstances and at different periods and thus will create (i.e. interpret) their own subjective meanings from physical phenomena (Saunders et al., 2016, p. 140).

6.4.4 Critical realism

Critical realism is a branch of the realist philosophy initially developed by Bhaskar (1978), and is described as transcendental realism primarily focused on the ontological perspective which explains that reality exists independently from our mind, regardless of what we see and experience, in terms of the underlying structures of reality that shape the observable events (Saunders et al., 2016). Similarly, Patomäki and Wight (2000) point out that in critical realism the world is composed not only of events, states of affairs, experiences, impressions and discourses, but also of underlying structures, powers and tendencies that exist, whether or not identified or known through one’s experience and/or discourse. Critical realism, then, differs from empirical and linguistic realism in viewing the world as, in part, composed of objects, including causal laws which are structured and “intransitive” (a term coined by Bhaskar). This intransitive dimension to the world is irreducible to events and their patterns.

6.4.5 Pragmatism

Pragmatism is not committed to any one system of philosophy and reality, and asserts that the research question should determine the research philosophy and that methods from more than one paradigm can be used in the same study (Collis & Hussey, 2014, p. 54). This is supported by Morgan (2014), who argues that pragmatism can be used as a paradigm for social research, regardless of whether that research uses qualitative, quantitative or mixed methods. Thus, a pragmatic researcher is more concerned with the outcome of the research and less with the specific methods employed to reach the outcome. This explanation is similar to Creswell's (2003, p. 4) that research is currently less concerned with what constitutes quantitative and qualitative research than with how research practices lie somewhere on a continuum between the two.

6.5 Paradigm adopted in this study

As it is possible for researchers to adopt a paradigm that is neither strictly positivist nor phenomenological in nature, most researchers today would prescribe to an intermediary paradigm using elements from the different philosophical paradigms (Collis & Hussey, 2014, p. 50). Initially, of all the paradigms discussed, this study largely supports the *post-positivist paradigm* on the basis that the study identified a theory and then tested the theory. However, in the view of this study, the researcher is testing a theory that may deduce different results from the original theory. The social nature of the subjects used in the current study may unknowingly prevent the results from being generalised. Because primary data is collected by means of a survey and then analysed using descriptive and inferential statistical methods, interpretivism is not appropriate. Moreover, pragmatism is typically suited for a mixed methods research design; however, as this study uses a cross-sectional and mono-method quantitative strategy, this paradigm is also not suited. Given this justification, the choice of paradigm most appropriate for this study is *post-positivism*.

6.6 Research approach

Deductive reasoning, informally referred to as the “top-down approach” (Trochim, 2006), is a study in which a conceptual and theoretical structure is developed from the literature and then tested by empirical observation (Collis & Hussey, 2014, p. 7; Saunders et al., 2016, p. 145).

For example, the researcher identifies an appropriate theory relevant to the study and intends to scientifically test the hypotheses and theory in the context of the study. This involves collecting data assisted by the theories that were initially identified by the researcher. Further, the researcher must specify how data can be collected in relation to the concepts that form the hypotheses (Bryman, 2001, p. 8).

Inductive reasoning, informally referred to as the “bottom-up approach” (Trochim, 2006), is the opposite of deductive reasoning. In inductive reasoning, theory is developed from observation of the empirical reality, thus moving from individual observations to statements of patterns, or law (Collis & Hussey, 2014, p. 7). Inductive reasoning is, by its very nature, more open-ended and exploratory (Trochim, 2006). For example, the researcher observes a particular problem or phenomenon in a particular context and forms a conclusion on what has contributed to that situation. Figure 6.3 depicts the deductive reasoning approach used in this study.

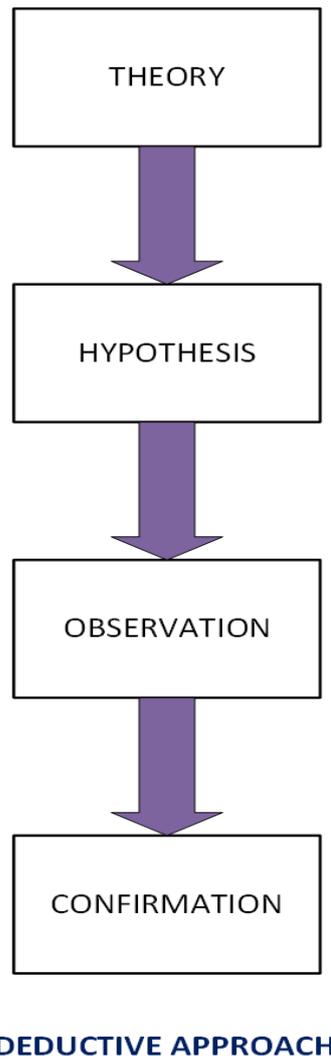


Figure 6.3: Deductive reasoning (adapted from Trochim, 2006)

Trochim (2006) notes that most social research involves both inductive and deductive reasoning processes at some time in the study. This view is also supported in this study, as during the different phases of this study design, both inductive and deductive reasoning were employed. To elaborate, inductive reasoning was used during the preliminary investigation which involved seeking and identifying relevant secondary data (exploratory), leading to the formation of the research model and associated hypothesis. Subsequently, deductive reasoning, by means of statistical analysis (confirmatory), was applied to evaluate the hypotheses of the study, as outlined in this chapter. This approach also aligns with the aforementioned choice of a *post-positivist* paradigm. Moreover, it also aligns to the factor analysis approaches discussed in Section 6.10.1.

6.7 Literature review process

According to Levy and Ellis (2006), undertaking in an effective literature review for one's particular study holds a number of advantages:

- An effective literature review assists the researcher to understand the existing body of knowledge in terms of where it was (the past), where it currently is (the present) and what needs to be known (the future).
- It provides a solid theoretical foundation.
- It presents and argues the existence of the research problem.
- It justifies the significance of the research, thus making a contribution to the existing body of knowledge.
- It provides guidance on how to structure appropriate research methodologies, approaches, goals and research questions pertaining to the study.

This section aims to discuss the procedures for identifying “what” type of literature sources were relevant to the study (e.g. conference papers, journal articles, books, theses), “which” literature works (i.e. scope) were appropriate for the problem being investigated (e.g. information security, human factors) and “where” these sources were located (e.g. Academic databases: IEEE Xplore, ScienceDirect, Emerald Insight, ACM), as well as to discuss the process that answers “how” the literature was organised and structured.

In general, the internet offers a plethora of information to researchers and users alike. This has been referred to as the “information explosion”, with the number of articles published on the internet increasing dramatically, with some having open access (Booth et al., 2016, p. 13). It is thus important to distinguish, in the sheer volume of information available, between what is relevant and what is irrelevant, as well as to reduce information overload (Booth et al., 2016). Hence, it is appropriate to suggest that one of the first procedures when commencing the literature investigation is to define the scope, thereby limiting the sources that are relevant to one's particular study (Collis & Hussey, 2014, p. 77). This is regarded as the *search strategy* and can be a challenging task given that today's IS research is interdisciplinary. This is noted by Webster and Watson (2002) who state that researchers may focus on a particular discipline while unintentionally excluding literature in another sub-discipline related to the IS literature. For example, this study includes literature from the information security discipline, but it also

considers literature from the psychology discipline (i.e. personality traits and habits) owing to the behavioural variables of this study. It is therefore important to ensure that the process of including relevant material pertaining to one's study is performed accurately without unintentionally excluding relevant literature.

One of the basic means of locating information on the internet is to use a keyword search by means of web search engines. However, in the world of scientific research, a basic "Google search" is not adequate to accurately locate scholarly work in a specific focus area. This is because anecdotal evidence, found in blogs and wikis, could be included in the results. In this study, Google Scholar was used as the starting point for a keyword search. Although it is regarded as a web search engine, it indexes the full text or metadata of "scholarly literature" across a wide variety of disciplines and sources. Besides the capability of locating keywords, the advanced features of Google Scholar allow the researcher the flexibility to "fine-tune" the search function to locate literature over a particular period of time. In this regard, the search was further refined to more recent works, as literature from 2011 to 2021 was identified for this study. Furthermore, multiple keyword searches were performed on specific subject areas related to the study using Boolean operators (noted by Collis & Hussey, 2014, p. 79). For example, in this study the search string used in some search engines and databases were ("phishing" OR "phishing attacks") AND ("Information processing" OR "heuristic-systematic processing" OR "cognitive processing") AND ("habits") AND ("personality traits" OR "Big Five personality traits") AND ("social media" OR "Facebook").

Other considerations in the search strategy were to select conference papers, however this was done with caution, based on their reputation and quality, as advised by Webster and Watson (2002). In addition, ResearchGate, a social networking site for scientists and researchers, was used to obtain access to conference papers and journal articles that might have been restricted on the host site. Elsevier's ScienceDirect academic database, in particular journals, such as *Computers & Security*, *Information and Management*, and *Computers in Human Behavior*, provided reliable works for the study. The metrics for these journals pertaining to their impact factor and CiteScore were also considered. To elaborate, at the time of writing, the journal *Computers in Human Behavior* had a CiteScore of 12.1 and an impact factor of 5.003. Additionally, these academic databases contained statistical analysis techniques, such as SEM, which were relevant to this study. Theses and dissertations were accessed from digital libraries such as the South East Academic Libraries System (SEALS).

SEALS is an academic library forum in the Eastern Cape Province of South Africa containing the member libraries of academic institutions that include the Nelson Mandela University (NMU), Rhodes University (RU), University of Fort Hare (UFH), and Walter Sisulu University (WSU). Additional sources had to be identified as this study fits predominantly in the psychology domain. As such, it was inevitable that overlaps would emerge within the current focus area of the study. This resulted in duplicates being created that needed to be identified and removed where necessary. For example, Vishwanath et al. (2018) discuss both habits and information processing in a single phishing study. Table 6.1 expands on the identification process and categorises the specific literature works that were examined for the study.

Table 6.1: Categories of literature work in the identification and screening process

Focus area	No. of articles	No. of articles after screening
Phishing	280	160
Personality traits	129	118
Social engineering/persuasion	59	54
Information processing	54	45
Habit	48	43
Computer self-efficacy	19	10
Social network threats	10	8
Social norms	6	6
TOTAL	605	444

Part of any research study, and not included in Table 6.1, is to investigate literature pertaining to the research design and methodology. For example, literature pertaining to guidelines on the use of structural equation modelling (n = 55) and articles adopting SEM (n = 61) were also necessary for this study.

In conclusion, all of the above-mentioned considerations were used to further enhance the search results, thus ensuring that relevant literature pertaining to the topic was not omitted. Once the sources were identified for the study, a process to *organise* the relevant literature into

the main categories of work/themes in a logical manner had to be undertaken. Hofstee (2006, p. 94) recommends using the funnel approach, as depicted in Figure 6.4. In this way, a process can be followed to categorise literature, thus improving the organisation of literature as well as the flow of arguments between the various literature sections.

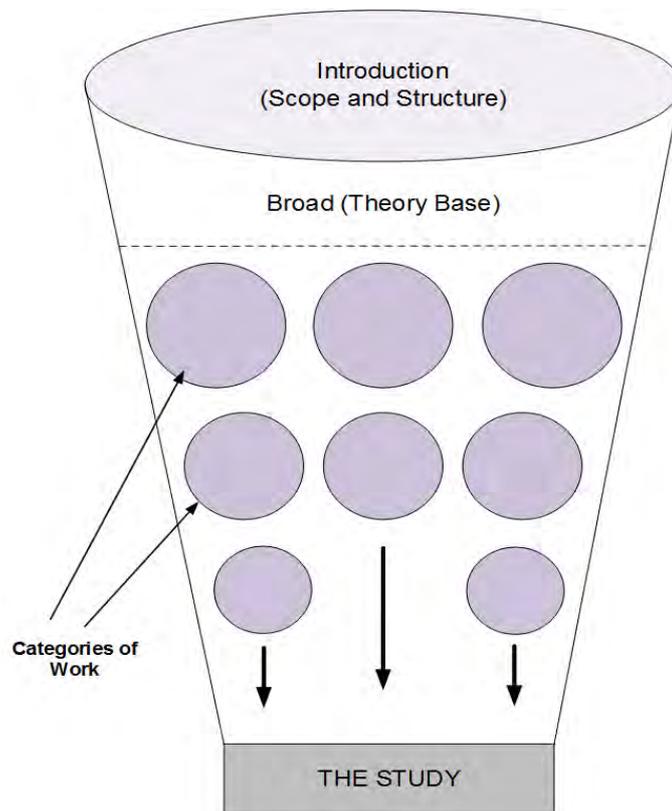


Figure 6.4: The funnel method used to structure a literature review (adapted from Hofstee, 2006, p. 96)

In this study, the broad categories of work that were identified in the literature review process were social engineering (SE) and phishing. More focused areas included personality traits, habits and information processing. Identifying the categories of literature works, from the broad to the specific, helped organise the structure of the literature chapters, as well as the content within the chapters. Subsequently, three main literature chapters emerged (Chapters 2–4) from the literature study.

6.8 Population and sample

The size of the sample influences the ability of the model to be estimated correctly, as well as the specification error to be identified. The determination of an appropriate sample size has been remarked by many scholars as a contentious issue for studies employing structural equation modelling (Kline, 2016; Rigdon et al., 2017; Schumacker & Lomax, 2016; Tanaka, 1987; Weston & Gore, 2006). Barrett (2007) goes further by advocating for structural equation modelling (SEM) publications to be withdrawn if samples are less than 200. Weston and Gore (2006) recommend a minimum sample size of 200 for any SEM study, provided the researcher anticipates no problems with the data. Coincidentally, Barrett (2007) provides an example, stating that the use of 200 first-year undergraduate university students is acceptable for a model using or invoking variables which are deemed universal amongst all members of the population, regardless of age or gender. In addition, Barrett (2007) cautions that determining a sample size based on the *power* of a model is futile, as there is no straightforward test of this kind. In addition, Schumacker and Lomax (2016, p. 120) state that the estimation of an adequate sample size and/or power in SEM “is complicated because theoretical models can have several variables or parameter estimates, and parameters are typically not independent in a model, and can have different standard errors”. In determining the size of a sample, the researcher has to consider a number of factors, namely the confidence one has in the data, the acceptable margin of error one will tolerate, the type of data statistical techniques one will be using for the analysis and the size of the target population (Saunders et al., 2016, p. 279). In this study, the sample size was computed using the method formulated by Yamane (1967). With a known population of 477 (N), a confidence interval at 95% and an error level at 0.05 (e), the sample size required for this study was 213.

The method used to collect primary data included an online questionnaire deployed using SurveyMonkey. Data were collected from respondents located at three university campuses in South Africa. The respondents selected in this study were 285 final-year students. Students were chosen because 65% of Facebook users are typically under the age of 35, are actively engaged on SNSs (Dixit & Prakash, 2018) and are driven by the need for a social presence and for entertainment value (Cheung et al., 2011). Moreover, the choice of final-year students was based on the notion that they would presumably be entering the workspace the following year and might bring security risks for their prospective employers. This is supported by Bailey et al. (2008), who found that college students were more susceptible to email phishing attacks.

The students thus should have had sufficient knowledge and experience in using SNSs (Vishwanath, 2015b) and, as such, it was anticipated that they had established particular habits when using SNSs.

6.9 Variable descriptions and measures

The measures and individual items for each construct were adopted from prior studies as they had been proven to be statistically reliable. All of the variable scales can be found in Appendix B. Some of the items for each of the variables were also designed in compliance with the recommendations of Podsakoff et al. (2012) to minimise bias. This is further discussed in Section 7.5.1. The variables used in this study are discussed in more detail next.

6.9.1 Personality traits

The public domain instrument known as the Big Five Inventory (BFI) scales test by John and Srivastava (1999) was used to determine the personality traits. This instrument consists of a total of 44 items scored on a five-point Likert scale (1 = disagree strongly to 5 = agree strongly). The BFI scales, fitting each of the traits, range from eight to ten items. The test determines into which of the five personality traits a person's personality predominantly fits. This measure was used to identify which personality characteristics or traits had an influence on habit and participants' ability to process phishing messages, which could ultimately lead to phishing victimisation. This personality test has been shown to have solid psychometric properties when compared to other even more comprehensive personality tests (John & Srivastava, 1999). This was also supported by Boudreaux and Ozer (2015) who, in their comparison of numerous Big Five scales, state that the BFI by John and Srivastava (1999) is considered highly reliable (Cronbach alpha coefficients range from 0.75 and 0.90, with an average above 0.80), stable over time (3-month retest coefficients range from 0.80 to 0.90, with a mean of 0.85), and possess convergent and discriminant validity with respect to other Big Five instruments. Moreover, this particular instrument has been adopted in other behavioural studies in information security (Pattinson et al., 2012; Van der Schyff, Flowerday, & Lowry, 2020).

6.9.2 Habit

Verplanken and Aarts (1999) recommend that researchers focus on habitual mindsets and automatic cue-response links instead of the associations between past and future behaviour. Accordingly, the 12-item Self-Report Habit Index (SRHI) of habit strength scale was used in this study to measure habit with the main focus on capturing *automaticity* (Verplanken & Orbell, 2003). As discussed in Section 2.5.1, many different types of behaviours can be performed in a social network environment. Accordingly, this study considers the behaviour of clicking, opening or sharing a link on SNSs as requested by a friend (as a form of compliance), as potentially leading to the formation of habits which can put them at risk to phishing. Items were scored on a five-point Likert scale (1 = disagree strongly to 5 = agree strongly). Verplanken and Orbell (2003) posit that measuring habits entails more than simply measuring the frequency of past behaviour and, as such, the SRHI is useful when one wants to determine the role of habit without measuring behavioural frequency. According to Gardner (2015), the SRHI focuses on aspects of automaticity such as lack of awareness (item: “I do without thinking”), lack of control (item: “that would require effort not to do”) and mental efficiency (item: “I have no need to think about doing”), behavioural frequency (item: “I do frequently”) and self-identity (item: “that’s typically ‘me’”). According to Gardner (2015), at the time of writing, the SRHI has been used in 119 studies (88%). Apart from the popularity of the scale in the health psychology and medicine discipline, it was also adapted by Vance et al. (2012) in the context of compliance with information security policies. This scale has also been used in the context of social media usage (Soror et al., 2021) and phishing (Vishwanath, 2015a).

6.9.3 Information processing

As discussed in Chapter 2, incorporating persuasion strategies into messages is an effective means of deceiving users into performing actions that will benefit the phisher. As the context of this study is SNSs, it was thus necessary to identify the appropriate phishing messages or stimuli associated with SNSs, specifically Facebook, in order to include this construct in the model. This objective was accomplished in Chapter 2. Facebook was chosen as it is the most popular SNS in the world and most student populations use the platform. Considering that 98.3% of Facebook users access the site via their smartphone (Tanhovska, 2021), actual

screenshots of stimuli were derived from the Facebook smartphone app. The survey instrument assessed respondents on six persuasive messages related to Facebook and personally obtained by the researcher. This is consistent with the approach of Butavicius et al. (2015), Chen et al. (2018) and Parsons et al. (2019), who used “actual” phishing emails. In an attempt to use stimuli that are of interest or liked by the respondents, South African content was utilised that the participants would be familiar with and potentially would be interested in clicking. This supports Petty and Cacioppo's (1986) assertion that persuasion is increased if the message is relevant to the audience. This was also confirmed in a study by Hassandoust et al. (2020), which found that respondents used heuristic processing when the messages they received were specific to their context. Table 6.2 describes the stimuli used in this study.

Table 6.2: Description of stimuli used to test information processing

Stimuli #	Description	Action required	Persuasion technique
Stimuli (S1)	Presents an opportunity to win a free store voucher worth ZAR1500. The voucher contains an expiry date.	Click/share	Authority and scarcity
Stimuli (S2)	The source offers an opportunity for people to win a silver Mercedes Benz vehicle. Two lucky draws. The draws are said to take place within the next two days.	Comment, like and share	Scarcity and social proof
Stimuli (S3)	Breaking news reporting that a famous local athlete Caster Semenya has died in a car accident. The video claims to show actual footage of the accident.	Click link	Curiosity
Stimuli (S4)	Opportunity to have financial freedom. Image shows a proof of payment received.	Comment with personal info (i.e. contact number)	Scarcity and social proof
Stimuli (S5)	Video thumbnail showing a person appearing to be robbed. The video indicates that it has been viewed 11 810 727 times.	Click play	Curiosity
Stimuli (S6)	Video thumbnail showing an altercation between workers at Marikana mines.	Click play	Curiosity

Although not part of Cialdini’s principles, most of the stimuli used fitted the “curiosity” technique which has been shown to be very effective in phishing (Blythe et al., 2011; Moody et al., 2017). None of the screenshots contained spelling errors, which the literature recommends as one of the cues that may assist in identifying phishing. As the primary focus of the study was not to determine which persuasion principle is most effective, not all persuasion principles were tested. The screenshots illustrated that a particular action was required from the user (e.g. to click on play). The purpose of including a variety of different phishing cases was to address the respondents’ potential bias as they might give more attention to some messages than others based on their interests or prior encounters. Heuristic processing

was measured by adopting a four-item scale used in prior research (Griffin et al., 2002; Vishwanath, 2015b), while systematic processing was measured using a three-item scale also adapted from prior research (Griffin et al., 2002; Vishwanath et al., 2011). Both the heuristic and systematic items were scored on a five-point Likert scale (1 = disagree strongly to 5 = agree strongly). The above-mentioned items in each of the stimulus were combined, thus consisting of a total of seven items per stimulus. Separating items according to whether they were heuristic or systematic could potentially influence respondents to respond in a way that they may consider morally acceptable rather than reflecting their true behaviour. Notably, Harrison et al. (2015) also measured systematic processing and heuristic processing separately.

6.9.4 Phishing susceptibility

Figure 6.5 presents a phishing email containing an attachment, purportedly originating from Facebook, which was used to test susceptibility to phishing. The email is designed to appear as if it originated from Facebook, with the address being `update@facebookmail.com`. It also employs the blue theme typically associated with Facebook branding. Similar to the study by Moody et al. (2017), this dependent variable was measured on a binary scale coded as 0 = not susceptible; 1 = susceptible. The items: "Reply to the email" and "Check the attachment because I am interested to know what my friend has to say" were considered to be items related to phishing susceptibility. Not susceptible was represented by the items: "Immediately delete the email", "Ignore the email" and "I do not trust this email". The item "Unsure" was considered to be a missing observation and not included in the analysis as it did not inform the exact position of the respondent's choice.

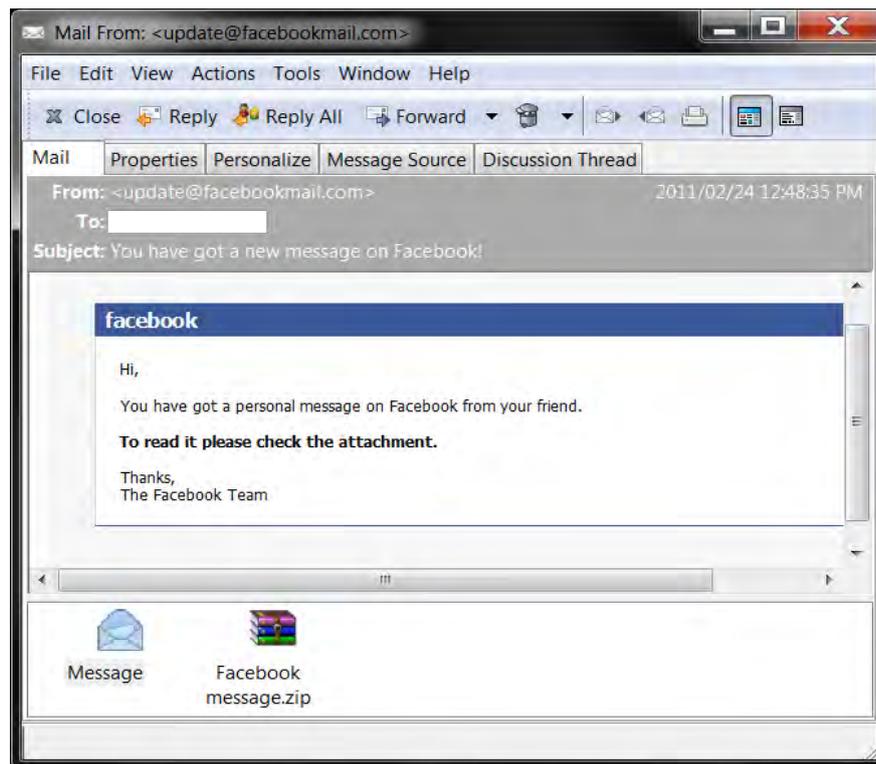


Figure 6.5: Phishing email purportedly originating from Facebook

6.9.5 Social norms

The scale consisted of six items scored on a five-point Likert scale (1 = disagree strongly to 5 = agree strongly). Respondents were required to rate their decision based on when to share a friend's post if they requested them to share it. Sample items include: "If it is my friend, then it is the friendly thing to do"; "If the post is popular (i.e. trending) and I know others will also find it interesting".

6.9.6 Computer self-efficacy

Compeau and Higgins (1995) developed a ten-item measure for computer self-efficacy (CSE); however, this measure is several years out of date and focused mainly on the use of general computer tasks and not on tasks related to phishing avoidance. Hocevar et al. (2014) developed social media self-efficacy scales that focused on an average of a person's (a) perceived social media skill, (b) confidence in the ability to successfully find information online, (c) level of

social media content production, and (d) level of social media content consumption. While the latter scale fits our context, it was also not deemed appropriate for the current study. Subsequently, this study defined computer self-efficacy scales as being more aligned to the study's context and considered items related to performing specific computer tasks, social media experience, general computer self-efficacy and their knowledge of some technical features required to protect or evaluate potential phishing attacks. This view is supported by Marakas et al. (2007), who state that measures for computer self-efficacy should be redesigned with articulated alignment to the situation under study. As such, this study measures task-specific computer self-efficacy. Marakas et al. (2007) define this as "an individual's perception of efficacy in performing specific computer-related tasks within the domain of general computing". The current study's scale consisted of eight items scored on a five-point Likert scale (1 = very poor to 5 = excellent) and required respondents to evaluate their abilities to perform certain tasks using a computer system.

6.9.7 Perceived risk

The scale consisted of four items measured on a five-point Likert scale (1 = disagree strongly to 5 = agree strongly). Sample items include: "There is little risk in sharing posts which instruct you to share to your profile (e.g. share this post and R100 will be donated to a charity)"; "There is little risk in accepting friend requests from strangers, as I can remove them later if I want to". As discussed in Section 2.6, the literature classifies various types of perceived risk. Van Schaik et al. (2018) suggest that most behavioural research on SNSs has focused primarily on privacy and not on security. This scale focused on the risks related to phishing in an SNS context. Certain measures of the scale were informed by Van Schaik et al. (2018), for example items related to social media behaviour, such as obliging with the sharing of posts and accepting friend requests from strangers. Another item (see Perceived risk item 3) measured respondents' overall awareness of risks pertaining to some form of personal loss that could potentially originate from SNSs (e.g. identity theft, reputational and financial loss). Two "control" items (see Perceived risk items 2 and 4 in Appendix B), informed by Van Schaik et al. (2018) and Nilsson (2009), were also used to ascertain whether the respondents perceived themselves as able to mitigate potential risks.

6.9.8 Demography

The collection of demographic data is important for any study. According to Hammer (2011), omitting demographics in studies stands the risk of assuming the stance of “absolutism”, which assumes that the phenomena of interest are the same regardless of age, race, socioeconomic status, education level and culture. As such, the questionnaire also contained items that captured information pertaining to the respondents:

- Demographics, including gender, age, course of study (multiple items)
- Duration, hours spent on SNSs (open-ended)
- Motivation/reasons for visiting SNSs (6 items)
- Dependency on SNSs (5 items).

However, while some of the demographic data was reported, its involvement in the theoretical model was not measured/included, either because 1) the data analysis confirmed it not to be valid and reliable, or 2) by including these constructs they had adverse effects on the rigour of the structural model. Regarding the latter, if one construct is added or removed it subsequently has a cumulative effect on all other constructs and simultaneously on their associated relationships between the constructs in the model.

6.10 Data analysis

As this study aimed to develop a theoretical model that entails examining the causal effect variables have on each other, covariance-based SEM was identified as the appropriate statistical technique. SEM is known for representing causal relations in multivariate data in the behavioural and social science disciplines (McDonald & Ho, 2002). SEM is not a single technique but rather a collection of related statistical procedures, allowing for theory building and model testing (Kline, 2016, p. 23). SEM can be considered a hybrid of *factor analysis* and *path modelling* (Fan et al., 2016), with factor analysis used to examine the interrelationships among the variables and path analysis used to test the hypothesised relationships among constructs (Weston & Gore, 2006). In this regard, SEM consists of two primary components, the *measurement model* and the *structural model* (Civelek, 2018; Hair Jr et al., 2010).

McDonald and Ho (2002) state that the latter is a composite of the measurement and path models. A more rigorous discussion on the analysis required to satisfy these two components of SEM is provided in Section 7.5 and 7.6 of the next chapter.

6.10.1 SEM approaches and techniques

According to Hair Jr et al. (2017), there are two types of SEM approaches used to estimate the relationships in a model, namely variance-based, also known as partial least squares structural equation modelling (PLS-SEM), and covariance-based (CB-SEM). According to Hair Jr et al. (2017), CB-SEM should be adopted when the goal is “theory testing, theory confirmation, or the comparison of alternative theories”. As such, CB-SEM is appropriate for this kind of research as the study utilised an existing theory (i.e. HSM) to test phishing susceptibility. SEM provides a way to test the relationships among observed and latent variables holistically and allows for theory testing even when experiments are not possible (De Carvalho & Chima, 2014; Savalei & Bentler, 2006). Observed variables are represented graphically by squares or rectangles, while unobserved variables are termed latent factors or constructs and are depicted graphically by circles or ovals (Schreiber et al., 2006). This is depicted in the theoretical model for the study (see Figure 7.1).

Factor analysis typically reduces the number of variables (i.e. data reduction procedure) used to explain a relationship or to determine which variables show a relationship (Yong & Pearce, 2013). Principal component analysis (PCA) and factor analysis are both multivariate techniques and are viewed as approaches for summarising and uncovering any patterns in a set of multivariate data, essentially by reducing the complexity of the data (Jolliffe & Cadima, 2016). *Principal component analysis* transforms a set of related (correlated) variables into a set of unrelated (uncorrelated) variables in an attempt to reduce the complexity of the data by decreasing the number of variables (Jolliffe & Cadima, 2016). It also identifies underlying “factors” that are responsible for the covariation among a group of independent variables. There are two types of factor analysis: exploratory and confirmatory (Hurley et al., 1997; Yong & Pearce, 2013). Exploratory factor analysis (EFA) aims to identify the smallest number of hypothetical constructs that can explain the covariation observed among a set of measured variables (Watkins, 2018). EFA explores the loadings of the variables in order to achieve the best model fit (Yong & Pearce, 2013). Factor analysis, unlike PCA, begins with a hypothesis about the covariance (or correlational) structure of the variables (Landau & Everitt, 2004).

Confirmatory factor analysis (CFA) confirms previously defined hypotheses between the variables from which a theory can potentially be formed (Hurley et al., 1997). CFA is a useful tool for assessing construct validity (Zikmund et al., 2013). This study used both types of factor analysis, as EFA was used in the pilot study to identify the dimensionality of items and to discard the items having low factor loadings.

Following this, CFA was used in the main study, with a different sample, to assess the reliability and validity of the latent variables. This approach is also recommended by Cabrera-Nguyen (2010). Factor loadings are correlation coefficients between observed variables and latent common factors. Eigenvalues are a good criterion for determining a factor. Accordingly, only factors with an eigenvalue equal to or greater than 1.0 were retained. The results were then rotated, employing the varimax with Kaiser normalisation method, to ascertain the loadings of each indicator on its respective construct.

Estimation of the SEM was conducted within the generalised linear model (GLM) framework. GLM estimators are maximum likelihood estimators that are based on a density in the linear exponential family (LEF). Moreover, GLM estimators are typically generalisations of nonlinear least squares and, as such, are optimal for a nonlinear regression model (as with the dependent variable which is binary). The non-Gaussian approach ensured a sufficient representation of the data in the modelling, thus ensuring the results were robust. In the analysis, the generalised structural equations model (GSEM) was employed, taking into account the binary dependent variable used to test phishing susceptibility. The advantage of GSEM is that it considers multilevel data structures, and one is able to use factor (categorical variable as compared to continuous variables) variable notation.

6.10.2 Reliability

Reliability concerns the accuracy of the measurement and the absence of differences if the research were repeated (Collis & Hussey, 2014, p. 52). In other words, it is the extent to which the method produces stable and consistent results which in turn demonstrate the rigour and trustworthiness of the study. *Internal reliability* is particularly important as in this study multiple-item scales were predominantly used. Cronbach's alpha coefficient and composite reliability is one of the most widely used tests in which the reliability is measured based on the

interrelationship between the observed item variables. Composite reliability and Cronbach's alpha coefficient values exceeding 0.70 is considered acceptable for reliability (Hair Jr et al., 2017).

6.10.3 Validity

Discriminant validity refers to the extent in which a construct differs empirically from other constructs. As such, it measures the degree of differences between the overlapping construct. Two approaches can be used to assess discriminant validity, namely the Fornell and Larcker criterion and heterotrait-monotrait (HTMT) ratio of correlation (Ab Hamid et al., 2017). In this study, the Fornell and Larcker criterion (Fornell & Larcker, 1981) was used; this compares the square root of the average variance extracted (AVE) of each construct with the correlation of the other latent constructs. Discriminant validity is established if the square root of the AVE of each construct has a greater value than the correlations with other latent constructs. In this regard, a factor loading exceeding 0.70 is considered adequate.

Convergent validity is the extent to which a measure correlates positively with alternative measures of the same construct (Hair Jr et al., 2017). According to Ab Hamid et al. (2017), convergent validity is established when the factor loading of the indicator, composite reliability (CR) and the average variance extracted (AVE) are considered. An AVE value exceeding 0.5 is deemed acceptable for convergent validity.

6.11 Ethical considerations

Saunders et al. (2016, p. 239) refers to ethics as the standards of behaviour that guide the researcher's conduct with regard to the rights of those who become the subject of their study, or are affected by the study. Collis and Hussey (2014, p. 30) refer to the moral values or principles that form the basis of a code of conduct. In both descriptions, it is evident that the researcher must follow a "guide" or "code of practice" to ensure that data collected from respondents is done so in an ethical way. In this regard, there are several professional bodies which provide guidelines. The following measures ensured that the process was carried out ethically:

- **Adherence to university requirements.** Approval for the collection of primary data was sought from the university's Ethics Committee. Ethical clearance was granted and registered with reference number FLO071SFRA01. Ethical clearance was also granted by the participating university at which primary data were collected.
- **Informed consent.** All necessary information pertaining to the purpose of the study and researcher details was shared with all the respondents before collecting primary data. The terms and conditions of the research project were displayed on the homepage of the survey instrument which required respondents to give their consent before they could commence with the survey (see Appendix C).
- **Confidentiality and anonymity.** Information collected from the respondents was treated confidentially. The research instrument was structured in such a way that no personally identifiable information was captured from the respondents. Although respondents were required to provide demographic information such as gender, age and course of study, this did not violate confidentiality as this information could not be traced back to the specific individual.
- **Honesty and transparency.** The respondents were fully informed about the purpose of the study and were made aware that the results of the study would be used for academic purposes and that the research contributed towards acquiring a PhD (see Appendix C). Further, they were informed that the questionnaire was strictly anonymous. They were also informed that any possible future publication of the study would most likely be available online for them to access.
- **Reciprocity.** The research had mutual benefit in the sense that respondents were made aware of an incentive in the form of a chance to win a prize voucher.
- **Dignity.** If the respondents felt uncomfortable in any way, they were able to withdraw their participation at any time during the survey process. They could also contact the researcher as his email address were provided on the survey homepage (see Appendix C).

6.12 Summary

The *research onion* introduced in Section 6.3 was used as a framework to appropriately organise and structure this chapter. The chapter discussed the paradigm for this study as being that of post-positivism. The study adopted both deductive and inductive approaches at various stages of the research design thus supporting the post-positivist philosophical view. This meant that the researcher initiated the study by first identifying a relevant theory that explained the phenomenon of the study. This was identified by means of a literature investigation from which specific research questions and sub-questions were formulated. The literature process, in terms of organising the literature, was discussed in detail. The strategy used in this study was that of a mono-method quantitative approach and, as such, primary data were collected by means of a cross-sectional survey and analysed using CB-SEM techniques. The results of the analysis are discussed in the next chapter.

7.1 Introduction

Chapter 5 introduced the proposed theoretical model of the study and its associated hypotheses. This chapter presents the results obtained from the statistical analysis techniques in order to ascertain whether the theoretical model is supported by the sample data. In this regard, PCA and CFA approaches were used to validate the measurement model. By accomplishing this, the researcher was able to ascertain how well the observed (measured) variables combined to identify underlying hypothesised constructs. This ultimately led to the formation of the theoretical model for the study along with its evaluation.

7.2 Results of the pilot study

As mentioned in the previous chapter, the purpose of conducting a pilot study was to ascertain the reliability and validity of the construct items and to identify potential problem areas and deficiencies in the research instrument. SurveyMonkey[®], an online survey tool, was used to collect primary data in both the pilot and the main study. The statistical software package, STATA 14, was used to conduct the various data analysis tests. The following sections present the results of the pilot study.

7.2.1 Sample size and demography

Data collection for the pilot study took place between 9 and 13 April 2018. A total of 25 responses were collected of which two were discarded due to incomplete responses. Several

demographic questions were asked at the beginning of the pilot questionnaire, including gender, age, course of study and duration of time spent on SNSs. From a gender perspective, 47.8% of the respondents were male ($N = 11$) and 52.2% were female ($N = 12$). The average age of the respondents was 21.39 years ($SD = 3.60$). In terms of the courses they were studying, 60.9% ($N = 14$) were enrolled in Information Technology and 39.1% in Mechanical Engineering ($N = 9$). Respondents reported spending an average of 7.25 hours per day ($SD = 5.99$) on SNSs (excluding WhatsApp).

7.3 Reliability assessment

As discussed in the previous chapter, the main purpose of the pilot survey was to ensure that the research instrument was reliable and valid prior to conducting the main study. Cronbach's alpha (CA) was used to evaluate the reliability of each of the constructs. CA values exceeding 0.70 are considered reliable.

Table 7.1: Summary of the results of the reliability tests for the pilot study

Constructs	Items	CA
Extraversion	8	0.695
Agreeableness	9	0.646
Conscientiousness	9	0.629
Neuroticism	8	0.745
Openness	10	0.640
Habit	12	0.732
InfoProcessing1	7	0.774
InfoProcessing2	7	0.749
InfoProcessing3	7	0.519
InfoProcessing4	7	0.755
InfoProcessing5	7	0.554
InfoProcessing6	7	0.558
Social norms	6	0.847
Computer self-efficacy	8	0.809
Perceived risk	4	0.811
Phishing susceptibility	6	0.733

As depicted in Table 7.1 some of the CA values for the constructs were not deemed reliable as they are below 0.7. Nevertheless, they were included in the final questionnaire for the main study in the expectation that a larger sample size would increase the CA value.

7.4 Results of the main study

7.4.1 Sample size and demography

Data collection for the main study took place between 4 June and 11 July 2018. This study used a convenience sample of *final-year* undergraduate students from a university located across three different sites. The total population consisted of 477 final-year students. Prior to data collection, ethical approval was granted by the university where the target sample was located. As the study aimed to achieve a 95% confidence level, a minimum of 213 users were required (Kothari, 2008). However, this study intended to collect as many responses as possible, ultimately managing to collect data from 285 respondents. Early analysis detected 70 cases to have incomplete responses and these were therefore removed. The final sample consisted of a total of 215 respondents of which 114 were male (53%) and 101 were female (47%). Respondents had a mean age of 22.6 years ($SD = 4.41$). Respondents reported spending an average of 5.69 hours per day ($SD = 4.47$) on SNSs (excluding WhatsApp).

7.4.2 Reliability and validity assessment

Following data collection for the main study, reliability and validity tests were conducted. As there are five classes of personality traits measured in a single instrument, it was important to establish the reliability and validity of the traits. As discussed in the pilot study, reliability tests were also conducted for the main study. In addition, as discussed in Section 6.10.1, two approaches were used to test validity, namely, PCA, a type of factor analysis, and Pearson's product moment correlation coefficient. As depicted in Table 7.2, the reliability and validity of the variables were analysed jointly for the personality traits construct.

Table 7.2: Results of the reliability tests for personality traits

Latent Construct	Original items	Items correctly loading	Items incorrectly loading	CA	Adjusted CA (revised items)	Factor loading (%)
Extraversion	1, 6, 11, 16, 21, 26, 31, 36	1, 6, 11, 21, 31, 36	12, 27	0.552	0.731 (7)	75%
Agreeableness	2, 7, 12, 17, 22, 27, 32, 37, 42	2, 17, 22, 32, 42	-	0.561	0.701 (8)	56%
Conscientiousness	3, 8, 13, 18, 23, 28, 33, 38, 43	8, 18, 23, 43	4, 12, 27, 29, 37	0.703	- (9)	45%
Neuroticism	4, 9, 14, 19, 24, 29, 34, 39	9, 14, 19, 24, 29, 34, 39	35, 43	0.734	- (8)	88%
Openness	5, 10, 15, 20, 25, 30, 35, 40, 41, 44	5, 10, 15, 20, 25, 30, 40, 44	3, 7, 13, 16, 26, 28, 33, 38	0.600	0.785 (9)	80%

However, after conducting the initial reliability and validity tests, some inconsistencies were detected. As depicted in Table 7.2, in each of the Big Five traits, some items did not belong to a particular trait. This was demonstrated by their factor loading showing their association to a different trait. Prior to adjustments, this also resulted in unreliable CA values for each particular construct. As such, Pearson's product moment correlation coefficient (r) was used to measure the strength and direction of the linear association that exists between two variables on a scale (Collis & Hussey, 2014, p. 275). Values can range from +1 to -1 (Zikmund et al., 2013, p. 562) with the following guidelines for interpreting the associations and their direction: small (0.1 to 0.3 and -0.1 to -0.3), medium (0.3 to 0.5 and -0.3 to -0.5) and large (0.5 to 1 and -0.5 to -1). The results of conducting this test on all five personality constructs are reported in Tables 7.3 to 7.7. The numbers indicated in the top row of these tables represent the item number in the questionnaire and are associated with that particular trait. Overall, these tables indicate that all the coefficients are statistically significant.

Table 7.3: Results of the validity test for Extraversion

	1	6	11	16	21	26	31	36
Pearson correlation	0.695*	0.393*	0.638*	0.473*	0.561*	0.494*	0.488*	0.586*
Significance (2-tail)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

* = $p < 0.05$

Table 7.4: Results of the validity test for Agreeableness

	2	7	12	17	22	27	32	37	42
Pearson correlation	0.464*	0.317*	0.432*	0.565*	0.449*	0.425*	0.588*	0.503*	0.570*
Significance (2-tail)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

* = $p < 0.05$

Table 7.5: Results of the validity test for Conscientiousness

	3	8	13	18	23	28	33	38	43
Pearson correlation	0.501*	0.513*	0.603*	0.605*	0.605*	0.408*	0.548*	0.558*	0.447*
Significance (2-tail)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

* = $p < 0.05$

Table 7.6: Results of the validity test for Neuroticism

	4	9	14	19	24	29	34	39
Pearson correlation	0.445*	0.608*	0.577*	0.601*	0.564*	0.567*	0.468*	0.655*
Significance (2-tail)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

* = $p < 0.05$

Table 7.7: Results of the validity test for Openness

	5	10	15	20	25	30	35	40	41	44
Pearson correlation	0.609*	0.510*	0.576*	0.634*	0.566*	0.523*	-0.280*	0.705*	-0.430*	0.487*
Significance (2-tail)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

* = $p < 0.05$

Table 7.8 reports on the results of the reliability tests conducted for the information processing construct. This construct consisted of six stimuli (labelled as InfoProcessing), each individually assessed for reliability. Following a few adjustments, the validity tests confirmed that the items numbered 1–4 represent heuristic processing and the items numbered 5–7 represent systematic processing. More detail pertaining to the descriptive statistics and loadings for each of the items can be found in Appendix A.

Table 7.8: Results of the reliability tests for the Information processing construct

Stimulus	Items	CA	Adjusted CA
InfoProcessing1	7	0.458	0.700 (5)*
InfoProcessing2	7	0.573	0.769 (5)*
InfoProcessing3	7	0.555	0.710 (5)*
InfoProcessing4	7	0.629	0.751 (6)*
InfoProcessing5	7	0.645	0.734 (6)**
InfoProcessing6	7	0.650	0.755 (6)**
TOTAL	42		0.875

Note: *items 2 and 3 were removed and only five items remained

**item 3 was removed and only six items remained

() indicates revised items

Following adjustments, the CA values for each stimuli ranged between 0.70 and 0.77. The overall CA score for the Information processing construct was 0.875, thus indicating

satisfactory reliability. Table 7.9 presents the results of the reliability tests for all the constructs included in the main study.

Table 7.9: Summary of the results of the reliability tests for the main study

Constructs	CA
Extraversion	0.731
Agreeableness	0.701
Conscientiousness	0.703
Neuroticism	0.734
Openness	0.785
Habit	0.882
InfoProcessing1	0.700
InfoProcessing2	0.769
InfoProcessing3	0.710
InfoProcessing4	0.751
InfoProcessing5	0.734
InfoProcessing6	0.755
Phishing susceptibility	0.733
Social norms	0.650*
Computer self-efficacy	0.754
Perceived risk	0.731

*Note: *item rounded to 0.7*

As summarised in Table 7.9, the resultant CA values indicate that the variables for the main study are deemed reliable.

7.4.3 Univariate analysis

To further enrich contribution of the main study, univariate (descriptive) analysis was performed on the dispositional factors of habit, social norms, and computer self-efficacy and perceived risk. The univariate data is displayed in the form of frequency distribution tables and is reported next.

Table 7.10: Distribution of the items related to the construct *Habit*

Item	Disagree strongly %	Disagree %	Neither agree nor disagree %	Agree %	Agree strongly %
I do frequently/often.	11.16	14.88	14.42	38.60	20.93
I do automatically.	23.72	22.79	20.00	22.33	11.16
I do without having to consciously remember.	22.33	22.79	19.53	25.12	10.23
That makes me feel weird if I do not do it.	35.81	29.77	10.70	18.60	5.12
I do without thinking.	37.21	27.44	8.37	18.14	8.84
Would require effort not to do it.	31.16	25.58	26.05	10.23	6.98
That belongs to my (daily, weekly, monthly) routine.	28.84	22.79	18.60	20.47	9.30
I start doing before I realize I'm doing it.	36.28	26.05	15.35	15.81	6.51
I would find hard not to do.	34.88	18.60	16.28	24.19	6.05
I have no need to think about doing.	24.65	21.86	19.53	21.86	12.09
That's typically "me".	30.70	20.00	19.53	21.40	8.37
I have been doing for a long time.	29.77	21.86	15.35	19.53	13.49

As reported in Table 7.10, more than half of the respondents (59.53%) indicated that when a friend posts a link on an SNS (e.g. video, image, news article) and requests of them to share it, they *frequently/often* click, open or share it. Depending on how often a friend requests this and the strength of the relationship to the friend, this could develop into habitual behaviour, especially as SNSs encourage sharing. In addition, it also indicates the effectiveness of the authority principle in getting the user to carry out tasks.

When measuring social norms, it is apparent from Table 7.11 that the distribution of most of the items ranges between "agree" and "agree strongly".

Table 7.11: Distribution of the items related to the construct *Social norms*

Item	Disagree strongly %	Disagree %	Neither agree nor disagree %	Agree %	Agree strongly %
If it is my friend, then it is the friendly thing to do.	17.97	9.68	13.82	31.80	26.73
It depends on what it is that I must share.	5.07	2.76	4.61	17.97	69.59
If it is a topic of interest to me personally.	5.53	2.76	8.29	32.26	51.15
The post is very popular (i.e. trending) and I know others will also find it interesting.	11.52	11.52	13.36	31.80	31.80
If I can see many of my friends or others have also already liked it.	24.88	20.28	14.29	21.66	18.89
If it could get me noticed with some likes from my friends.	33.64	17.51	15.21	19.82	13.82

The respondents (58.53%) agreed that if their friend were to request them to share a post, they would consider sharing it because “it is the friendly thing to do”. Additionally, 63.6% agreed that they would share a post if it were popular or trending and perceived that others would also find it interesting. These behaviours can be leveraged by phishers to comply with requests that could benefit them.

As depicted in Table 7.12, for all the items related to computer-self efficacy, it is apparent that respondents perceived their computer abilities, specifically practical tasks such as typing a document, surfing the web, email, and using social network sites, as ranging between good to excellent.

Table 7.12: Distribution of the items related to the construct *Computer self-efficacy*

Item	Very poor %	Poor %	Average %	Good %	Excellent %
Using a desktop computer to type a document (e.g. assignment, CV, report).	0.72	2.15	17.56	41.94	37.63
Using a web browser (e.g. Chrome, Explorer, Firefox) to search for information on the internet.	0	1.80	9.35	42.81	46.04
Using the features of an email client app (e.g. Gmail, Yahoo) to send/receive messages and download/upload attachments	0	6.83	19.78	36.33	37.05
Identifying different file extensions (e.g. .docx, .xlsx, .pdf, .rar, .zip).	0.72	13.31	34.89	31.29	19.78
Using social network websites (e.g. Facebook, Twitter, and Instagram) to post and interact with other users.	0.36	1.44	13.67	30.22	54.32
Checking the security settings of a website to determine if it can be trusted as safe/original.	3.24	23.74	35.97	28.78	8.27
Identifying safe web links/URLs.	4.68	29.14	37.05	20.86	8.27
Installing software on a desktop/laptop computer.	7.55	15.83	23.74	28.42	24.46

The respondents rated their ability to interact with others on SNSs between good and excellent (84.54%), which indicates that the respondents are active social media users. However, pertaining to the items related to security aspects such as identifying safe URLs, viewing security settings on websites, identifying various file extensions, this optimism reduced (between average to very poor). Having competency in each of these areas is necessary to combat phishing as in both email and social media, phishers' direct users to spoofed websites.

Table 7.13: Distribution of the items related to the construct *Perceived risk*

Item	Disagree strongly %	Disagree %	Neither agree nor disagree %	Agree %	Agree strongly %
There is little risk in sharing posts which instruct you to "share" to your profile (e.g. share this post and R100 will be donated to a charity).	28.46	14.23	10.38	16.54	30.38
There is little risk in accepting friend requests from strangers, as I can remove them later if I want to.	24.23	14.23	9.62	17.69	34.23
There is little risk that I can be personally affected on social networking websites (e.g. losing money, identity theft).	32.31	9.62	7.31	13.08	37.69
I am able to protect myself against threats on social networking websites as I have control of my account.	6.92	13.46	11.15	27.31	41.15

As reported in Table 7.13, 46.92% of the respondents did not perceive any risks in sharing posts that instruct them to share to a profile, while 10.38% were unsure if they would share. More concerning is that more than half of the respondents (51.92%) perceived little risk in accepting friend requests from strangers. Moreover, 50.77% agreed that there is little risk of being personally affected by any loss on SNSs. Respondents (68.46%) also seemed optimistic about their abilities to protect themselves against social network threats.

Table 7.14: Univariate results of construct *Phishing susceptibility*

Item	Frequency	%
Reply to the email	17	7.91
Immediately delete the email	21	9.77
Check the attachment	86	40.00
Ignore the email	26	12.09
I do not trust this email	44	20.46
Unsure	21	9.77
TOTAL	215	100

As Table 7.14 indicates, the respondents were susceptible to a phishing email originating from Facebook, as 40% indicated that they would check the attachment. Moreover, only 20.47% of the respondents were suspicious of the phishing email, while 9.77% were unsure.

As mentioned in Section 6.10, SEM consists of the *measurement model* and the *structural model*. The next section focuses on a discussion of multivariate analysis.

7.5 Measurement model evaluation

The SEM measurement model allows the researcher to evaluate how well the observed (measured) variables combine to identify underlying hypothesised constructs. As mentioned in Section 6.10.1, CFA was used to test the reliability and validity of the measurement model. Owing to the Cronbach's alpha (CA) limitations, it is technically more useful for researchers to apply composite reliability (CR) values, as these take into consideration the different outer loadings of the indicator variables (Hair Jr et al., 2017). Similar to the CA requirements discussed earlier, constructs exceeding 0.7 are deemed acceptable for internal reliability. This evaluation ascertained whether the instrument produced consistent results. As reported in Appendix A, apart from the *agreeableness* trait, the constructs exceeded the recommended level of 0.7 for CR, thereby establishing acceptable reliability of the scales.

In terms of validating the measurement model, both the convergent validity and the discriminant validity were assessed. For convergent validity, the individual item loadings of the constructs and their associated items (as indicated in Appendix A) were inspected and secondly the average variance extracted (AVE) scores were computed. Factor loadings exceeding 0.5 were deemed acceptable while those below 0.5 were subsequently dropped from the model. Apart from the Heuristic processing variable, the AVE values of the constructs were above the recommended threshold of 0.5, thus indicating good convergent validity (Bagozzi & Yi, 1988). According to Fornell and Larcker (1981), if the AVE for a construct is below 0.5, and the composite reliability is higher than 0.6, then the convergent validity of the construct is still considered adequate.

To assess discriminant validity, the Fornell-Larcker criterion was performed (Fornell & Larcker, 1981) and presented in Table 7.15, in the form of a correlation matrix. This test is used to evaluate whether the square root of the AVE of each variable (on the diagonal) is greater than the correlation coefficients it shares with other variables in the same measurement model. The results show that the square root of all the AVE value were above 0.7, except for the Systematic processing variable, where the square root of the AVE equals 0.694. However, this construct maintains discriminant validity as the diagonal value is greater than the off-diagonal values. In accordance with the guidelines provided by Weston and Gore (2006), the

descriptive statistics, such as the mean and standard deviations, were included in the matrix. This allows for other researchers to duplicate the results and independently assess model fit.

Table 7.15: Descriptive statistics and correlations

Constructs	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12
1. Extraversion	3.519	0.721	0.708											
2. Agreeableness	3.978	0.715	0.156*	0.720										
3. Conscientiousness	3.691	0.854	0.321*	0.352*	0.833									
4. Neuroticism	2.661	0.701	-0.303*	-0.388*	-0.437	0.756								
5. Openness	3.610	0.277	0.410*	0.119*	0.338	-0.192*	0.951							
6. Heuristic processing	2.901	0.290	0.021	0.048	0.031*	0.002	0.093	0.760						
7. Systematic processing	3.080	0.565	0.361*	0.017	-0.013	0.027	0.187*	0.491*	0.694					
8. Habit	3.908	0.256	0.005*	0.117*	-0.139*	-0.130	0.029*	0.146*	0.223*	0.710				
9. Computer self-efficacy	3.394	0.490	0.229*	0.049	0.231*	-0.050	0.315*	0.068	0.029	-0.057	0.819			
10. Social norms	2.649	0.599	-0.007	-0.097	-0.100*	0.145*	0.038	0.073*	0.119*	0.308*	-0.040	0.800		
11. Perceived risk	2.093	0.160	0.015*	0.110	0.350	0.040*	-0.024	-0.104*	-0.093*	0.312*	0.100*	0.039	0.783	
12. Phishing susceptibility	0.651	0.053	0.170*	0.031*	-0.121*	0.033	0.122*	0.047	0.022	0.093	-0.143*	0.003*	0.051	1

Note: the square root of the AVEs is represented in bold, appearing down the diagonal

* Indicates significant correlation coefficients $p < .05$

7.5.1 Common method variance

As a self-reported questionnaire was the sole method used to collect data from the respondents in a single sitting, it is possible that this could bring inherent problems and limitations to the interpretation of the results of this study (Woszczyński & Whitman, 2004). For example, the questionnaire was lengthy as it consisted of many items (122 items) and as such required substantial cognitive effort to generate accurate responses, thereby increasing the risk that respondents might feel fatigued and consequently less willing to respond accurately towards the end of the survey. As a result, respondents might have a tendency to provide consistent, yet false, answers within the item scale measures and possibly across other scale measures (known as a consistency motif). All of these concerns are caused by the effect of common method variance (CMV) (Podsakoff & Organ, 1986). CMV is “attributable to the measurement method rather than to the constructs the measures represent” (Podsakoff et al., 2003). Accordingly, there is the prospect that the tested relationships among the constructs may be inaccurate as a result of the effect of CMV, which may potentially lead to incorrect conclusions concerning the reliability and validity of the item scales this study aims to measure.

In order to mitigate these concerns, Podsakoff et al. (2003) discuss several sources of bias that could potentially influence the validity of the research findings. Two main approaches can help overcome CMV – procedural and statistical approaches. The procedural, or preventative, approach (an *ex ante* technique) is the most preferred and is applied early in the research design stage. The statistical approach (an *ex post* technique) is conducted in the empirical stage to detect or possibly eliminate CMV (Chang et al., 2010; Podsakoff et al., 2003). Following the guidelines of Podsakoff et al. (2003), the following *procedural* strategies in this study were applied to mitigate CMV. First, an initial version of the survey was pilot-tested to establish whether the research instrument could be considered reliable, accurate and understandable. The instrument included synonyms in parenthesis (where necessary) to ensure questions were clear and unambiguous. For example, in the personality trait scale the item: “can be tense” (i.e. nervous, anxious)”. Respondents were instructed to provide feedback relating to any misinterpretations about what the questions expected of them. To encourage honest responses, the respondents were informed about the purpose of the study before they started the survey, that participation was voluntary, that there were no right or wrong answers, and that they could

withdraw from the survey at any time. Moreover, that the data was collected anonymously and no identifiable personal information was required from them.

In respect of statistical remedies (an ex-post technique), two statistical tests were performed during the empirical stage to ascertain whether CMV posed any concerns. First, the Harman's single-factor test was used to determine whether variance in the data could be largely attributed to a single factor (Podsakoff et al., 2003). If the total variance for a single factor is less than 50%, this test suggests CMV to be of no concern. The results show that the largest variance explained by a single factor in the model was 10.31%, indicating that none of the emergent factors could explain the majority of the covariance. Second, Bagozzi et al. (1991) suggest that CMV can have an effect on the discriminant validity of the constructs (see Section 6.10.3). In this regard, the procedure by Pavlou et al. (2007) was used to examine the correlation matrix (shown in Table 7.15) and to determine whether any of the correlations between any pair of constructs exceeded 0.9 – all the correlations were subsequently found to be below 0.9. Based on this examination, it was concluded that the measurement model was both reliable and valid as it satisfied the requirements from a convergent and discriminant perspective.

7.6 Structural model evaluation

The aim of the hypotheses of this study was to establish the extent to which the constructs influence susceptibility to phishing on SNSs. As the measurement model demonstrated sufficient construct validity and reliability, path modelling analysis was conducted. The goal of path analysis, and more generally of SEM, is to determine the extent of which the proposed model, which is a set of specified causal and non-causal relationships among the latent variables, accounts for the observed relationships among these variables. The observed relationships are usually the covariances, summarised in the sample covariance matrix (McDonald & Ho, 2002). The relationships can be tested using path analysis techniques to depict how one factor influences another. The structural model specifies the hypothesised relationships among latent variables representing the causal and consequent constructs of a theoretical proposition (Lowry & Gaskin, 2014).

Relying on *p*-values as a sole means to justify significance has received criticism (Zhu, 2016). Although significance tests are important, it is also recommended that effect sizes (f^2)

should be reported, as they offer insight into the magnitude of the actual size of an effect (Bowman, 2017) and therefore help researchers to assess the overall contribution of a research study (Sullivan & Feinn, 2012). The guidelines for assessing Cohen's f^2 are values of 0.02–0.14, 0.15–0.34, and 0.35 and above, which respectively represent small, medium and large effects of an exogenous latent variable on an endogenous latent variable (Cohen, 1988). Effect size values of less than 0.02 indicate that there is no effect.

7.6.1 Path analysis and hypotheses outcomes

Based on the results of the path coefficient analysis, presented in Table 7.16, apart from the Openness trait, the remaining Big Five traits of Extraversion, Agreeableness, Conscientiousness and Neuroticism show statistically significant relationships with habit. The results of the path analysis revealed the following outcomes:

- **Hypothesis 1a** proposed that the *Extraversion* trait will be positively related to performing a *Habit* on SNSs. Extraversion revealed a positive influence on habit ($\beta = 0.245$, $p < 0.05$, small effect) thus supporting H1a.
- **Hypothesis 1b** posited that the *Agreeableness* trait is positively related to performing a *Habit* on SNSs. Although statistically significant, agreeableness had a negative influence on habit ($\beta = -0.161$, $p < 0.05$, small effect), thus H1b is not supported.
- **Hypothesis 1c** argues that the *Conscientiousness* trait is negatively related to *Habit* on SNSs. Conscientiousness was also found to have a negative influence on habit ($\beta = -0.167$, $p < 0.05$, small effect), thereby supporting H1c.
- **Hypothesis 1d** stated that the *Neuroticism* trait will be negatively related to *Habit* on SNSs. The results show that neuroticism ($\beta = 0.227$, $p < 0.01$, small effect) positively influences habits. Although this relationship is statistically significant, its direction does not support the hypothesis, thus H1d is not supported.

- **Hypothesis 1e** posited that the *Openness* trait will be positively related to *Habit* on SNSs. However, the results show that openness is statistically not significant to habit ($\beta = 0.183, p > 0.10$, no effect). As such, H1e is not supported.
- **Hypothesis 2a** posited that *Habit* will have a negative influence on *Systematic processing*. As the results indeed show habit to have a negative influence on systematic processing ($\beta = -0.269, p < 0.001$, small effect), support is provided for H2a.
- **Hypothesis 2b** posited that *Habit* will have a positive effect on *Heuristic processing*. The results show that habit is positively related to heuristic processing ($\beta = 0.207, p < 0.01$, small effect), thus providing support for H2b.
- **Hypothesis 3** proposed that *Habit* is positively related to *phishing susceptibility*. The results show habit to have a positive effect on phishing susceptibility ($\beta = 0.209, p < 0.001$, small effect) thus supporting H3.
- **Hypothesis 4** posited that *Systematic processing* will be negatively related to *phishing susceptibility*. Systematic processing had a negative influence on phishing susceptibility ($\beta = -0.277, p < 0.001$, small effect) thus supporting H4.
- **Hypothesis 5** stated that *Heuristic processing* will be positively related to *phishing susceptibility*. The study indeed determined that heuristic processing had a positive effect on phishing susceptibility ($\beta = 0.172, p < 0.05$, small effect), thus supporting H5.
- **Hypothesis 6** proposed that *Social norms* will be positively related to *phishing susceptibility*. As social norms has a positive effect on phishing susceptibility ($\beta = 0.098, p < 0.05$, small effect) H6 is thus supported.
- **Hypothesis 7** put forward that *Computer self-efficacy* will be negatively related to *phishing susceptibility*. Computer self-efficacy was found to have a negative effect on phishing susceptibility ($\beta = -0.118, p < 0.01$, small effect) thus supporting H7.

- **Hypothesis 8** argued that *Perceived risk* is positively related to *phishing susceptibility*. However, perceived risk was found not to be statistically significant with regard to phishing susceptibility ($p > 0.10$) and, hence, this hypothesis was not supported.

Overall, the results of the path analysis support nine of the thirteen hypotheses proposed in the study. Table 7.16 reports on the analysis.

Table 7.16: Path analysis and hypotheses outcomes

Test	Path	β	S.E.	<i>t</i> -value	<i>P</i> -value	f^2	Outcome
H1a	Extraversion→Habit	0.245	0.093	2.624	0.010	0.064	Supported
H1b	Agreeableness→Habit	-0.161	0.072	-2.242	0.022	0.024	<i>Not supported</i>
H1c	Conscientiousness→Habit	-0.167	0.076	-2.202	0.030	0.045	Supported
H1d	Neuroticism→Habit	0.227	0.070	3.256	0.001	0.051	<i>Not supported</i>
H1e	Openness→Habit	0.183	0.165	1.109	0.125	0.000	<i>Not supported</i>
H2a	Habit→Systematic	-0.269	0.065	-4.083	0.000	0.078	Supported
H2b	Habit→Heuristic	0.207	0.062	3.339	0.001	0.046	Supported
H3	Habit→ Phishing Susceptibility	0.209	0.078	2.685	0.007	0.068	Supported
H4	Systematic→Phishing Susceptibility	-0.277	0.071	-3.916	0.000	0.072	Supported
H5	Heuristic→Phishing Susceptibility	0.172	0.038	4.514	0.020	0.109	Supported
H6	Social norms→Phishing Susceptibility	0.098	0.018	2.105	0.035	0.023	Supported
H7	Computer self-efficacy→Phishing Susceptibility	-0.118	0.039	-3.005	0.003	0.081	Supported
H8	Perceived risk→Phishing Susceptibility	0.152	0.149	1.020	0.109	0.004	<i>Not supported</i>

Following the path analysis and the outcomes of the hypotheses tests, presented in Table 7.16, the structural model was created and is depicted in Figure 7.1.

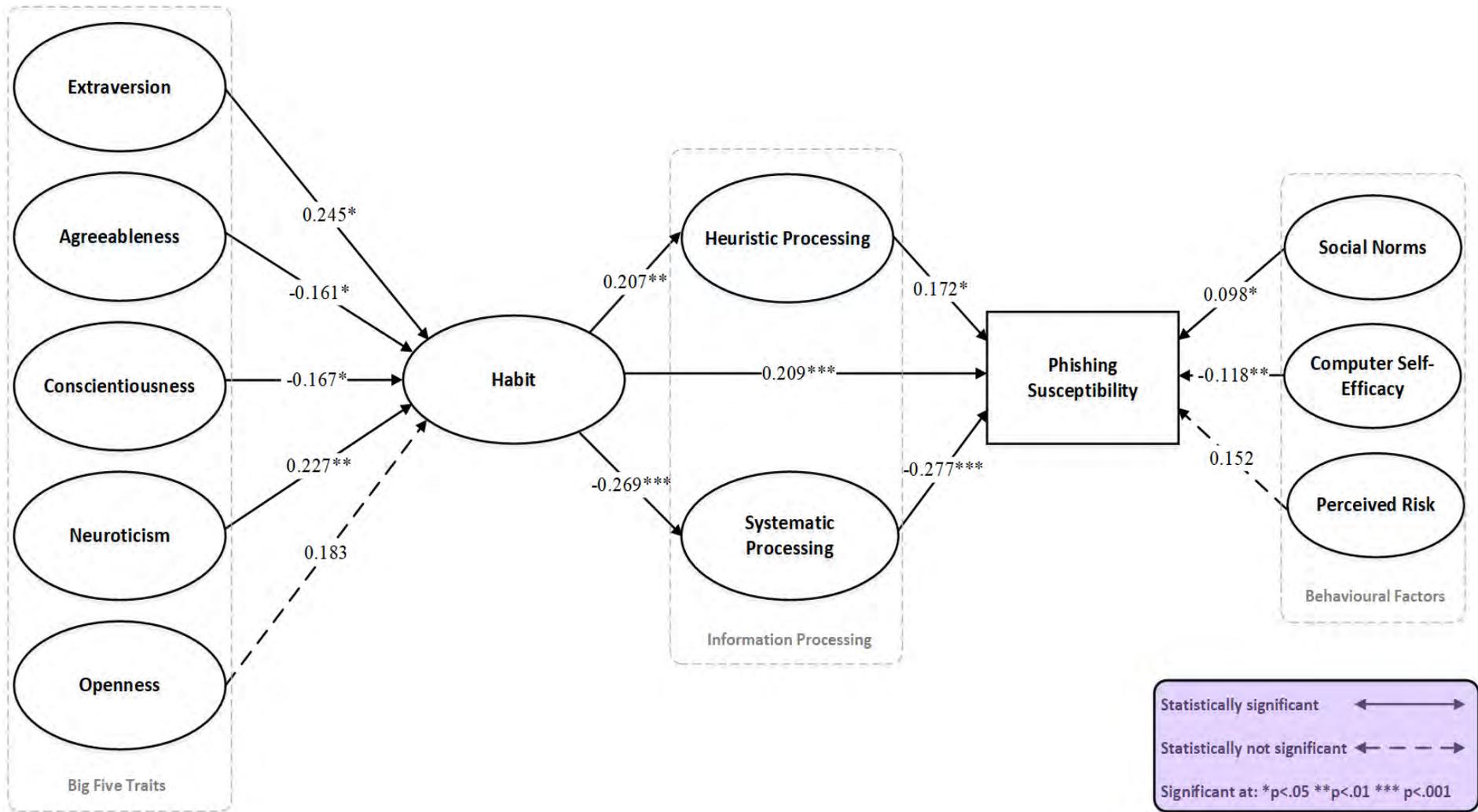


Figure 7.1: The Personality-Habit-Information Processing (PHIP) model

Figure 7.1, depicted as a path diagram, represents the theoretical model of this study, demonstrating the predictors of phishing susceptibility on SNSs in terms of personality traits, habits, information processing as well as other behavioural factors. The model also indicates the correlation coefficients (β) and the significance of the relationships (*) between the variables.

7.6.2 Evaluation of model fit

According to Weston and Gore (2006), researchers should evaluate fit in terms of the significance and strength of estimated parameters and how well the overall model fits the sample data. The measures that assess the compliance of the models with the data are known as fit indices or fit statistics (Civelek, 2018). In addition to the sample size debate in SEM, there has been considerable disagreement and debate over what constitutes acceptable threshold values for adjudicating acceptable model fit in SEM (Gefen et al., 2011). Barrett (2007) controversially advocates for an outright ban on approximate fit indexes and suggests that “the chi-square test is the ONLY statistical test for SEM models”. This view is supported by Kline (2016, p. 269), who recommends that researchers should at the very least report on the following fit indices: the model chi-square (χ^2) test with its degrees of freedom (*df*) and *p*-value, the root mean square error of approximation (RMSEA), the standardised root mean square residual (SRMR) and the comparative fit index (CFI). Several alternative fit indices (e.g. χ^2/df) exist but have been shown to be flawed and it is recommended that they not be relied upon (Gefen et al., 2011). As such, this study adopted the recommendations made by the aforementioned authors. Assessing overall fit in CB-SEM typically begins with the χ^2 test and its *df* (Gefen et al., 2011). The χ^2 test can be described as an accept–support test, with a “significant” result indicating that the model does not fit the sample data (Kline, 2016, p. 265). Thus, in contrast to the traditional significance measures, a non-significant χ^2 result at the $> .05$ threshold is desired to achieve a good fit between the variance and covariance matrix (Barrett, 2007). The χ^2 test in this study achieved an acceptable fit: $\chi^2 = 22.591$, *df* = 16 and *p* = 0.125. In terms of goodness-of-fit measures, RMSEA is an absolute fit index which evaluates the extent to which a hypothesised model differs from a perfect model. An RMSEA value of .00 indicates that the model exactly fits the data (Kline, 2016, p. 273). The SRMR is an approximate fit index that tests the difference between the observed correlation and the model implied correlation matrix. Similar to the RMSEA, an SRMR value of zero indicates perfect

fit (Hooper et al., 2008) and a value as high as .08 is deemed an acceptable fit (Hu & Bentler, 1999). CFI is an incremental fit index and compares the sample covariance matrix with the null model. In order to ensure that misspecified models are not accepted, a value exceeding 0.9 is required (Hu & Bentler, 1999). Table 7.17 summarises the fit index measures with their associated acceptable thresholds.

Table 7.17: Fit indices of the model

Fit Indices	Model Value	Acceptable standard
χ^2	22.591 (<i>df</i> =16; <i>p</i> = 0.125)	<i>p</i> > 0.05
RMSEA	0.044	< 0.08
SRMR	0.051	< 0.08
CFI	0.903	≥ 0.9

The values for all the fit indices satisfied the required criteria thus validating that the theoretical model for this study fits the data satisfactorily.

7.7 Summary

The aim of this chapter was to provide a detailed presentation of the results following the data analysis. Initially, a pilot study was conducted in order to ascertain whether the survey instrument was considered reliable. This was followed by data collection for the main study. A discussion on the sample size determination and motivation for the target sample was also provided. Following this, the results concerning the reliability and factorial validity and the necessary adjustments to the constructs were presented. SEM was identified as the appropriate technique in this study to analyse the structural relationships between multiple constructs. In this regard, a detailed discussion on the measurement model and structural model was provided. The chapter concluded with an overview of the path estimates of the model and the outcomes of the hypothesis tests. In addition, the results of the fit indices give support for the structural soundness of the model. The next chapter provides a detailed discussion of these results and how they contribute to addressing the research questions posited in earlier in the study.

8.1 Introduction

This study developed a model which can help identify behavioural characteristics that make users susceptible to phishing on SNSs. Thirteen hypotheses were proposed in Chapter 5 with the objective of investigating the influence of the Big Five personality traits and their effect on habits. Moreover, the extent to which habits influence information processing was examined as these ultimately determine the outcome of phishing susceptibility. The previous chapter reported several statistically significant findings and presented the theoretical model in the form of a path diagram. The aim of this chapter is to provide a discussion of the results and, in doing so, it provides further insights into how the results contribute to addressing the research problem.

8.2 The extent to which the Big Five personality traits influence habit

Chapter 3 highlighted that there is a lack of research in the information security domain that investigates the influence that the Big Five personality traits have on habit. Research in this area has been mainly conducted from a health perspective (Gardner, 2015). For example, Intiful et al. (2019) investigated the relationship between personality traits and dietary habits. Given the absence of research investigating the influence of the Big Five personality traits on habit, this study utilised literature stemming from Facebook usage. Literature has shown that an individual's personality trait has correlations to very specific Facebook activities (Sumner et al., 2011). In this regard, much of the literature on Facebook usage, which included personality traits, discussed concerns pertaining to Facebook addiction, dependency or intensity. The literature in this area highlights that excessive frequency and performance of a

behaviour on SNSs, as antecedents of habit, can potentially lead to a form of addiction or dependency. In addition, this study leveraged on literature, drawing on Maslow's hierarchy of needs (Maslow, 1943) and its relationship to the Big Five traits, in which it has been argued that the quest to satisfy certain needs on SNSs could motivate and foster habitual behaviour. Maslow's hierarchy of needs is a motivational theory in psychology comprising a five-tier hierarchical model of human needs, often depicted as a pyramid. From the bottom of the hierarchy upwards, the needs include physiological (food, water, rest and clothing), safety (job security, protection from danger), love and belonging (intimate relationships and friendship), esteem, and self-actualisation (Maslow, 1943). This study posits that SNSs are one of the sources that could help satisfy certain needs and desires. The high engagement that results from utilising the features on these sites may satisfy these needs, thus fostering the habit. This is possible, as in Section 2.5.1 satisfaction is pointed as being an antecedent of habit. Moreover, this may also bear some resemblance to the uses and gratifications theory (UGT), which asserts that the use of a particular media is goal-directed to satisfy certain wants and needs (Katz et al., 1973). Moreover, Kircaburun and Griffiths (2018) point out in their study that personality traits may influence how individuals experience gratification.

8.2.1 Extraversion and its influence on Habit

As *extraverts* are typically positively disposed, social and enthusiastic, it was anticipated that this trait would be more likely to develop habits from the frequent engagement with other users on SNSs. The results ($\beta = 0.245$, $p < 0.05$, small effect) support this view and align with the findings of Van der Schyff, Flowerday, Kruger, et al., (2020), who found that people with the extraversion trait do develop intensive use of Facebook. Wehrli (2008) and Correa et al. (2010) likewise found extraversion to be positively related to the use of SNSs. Grandiose narcissism is shown to correlate with high levels of extraversion (Zajenkowski & Fronczyk, 2020) and Facebook specifically gratifies the narcissistic individual's need to engage in self-promoting and superficial behaviour (Ryan & Xenos, 2011). This is expected, as extraversion is strongly correlated with Maslow's need for esteem (Montag et al., 2020). It can accordingly be argued that individuals with the extraversion trait are prone to spending more time on SNSs, posting photos of themselves, interacting with their friends, and regularly checking for notifications. More concerning is the fact that intensive usage can lead to SNS addiction (Balcerowska et al., 2020). As a result of these behaviours, this study argues that users with this trait could form

habitual behaviour such as clicking on malicious post or links on SNSs, which could put them at risk of phishing.

8.2.2 Agreeableness and its influence on Habit

In view of the fact that *agreeable* individuals are described as trusting, friendly and respecting of others' feelings and beliefs, and that this trait is strongly correlated with Maslow's need for belongingness (Montag et al., 2020), it was expected that users with this trait would likely develop habits, with high engagement, to satisfy this need within the SNS community. Contrary to the study's expectations, however, the results revealed that agreeableness was negatively related to habits ($\beta = -0.161$, $p < 0.05$, small effect). This result is similar to that of Ross et al. (2009) and Amichai-Hamburger and Vinitzky (2010), who found no relationship between this trait and Facebook usage. Furthermore, Rajesh and Rangaiah, (2020) found that agreeableness is not related to Facebook addiction. These findings suggest that this trait might be less prone to forming habitual behaviours on SNSs. Rolland (2002) found agreeableness to be sensitive to the cultural background of the individual, which may explain why the study results found agreeableness to have a negative correlation to habit. Furthermore, negative experiences may have an adverse effect on agreeableness as well as on phishing susceptibility (Parrish Jr et al., 2009). Thus, it is possible that the background or current circumstances of the respondents may have had an influence on their disposition at the time of their participation in the survey.

8.2.3 Conscientiousness and its influence on Habit

The *Conscientious* trait is associated with individuals who are self-disciplined, goal-oriented and, more importantly, thorough and cautious. Conscientious users are generally risk adverse and less prone to SNS addiction (Balcerowska et al., 2020; Tang et al., 2016). Sumner et al. (2011) found those with the conscientious trait to be less likely to join Facebook groups, were negatively correlated to the use of profanity and more likely to use correct spelling and grammar. As such, these users could probably identify phishing cues such as poor spelling. Accordingly, this study expected users with this trait to be less likely to develop a habit. In support of this result, the conscientiousness trait was found to be negatively correlated to habit

($\beta = -0.167$, $p < 0.05$, small effect), thus supporting the notion that conscientious users are less likely to be vulnerable to phishing attacks.

8.2.4 Neuroticism and its influence on Habit

As individuals who are *neurotic* are typically depressed, lack self-esteem and are anxious, it was expected that they would have a tendency to resist adopting technologies (Özbek et al., 2014). Although Halevi et al. (2013) found this trait most at risk of email-based phishing, as the current study context involved SNSs and prior literature has shown this trait to be less engaged with technology, it was expected that neurotic individuals would be less likely to be habitually engaged on SNSs. However, the results showed this not to be the case ($\beta = 0.227$, $p < 0.01$, small effect). This might indicate that while the neurotic trait may be risk averse with regard to privacy concerns, such as sharing information, this does not necessarily indicate that they will have no interaction on SNSs. This explanation would also support Mehdizadeh's (2010) finding that users with low self-esteem are more engaged and active in promoting themselves on Facebook. This is because the neurotic trait relates to the narcissism trait in terms of self-promotional content on Facebook (Mehdizadeh, 2010). Moreover, Hughes et al. (2012) found neuroticism to be positively associated with Facebook usage. The latter also supports the findings of Wehrli (2008) who, contrary to expectations, also found neuroticism to be associated with more participation on StudiVZ, a German-based SNS. Ross et al. (2009) found that individuals higher in neuroticism were less willing to share personal information on Facebook, and those low in neuroticism preferred posting photos on Facebook timeline. However, Sumner et al. (2011) found that the more neurotic a person is, the more photo albums they have on Facebook. These similar findings may indicate that owing to emotional instability, users with this trait will be motivated to use SNSs to compensate for their lack of interpersonal skills (Moore & McElroy, 2012), as well as, from the perspective of Maslow's needs, to avoid loneliness by garnering a sense of belonging on SNSs (Butt & Phillips, 2008). More concerning, loneliness was found to be positively related to Facebook addiction (Rajesh & Rangaiah, 2020). Like extraverts, this trait too could potentially lead to developing addictions as a result of excessive habit performance.

8.2.5 Openness and its influence on Habit

Openness is associated with individuals who have an appreciation for new experiences and are respectful of others' differences in ideas and beliefs (Uebelacker & Quiel, 2014). As this trait is associated with being sociable on Facebook (Ross et al., 2009), the study expected high engagement leading to the development of habits. Sumner et al. (2011) found the openness trait to willingly share personal information about themselves, including lengthy biographies, and their hobbies and interests, and such people are inclined to write wall posts and comment on other's posts. Moreover, individuals high in openness have been found more likely to commit security policy violations (Johnston et al., 2016). Although this trait showed a positive association to habit, the study hypothesis was not supported, as the results exhibited a statistically non-significant result. In terms of Maslow's needs, openness was not significantly correlated with the physiological and the safety and security needs (Montag et al., 2020). Given that the context of this study involves protection against phishing attacks, this might explain why this study also exhibited a non-significant result ($\beta = 0.183$, $p > 0.10$, no effect). In the context of SNSs, Van der Schyff, Flowerday, Kruger et al. (2020) also found the openness trait as having no significant relationship with Facebook usage.

8.3 The extent to which Habit influences Information processing

As mentioned in Section 4.3, the HSM is a dual-process theory. According to Vishwanath (2015b), information processing is a parallel process that, unlike automatic responses, causes some amount of conscious reflection. As such, Vishwanath (2015b) argues that if habits and heuristic processing lead to individuals clicking on a phishing link, it is possible that the two are related and part of the same process. This view may support the description by Gardner (2015), who states that dual-process models propose two parallel processing systems which are ultimately two pathways of behaviour. Gardner (2015) states that of these, habit operates in the impulsive pathway which prompts rapid and efficient behavioural response but reduces foresight. The description by Gardner (2015) indicates that habit is a shallow form of processing with the focus on efficiency, thus fitting the characteristics of heuristic processing.

This study investigated the relationship between habit and information processing and moreover the direct effect of habit on phishing susceptibility. The results show habit to have a negative influence on systematic processing ($\beta = -0.269$, $p < 0.001$, small effect). This indicates that individuals presented with phishing messages and acting on habitual behaviour, will not make an effort to use critical thinking to evaluate the quality and authenticity of the content and its arguments. On the other hand, habit has a positive influence on heuristic processing ($\beta = 0.207$, $p < 0.01$, small effect) which indicates that individuals evaluating phishing messages, and who possess a habit, will react quickly to the message and rely on superficial cues to reach a decision. On this aspect, the results also showed that habit is positively related to phishing susceptibility ($\beta = 0.209$, $p < 0.001$, small effect). This supports Vishwanath's (2015a) suggestion of a possible joint influence of habit on cognitive processes and essentially answers Vishwanath's concern that habit and heuristic processing could essentially be measuring the same process; that is, if one is performing a behaviour out of habit, one is applying heuristic processing at that particular moment. In both cases, the results show that an individual is likely to be susceptible to phishing. Coincidentally, the coefficient values (see Figure 7.1) for these two relationships in this study show strong similarity as they have almost identical values.

8.4 The extent to which Information processing influences phishing susceptibility

The presence of visual elements in persuasive stimuli has been found to be most effective, as it has a stronger effect on the user (Ferreira & Teles, 2019). It is thus not surprising that this study found heuristic processing to have a positive influence on phishing susceptibility ($\beta = 0.172$, $p < 0.05$, small effect). Stimuli on Facebook predominantly consist of graphics, with less textual information, thus not allowing users to perform a thorough evaluation, assuming they are motivated to do such. This also supports prior literature which states that persuasive stimuli encourage heuristic processing. Moreover, this study found systematic processing to have a negative influence on phishing susceptibility on SNSs ($\beta = -0.277$, $p < 0.001$, small effect) which supports prior literature that shows that users who apply systematic processing to persuasive messages such as in phishing will be less susceptible to phishing, as they will apply more cognitive effort when evaluating the message (Vishwanath et al., 2018).

The univariate analysis indicated that 40% of the respondents in this study fell victim to the phishing email, while only 20.47% indicated that they were suspicious of it. This is similar to the results of Oliveira et al. (2017), who found more than 40% of internet users were highly susceptible to spear phishing email attacks. This study reported that 7.91% indicated that they would reply to the email. This remains a risk, as phishers could use this as an opportunity to establish trust with the victim and possibly use other methods to persuade the user into performing actions to gather more information. For example, an email response could contain the victim's email signature, so the phisher could phone the victim and use other persuasion techniques to gather personal information from them. Both the univariate and multivariate analysis indicate concerns for organisations in this regard, as all it takes is for one user to fall victim to a phishing attack, which could compromise the entire organisation.

8.5 The influence of dispositional factors on phishing susceptibility

According to Russell (2014) behaviour is considered by psychologists to be a product of both the *situation* (e.g. cultural influences) and the *subject* (internal characteristics of the person). Situationism holds that one's behaviour and actions are influenced by the environment and the surroundings of the subject. In contrast, dispositionism posits that our behaviour is determined by internal factors such as personality traits. Accordingly, dispositional factors play an important role in determining why it is important to understand how both the environment and an individual's internal biases affect their behaviour. As mentioned in Chapter 5, Workman (2008) found that individuals who are high in normative commitment feel obligated to reciprocate the SE requests incorporated in phishing attacks, such as receiving free gifts or divulging confidential company and user information. In the study by Algarni, Xu, Chan, and Tian (2014), it was found that social network users may unintentionally participate in distributing SE messages on SNSs as a result of socio-psychological factors. Oblivious to the fact that the messages were malicious, participants were "sharing", "liking", and "retweeting" messages for reasons such as to help others in need and to show that they were accommodating of others' religious beliefs. Posts included job offers and prizes or deals that required victims to provide personal information. This behaviour is in line with the "reciprocity" principle. In the study by Algarni, Xu, and Chan, (2014), one of the participants reported: "Some of the users in my friend list always like and write good comments on my photos or posts, and I usually do the same for them to keep them around." The aforementioned reasons show that users may be inclined to please others on social networks and, furthermore, through social

interaction may copy behaviour that they see in others. This response is in line with the result found in the univariate analysis in this study, where most respondents agreed that they shared a friend's post because it is "the friendly thing to do".

The results of this study therefore support the notion that *social norms* positively influence phishing susceptibility ($\beta = 0.098$, $p < 0.05$, small effect). As such, users who feel obligated to share posts, or have a tendency to comply with requests from others on SNSs, are more at risk for phishing attacks on this platform. This is not surprising, as a study by Mattke et al. (2020), which investigated user behaviour on Facebook, found that intentions to click on sponsored content is influenced by observing a "like" from a close friend.

Pertaining to *computer self-efficacy*, the univariate analysis revealed that the respondents considered themselves above average in performing most general computer-related tasks. This was expected as the respondents are students in their final year of university and frequently use computers to create and submit assignments, and use the internet to conduct research for assignments and use SNSs. However, the analysis also showed that the respondents considered themselves less confident (below average) in their ability to handle tasks that related directly and indirectly to protecting themselves against phishing threats. For example, understanding the apps associated with certain file extensions has been shown to be important in protecting against security threats. Microsoft Office and Adobe Acrobat files, for example, represented 94% of malware attachments in emails in 2020 (Bhardwaj et al., 2020).

The results of this study support the hypothesis that individuals with *computer self-efficacy* will be less at risk of phishing, as computer self-efficacy was found to have a negative effect on susceptibility to phishing ($\beta = -0.118$, $p < 0.01$, small effect). This indicates that individuals who have the competency to perform certain computer-related tasks and other validation checks in determining the authenticity of phishing messages, are less susceptible to phishing. This supports other studies, as Iuga et al. (2016) found years of computer usage have a statistically significant impact on the detection rate of phishing emails and Alseadoon et al. (2015) found that email experience increases victims' suspicion of phishing emails. Wright and Marett (2010) are of the view that internet experience and increased security knowledge resulting from training should make users less susceptible to phishing. Moreover, Pattinson et al. (2012) maintain that the more familiar people are with computers, the better they manage phishing. Algarni et al. (2015) and Albladi and Weir (2018) found that computer and security

knowledge decrease a user's susceptibility to phishing attacks. These results contradict those of Dhamija et al. (2006), who found in their study that neither previous experience nor hours of computer usage had any effect on whether users distinguished legitimate and spoofed websites or not. Similarly, in a study by Moody et al. (2017), it was found that internet experience presented contradictory findings. They found that the more individuals used the internet, the more likely they were to click on links in unsolicited emails. Moreover, Moody et al. found that the more time spent on the internet, the more knowledge that would be acquired on internet threats and thus the better equipped such persons would be to identify those potential threats and attacks.

Phishing on SNSs is shown to provoke different risk perceptions compared to email-based phishing attacks (Alqarni et al., 2016). Pertaining to risk perception, Moody et al. (2017) found unexpectedly that individuals who believe clicking on links in emails is risky behaviour did not transfer this to their behaviour, as they were also prone to clicking on links. This might explain why no statistically significant relationship with *perceived risk* was found in this study ($p > 0.10$).

8.6 Summary

Prior literature has shown that individuals with certain personality traits, apart from conscientiousness, have vulnerabilities that can be exploited by deception. Empirical research has also revealed that the extraversion and agreeableness traits share very similar characteristics, particularly in terms of Facebook engagement. The results of this study show that extraversion and neuroticism, although they would appear to be polar opposites in terms of their trait descriptions, may be motivated by very similar goals or needs (Montag et al., 2020) which could potentially reinforce the habit to click or share social media posts. For example, extraverts have a natural social need to be involved and engaged with other Facebook users. On the other hand, neurotic users might develop habits because of their loneliness in the physical world, and thus may be more motivated to compensate for this need on SNSs. Moreover, SNSs could help them overcome their lack of self-esteem, finding SNSs a safer or more private environment in which to interact with others without awkward face-to-face interactions (Kuss & Griffiths, 2011). Accordingly, loneliness and a quest for belonging, as related to literature on Facebook usage, might tie these two traits together in terms of their

relation to phishing susceptibility and habits (Amichai-Hamburger & Ben-Artzi, 2003). This need for belonging can also drive certain behaviours such as users sharing their own posts and receiving gratification from these from others in the form of “likes” in regard to what they posted. As such, some scholars have investigated risky behaviour by considering both the Big Five traits and narcissism (Sudzina & Pavlicek, 2017).

Overall, the results on the relationship between personality traits and habit closely resemble those of Tang et al. (2016), who found agreeableness and conscientiousness to be negatively associated with Facebook addiction. This study also found these same traits to be negatively related to increased Facebook usage (Andreassen et al., 2013; Hussain et al., 2019; Ryan & Xenos, 2011). Ideally, personality characteristics that support reserved behaviour, low impulsivity and distrust are less at risk of phishing (Mayhorn et al., 2015). However, as context may influence the way people react (e.g. SNSs, emails, text messages); an individual who may be considered a “low risk” personality profile may be just as likely to fall victim to phishing when particular persuasion principles are used or the individual is put under pressure. For example, Canham et al. (2019) state that the characteristics of the conscientiousness trait can also be leveraged by phishers. According to Canham et al. (2019), one study to which they refer, deliberately exploited this trait by sending the target an attachment that requested them to correct an error on a timesheet. The current study’s findings also lend support to the heuristic-systematic processing model (HSM) and phishing susceptibility, as this study showed that users who apply heuristic processing will be vulnerable to phishing. The next chapter formally concludes the study.

9.1 Introduction

The previous four chapters have described the development of a model which identifies factors that make a user vulnerable to phishing on SNSs. The aim of this chapter is to formally conclude the study by summarising the main findings and discussing the way in which the research questions proposed in Chapter 1 have been accomplished. This chapter begins with a summary of each of the chapters while also highlighting gaps and concerns found following the review of the literature. The chapter also acknowledges the limitations of the study and in so doing reveals opportunities for future research. Importantly, the chapter also states the contributions of this study and lists publications emanating from the study.

9.2 Chapter summaries

Chapter 1 provided a complete overview of the study. The context of the study was set by first highlighting the worldwide popularity of social network sites (SNSs) and how these virtual communities foster a culture of information sharing and trust between entities on these sites. Importantly, there is a darker side, with the platform being used to propagate misinformation, leading to some posts being malicious and harmful.

The chapter continued by providing a background on phishing and with a discussion on the current measures and approaches used to mitigate this threat. In this regard, mitigation procedures used by organisations mainly include technological controls, enforcing security policies and conducting SETA programmes. From a scholarly perspective, researchers have focused on phishing detection by developing anti-phishing controls (e.g. browser plug-ins) and

machine-learning techniques for the detection and prediction of phishing websites. Moreover, researchers have investigated the effectiveness of training interventions and how such interventions influence users' responses to phishing. As all of these approaches have been shown to have their limitations, attention was drawn to adopting and extending existing theories and models with a focus on understanding human behaviour. It was pointed out that because not all people behave in the same way when faced with phishing, scholars have turned their attention to exploring individual differences (i.e. dispositional factors) to help characterise user behaviour and narrow down the problem. The chapter then described the problem and defined the study objectives and the associated research questions, as well as the methodological approach for the study.

The literature review spanned three chapters (Chapters 2–4) which established the theoretical foundation for the study. Overall, the literature review confirmed that many of the vulnerabilities associated with phishing stem from behavioural factors. This correlates with the literature that holds that “humans are the weakest link” in protecting against most security threat agents. The effectiveness of phishing lies in the use of social engineering (SE) techniques, particularly the use of persuasion, which aim to exploit certain behavioural tendencies in users. As such, **Chapter 2** appropriately provided a literature review on SE and its approaches. In doing so, it revealed several main findings. As SE attack methods are complex, constantly evolving and interrelated, researchers have created taxonomies, models and frameworks to identify appropriate countermeasures to mitigate phishing. However, despite this, disparities continue to exist between the taxonomies as a result of the original problem they attempted to address – the evolving nature of phishing attacks.

The chapter then focused on discussing Cialdini's persuasion principles and showed, in the form of real-world stimuli, how each of these principles can be executed on Facebook. This was necessary in order to identify the techniques and type(s) of phishing attacks, thereby addressing the *first research sub-question* of the study. Phishers can use a combination of persuasion principles in a single message in order to increase deception. In such cases, identifying which specific persuasion principle is the most effective on a user is almost impossible. According to Mansfield-Devine (2018), phishing exploits curiosity, fear or simple lapses in attention. It was noted that Cialdini's persuasion strategies do not take into account fear and curiosity and the influence these can have on users – especially social media users.

The chapter also introduced dispositional factors (internal factors) such as habit, perceived risk, computer self-efficacy, and social norms, as these variables can also potentially influence users' responses to phishing. Of these, habit is an important behavioural aspect in this study as users could be highly engaged in performing behaviours such as liking, sharing and clicking on links on SNSs. Studies suggest that more than half of all media behaviours are habitual (Wood et al., 2002). By definition, because habits are characterised as “mental scripts” and “automatic” behaviours, it is possible that this variable is related to information processing, specifically heuristic processing. The gratification that may result from frequently performing the aforementioned activities, compounded by past behavioural frequency, can reinforce the *automaticity* of the behaviour. As a result, this could also influence users to overlook suspicious phishing messages on SNSs. Given the lack of literature on habit in the context of both phishing and SNSs, this inspired the exploration in this study.

Computer self-efficacy is described as a user's judgement of their capability or competency to use computer-related technology. This study took the position that users who are competent in using computers will be in a better position to recognise and protect themselves against phishing attacks. As social media users may perceive themselves as being pressured or obligated by their friends into performing certain behaviours such as sharing, responding or liking posts, this study considered social norms to be another influential variable that has an effect on user behaviour. Because the Facebook platform inspires the trust compared to other popular SNSs (Fogel & Nehmad, 2009), users may not perceive threats on this platform, thus supporting the need to consider perceived risk as another influential variable. The aforementioned factors of computer self-efficacy, perceived risk and social norms are variables included in the theoretical model by Davis and Tuttle (2013). Similarly, Bashir and Madhavaiah (2015) extended the TAM and examined the influence of self-efficacy, social influence and perceived risk as factors that can affect internet banking adoption.

Chapter 3 introduced personality traits with a specific focus on the widely popular five-factor model (FFM) to categorise people's anticipated behaviour in association with each trait. The aim of the chapter was to determine the personality traits that are most vulnerable to phishing. By doing so, this addressed the *second research sub-question* for the study. Recent trends and developments in the phishing domain, involving the use of personality traits, were presented. The review showed that personality traits overlap across many focus areas; however, it became apparent that the literature has focused mainly on the influence of

persuasion principles and gender on personality traits. Nevertheless, despite the rapid growth and utilisation of personality traits in phishing research, the literature review revealed contradictory findings in terms of which traits are more or less susceptible to phishing. This may be explained by other behavioural motivations and influences that may alter the behaviour of users who have these traits. As SNSs are platforms that typically encourage users to be open by sharing their experiences on a daily basis, most studies on personality traits in these contexts have focused on privacy-related issues. Importantly, the literature review also revealed a lack of studies that have investigated the influence of the Big Five traits on habit (Vishwanath, 2015b). This lack is also expressed by Wood (2017), who states that “habit is largely missing from modern social and personality psychology”.

Various studies have shown that most SETA campaigns are inadequate for sustaining behavioural change (Jensen et al., 2017). Accordingly, other influential variables were considered which could explain why users, intentionally or unintentionally, ignore certain cues and characteristics which can help them identify phishing. As such, **Chapter 4** discussed the psychological frameworks in the information security discipline commonly used to help understand, explain or predict human behaviour under different situations and in different contexts. This was followed by introducing Heuristic-Systematic Model (HSM) of information processing which served as the underlying theory for the current study. The HSM proposes that individuals can process persuasive information in two modes, either heuristically or systematically, when determining the validity of a message. If users consider determining the validity of a phishing message on an SNS to be time-consuming, difficult or unimportant, this may result in their using heuristic processing. Using an example, a comparison of an email and a Facebook post showed the characteristics of the process that a user would typically use in identifying or evaluating the authenticity of a message.

The chapter further showed that it is difficult to distinguish a Facebook phishing post from a legitimate one even when using systematic processing. The literature review indicated that the HSM has been applied to examine how users respond to certain characteristics in phishing messages. Moreover, prior literature showed that studies that utilised the HSM extended it by incorporating variables from other popular theoretical models in order to create new models and to examine their effects on information processing. For example, Vishwanath et al. (2018) included email habits in their model, but did not examine their direct effect on information processing. This creates an opportunity for further investigation, considering that both habit

and heuristic processing have been described as a form of “automatic” behaviour. For example, most users open and read their text messages reflexively, and thus do not expect to receive phishing or malicious messages on their smartphones (Volkman, 2019a).

Chapter 5 provided a discussion on the behavioural influence of personality traits, habits and information processing and their influence on phishing susceptibility. Moreover, dispositional factors including social norms, perceived risk and computer self-efficacy and their influence on phishing susceptibility provided additional support for the study hypotheses. This was followed with a formal presentation of the hypotheses which attempt to explain the relationships between the constructs that subsequently translated to the proposed research model for the study.

Chapter 6 provided a detailed discussion of the research design and methodological approach used in this study. The study adopted a post-positivist philosophical paradigm and justified this choice of paradigm based on the study’s approach to theory development. This was followed by a discussion on the literature review process. A discussion on the research strategy was provided, in particular on how the sampling strategy was formulated. The variables and their associated measures adopted for this study were discussed. Importantly, a discussion on the statistical analysis approaches and methods used to ensure validity and reliability of the constructs were discussed. The chapter concluded with a discussion on how the collection of primary data adhered to the ethical guidelines and procedures.

Chapter 7 presented the survey results of both the pilot and the main study. The chapter focused on reporting the results of multivariate analysis which employed a covariance-based structural equation modelling (CB-SEM) approach that resulted in the development of both the measurement and the structural model. The chapter then reported on the outcomes of the path analysis, specifically the correlation coefficients, effect sizes, t-values and statistical significance. The study subsequently showed support for nine of the thirteen hypotheses proposed in the study. This was followed by a presentation of the structural model depicting the hypothesised associations between each of the constructs. The chapter concluded by evaluating the fit of the model, using several fit indices, in order to provide evidence that the model was developed with rigour and is structurally sound.

Chapter 8 focused mainly on a discussion of the relationships found in the multivariate results presented in Chapter 7. The proposed hypothesised relationships between the constructs served as the basis for organising the discussion in the chapter. This discussion provided evidence to support the outcomes of each of the hypothesised relationships, irrespective of the outcome of these relationships.

9.3 Research questions revisited

Chapter 1 revealed that because phishing has for several decades remained one of the most pertinent threats facing organisations today, it is apparent that the current methods used to address phishing are inadequate. Social network sites (SNSs), which are popular the world over and widely accepted as a fun and useful tool, may be leveraged by phishers to exploit the behavioural tendencies of users on these sites. Inadvertently, users are their own worst enemies as their personality traits and frequent engagement on SNSs may result in them developing certain behaviours or reinforcing existing behaviours such that they evolve into habits, thus resulting in unintended risks for themselves, their friends and, potentially, their organisations. Owing to individual characteristics, such as personality traits, each user has their own set of preferences and personal motivations and, consequently, this will have an influence on how they react to various messages on SNSs. This, coupled with habit, can influence a user's judgement to overlook certain malicious messages on SNSs or dismiss them with insufficient scrutiny. Hence, the objective of this study was to develop a theoretical model that could help to identify users who are susceptible to phishing on SNSs. In "identifying" such users, this study centred on dispositional factors relating to personality traits and habit and how they influence information processing. To achieve this, seven research sub-questions (see Section 1.5) had to be answered, most of which focused on the extent to which the model constructs of the study influence behaviour. Collectively, these subsequently answered the following main research question:

Main question. *To what extent do certain dispositional factors, such as personality traits, habit and information processing, influence a user's response to phishing messages on social networking sites?*

The following section provides an overview as to the way in which the sub-questions were answered:

- **Sub-question one.** *What type of messages are social media users exposed to which can put them at risk of phishing attacks?*

The first sub-question was answered by means of a literature investigation which aimed to identify common phishing messages or stimuli employed on Facebook. Examples, in the form of screenshots found on Facebook were reported in Chapter 2. The stimuli were categorised according to Cialdini's six principles of persuasion, as these have been shown to be effective in influencing users on SNSs. Identifying the type of stimuli also had an influence on the information processing construct (and other subsequent sub-research questions) as these types of stimuli were used in this study to evaluate information processing.

- **Sub-question two.** *Users with which personality traits are more susceptible to phishing attacks than others?*

The second sub-question was also answered by means of a literature investigation into the Big Five personality traits and their influence on how users respond to phishing. In an attempt to determine which trait is more susceptible to phishing, a significant proportion of the literature investigated how such users with certain Big Five traits will respond to each of the persuasion principles (those of Cialdini). The literature review confirmed that personality traits are a significant predictor of susceptibility to phishing, although with contradictory findings across numerous studies. The contradictions in themselves have spurred scholars to research more on this aspect in order to confirm or reject others findings. Nonetheless, some patterns of risky traits emerged and the description of each trait helped bring to light the traits identified for this study that could be most at risk.

- **Sub-question three.** *To what extent do personality traits influence habit when users are presented with phishing messages on social networking sites?*

This sub-question was answered by the study's empirical investigation and corresponds to hypothesis 1a to e. More specifically, the following was established:

- *Extraversion* has a positive influence on *Habit* ($\beta = 0.245^*$, $p < 0.05$) with a small effect on the explanatory power.
- *Agreeableness* has a negative influence on *Habit* ($\beta = -0.161^*$, $p < 0.05$) with a small effect on the explanatory power.
- *Conscientiousness* has a negative influence on *Habit* ($\beta = -0.167^*$, $p < 0.05$) with a small effect on the explanatory power.
- *Neuroticism* has a positive influence on *Habit* ($\beta = 0.227^{**}$, $p < 0.01$) with a small effect on the explanatory power.
- *Openness* is found to be statistically non-significant on *Habit* ($\beta = 0.183$, $p > 0.10$) with no effect on the explanatory power.
- **Sub-question four.** *To what extent does habit influence information processing when users are presented with phishing messages on social networking sites?*

This sub-question was answered by the empirical investigation of the study and corresponds to hypothesis 2a and 2b. Accordingly, it was established:

- *Habit* has a negative influence on *systematic processing* ($\beta = -0.269^{***}$, $p < 0.001$) with a small effect on the explanatory power.
- *Habit* has a positive influence on *heuristic processing* ($\beta = 0.207^{**}$, $p < 0.01$) with a small effect on the explanatory power.
- **Sub-question five.** *How does habit influence susceptibility to phishing?*

This sub-question was answered by the empirical investigation of the study and corresponds to hypothesis 3. Subsequently, it was established:

- *Habit* has a positive influence on *phishing susceptibility* ($\beta = 0.209^{***}$, $p < 0.001$), exhibiting a small effect on the explanatory power.

- **Sub-question six.** *How does information processing influence phishing susceptibility when users are presented with phishing messages on social networking sites?*

This sub-question was answered by a literature investigation as presented in Chapter 4 and by the empirical investigation of the study and corresponds to hypotheses 4 and 5. Accordingly, the following was established:

- *Systematic processing* has a negative influence on *phishing susceptibility* ($\beta = -0.277^{***}$, $p < 0.001$), exhibiting a small effect on the explanatory power.
- *Heuristic processing* has a positive influence on *phishing susceptibility* ($\beta = 0.172^*$, $p < 0.05$) and a small effect on the explanatory power.
- **Sub-question seven.** *To what extent do dispositional factors of social norms, computer self-efficacy and perceived risk influence phishing susceptibility?*

This sub-question was answered by the empirical investigation and relates to evaluating hypotheses 6 to 8 of the study, specifically:

- *Social norms* have a positive influence on *phishing susceptibility* ($\beta = 0.098^*$, $p < 0.05$) and a small effect on the explanatory power (addresses hypothesis 6).
- *Computer self-efficacy* has a negative influence on *phishing susceptibility* ($\beta = -0.118^{**}$, $p < 0.01$) and a small effect on the explanatory power (addresses hypothesis 7).
- For hypothesis 8, *perceived risk* was found to have a statistically non-significant effect on *phishing susceptibility* ($p > 0.10$).

9.4 Contributions of the study

Using SEM, this study developed a theoretical model with the aim of identifying users who are susceptible to phishing on SNSs. This study makes several contributions.

9.4.1 Contribution to theory

This is the first empirical study, in one comprehensive model, investigating the relationship between the Big Five personality traits and habit and its effect on information processing, which can influence susceptibility to phishing on SNSs. In addition, the model included other behavioural factors such as social norms, perceived risk and computer self-efficacy. Much like all theoretical models, including the TPB and the TAM, the study allows other researchers to expand on these models or theories to include other variables such as gender, culture and the like, which could potentially offer further insights into phishing susceptibility and contribute to the body of knowledge. For example, Spottswood and Hancock (2017) found *social norms* on SNSs can trigger heuristic processing.

As prior literature has shown that Maslow's hierarchy of needs has a relationship to the Big Five traits, this study argued that identifying certain needs associated with a particular trait could bring to light factors that contribute to habit formation. Classifying potential motivating factors such as self-esteem, boredom, loneliness and a need to belong (belongingness) as leading factors for performing a habit on SNSs offers interesting prospects for future research.

Much behavioural research focuses on the individual's attitude; however, this study focuses on various dimensions of the individual's personality which may generate interesting behavioural insights. Shropshire et al. (2006) state that using personality traits has two benefits. Firstly, as alluded to in Chapter 3, personality traits are stable over time (Funder, 2001) and as such can be used to make relatively longer-term projections of a behaviour than is possible with the attitude construct. Secondly, whereas an individual cannot have an attitude towards a technology that they are unaware of, they will always have a personality that can be measured. As such, personality traits provide a useful way of classifying risky users. For example, a study by Bagby et al. (2007) found that compulsive gamblers possessed the neuroticism trait and were significantly lower on the conscientiousness trait. Thus, users with this trait who are addicted to gambling might be risk takers, and as such might be more susceptible and more easily enticed to click on phishing posts related to winning money or to financially related opportunities.

There is a lack of studies investigating the influence of personality traits on habit. As mentioned earlier, this has been pointed out by Wood (2017), and by Vishwanath (2015a) who states that the “impact of habit and information processing on phishing deception is unclear”. This study showed that habits put users directly at risk of phishing. Interestingly, this study revealed that habit uses a form of heuristic processing due to the automaticity aspect of the behaviour. As a result, the inclusion of the habit construct in this study also makes a contribution. Valecha et al. (2015) state that there is a need for future research to investigate whether social media users assess the validity of a phishing message using systematic or heuristic or a combined approach. This study addressed this and found that heuristic processing puts users at risk to phishing.

Literature on phishing and SETA programme interventions has focused predominantly on email-based phishing and spoofed websites, and hence a lack of attention has been given to phishing conducted on SNSs. As such, the context of the study also offers a contribution. Other studies have not made use of the real-world phishing stimuli found on Facebook and moreover have not shown how each stimulus utilises Cialdini’s six persuasion principles. As such, the stimuli presented in Section 2.4 offers a contribution to experimental design.

9.4.2 Contribution to practice

The use of SNSs in organisational settings and for personal communication has blurred the distinction between work and personal time, thus offering further opportunities for data breaches in the workplace. Literature predominantly focuses on addressing human behaviour from an organisational perspective and not considering the risks users bring from a “home user”-level perspective (Ying Li & Siponen, 2011). It is possible that social media users develop certain habitual behaviours which are transferred from home or which develop in their personal time. As such, organisations could overlook these behaviours and also neglect this aspect in their training interventions, as they may not be able to detect or observe these behaviours (e.g. duration spent on Facebook) in the work environment, thus bringing security risks (Silic & Back, 2016).

As alluded to earlier, ultimately, having some effect on human behaviour will require education and training interventions. As mentioned by Lawson et al. (2020), it is not possible to develop and subsequently utilise a phishing susceptibility model to protect potential targets

without first identifying the areas of greatest susceptibility. As prior literature has criticised the effectiveness of SETA interventions as being too general, the findings of this study may assist organisations in the customisation of an individual anti-phishing training programme to target specific dispositional factors in vulnerable users (Burns et al., 2019). By using an instrument similar to the one used in this study, pre-assessments or simulated attacks could be used to determine and classify certain risk profiles by also considering their personality traits. Similarly, Mitnick and Simon (2002) suggest that organisations should carry out “auditing and testing” on employees in order to determine their susceptibility to SE attacks. Organisations could examine employees’ preferences to, for example, determine their interest in free gifts, movie genres, employment opportunities, financial stability and the like. These preferences could help to identify potential behavioural vulnerabilities that phishers could use to persuade victims in both emails and on SNSs. Once identified, more targeted and explicit training interventions could be conducted with the associated vulnerable groups to address employees’ personal sets of vulnerabilities with consideration of their personality traits. This approach has also been recommended by Mayhorn et al. (2015). Jensen et al. (2017) introduced “mindfulness” techniques into training programmes as an approach to teach individuals to allocate attention “dynamically” during message evaluation and anticipate judgement of suspicious messages. Ultimately, such training interventions are aimed at improving systematic processing in order to better detect phishing (Johnson et al., 2001). Another user-centred approach is to incorporate the use of personas into the awareness training itself (Ki-Aries & Faily, 2017). Based on each of the personality traits, a fiction-based or role-based perspective can be used in the design process to create the personas (Ki-Aries & Faily, 2017).

In addition, organisations can use the knowledge of habit formation to break the cycle of habitual behaviour. In this regard, SETA programmes may assist in replacing the insecure behaviour and establishing better behaviours by inserting a new routine between the cue and reward (Pfleeger et al., 2014). Other individual factors such as age, gender and technical expertise may also have an impact on the type of training that yields the best results (Jampen et al., 2020). In addition, Gardner (2015) states that self-regulatory skills training may be a valuable addition to interventions aimed at modifying habits via reflective motivation change. The importance of considering educational interventions aimed at improving users’ cognitive skills to avoid internet deception has been identified in early work (Grazioli, 2004). On the contrary, strong habits are not easily changeable by mere informational interventions but

involve disrupting the environmental factors that automatically prompt habit performance (Verplanken & Wood, 2006).

Harrison et al. (2015) recommend that anti-phishing interventions be developed in such a way as to also include educating individuals on the use of richness and presence cues in emails and the risks these could impose. In this regard, with a focus on information processing, training users to become more vigilant with regard to graphics, logos, and other elements may assist the user to judge the authenticity of an email. However, this study pointed out, as reported in the study by Vishwanath et al. (2011), that this could have an adverse effect in terms of which the user may focus only on systematic processing and thereby neglect the role of heuristics. From the literature, it is apparent that both modes of processing have a role to play if one is to conduct a thorough assessment of a phishing message. As a result, the stance taken in this study was to focus on the behavioural perspective, instead of the user interface perspective, which might help to identify to what extent the former would influence the two modes of processing.

9.5 Limitations and future research directions

While this study makes several important theoretical contributions and offers interesting insights into various spheres of behavioural research, there were several limitations that create possibilities for future research.

9.5.1 Sample size and demography

The study assumed that respondents might have pre-existing knowledge or experiences of phishing threats. The study used a convenience sampling method and cross-sectional study that involved only students within one particular university. As echoed by Butavicius et al. (2015), such a sample may not necessarily reflect the abilities of the wider population and, therefore, limits the generalisability of the study findings. In addition to increasing the sample size, it would be beneficial to collect data from an international population to determine whether the model could be generalised across different social and cultural contexts. Moreover, how “employees” respond to phishing attacks in the SNS context needs to be further investigated (Silic & Back, 2016). Demographics such as age and gender were not included in the

theoretical model. Although Mohebzada et al. (2012) suggest that demographics are not conclusive in predicting attack susceptibility, it would be interesting to see if certain personality traits, in performing a habit, would be influenced by the person's gender. This is not inconceivable as prior literature has shown that females habitually practice security compliance in their workplace (Nord et al., 2020).

9.5.2 Context

This study had a focus on Facebook as it is the most popular SNS and it offers many technical features that can be exploited by phishers. Although other SNSs have similar features to Facebook, SNSs are designed to fit a particular purpose which also creates opportunities for phishers to exploit particular information related to specific sites. For example, Cooper and Naatus (2014) state that LinkedIn users are comfortable having their personal information viewed by others. As such, if a user is applying for employment on this site they might be prepared to relinquish personal information such as their identity document/passport, curriculum vitae and copies of their qualifications to a phisher impersonating a recruiter. In contrast, if the same information were requested on Facebook, the user might be more reluctant to give it out. It is thus possible behaviours exhibited by users on one platform, such as Facebook, are not necessarily the same on other SNSs. This creates opportunities for future research to test if certain behaviours are consistent across other SNSs used by the same sample.

9.5.3 Experimental design

In adherence to ethical guidelines and procedures, an actual phishing experiment could not be conducted in a Facebook environment. As such, the phishing messages or stimuli used to test information processing in the study originated from the researcher's Facebook account. It is possible that this creates bias as the stimuli originated from acquaintances connected to the researcher and not to the respondents. Prior literature indicates that individuals are more likely to respond to posts from their friend connections than from unknown entities.

Moreover, that it was assumed that respondents would address the section on information processing in the survey using the same amount of time and attention to detail as they would in the Facebook environment. As the instrument was a survey, the measures used were indirect and consequently could not measure response time, which could be particularly important with

regard to information processing. However, a study by Bayl-Smith et al. (2020) points out that although participants spent more time evaluating an email (using systematic processing) and correctly identified phishing features (albeit a genuine email), this did not improve their ability to correctly detect actual phishing emails. Thus, the amount of time spent assessing for phishing cues does not conclusively prove that they will not fall victim to such phishing attacks. Although the survey instrument consisted of several persuasive messages, the instrument assessed respondents on a single message at a time. In the SNS environment, users would be exposed to a greater number of posts on a timeline at once, containing both textual and graphical elements, and viewing this at different times of the day. The messages were also mainly graphics which trigger heuristic processing and have been shown to be more effective for phishing attacks (Vishwanath, 2017). It is possible that users could dismiss posts that may pass through their timeline without applying any mode of processing, thus making no decision.

As surveys and observations have been shown to capture different factors concerning security behaviours in phishing experiments (Flores et al., 2014), a multimethod approach might be more suitable for this study design, in particular to observe which specific habits users are predominantly performing, for example liking, sharing or clicking on links. Future research can explore the direct effect of habit on phishing susceptibility to determine whether its influence is partial or full. In this regard, a partial or full mediation analysis of the model can be tested for moderator–mediator variable distinction (Baron & Kenny, 1986). For each of the Big Five traits, future research can investigate the approaches, mechanisms or content needed for security training interventions to be effective.

9.5.4 Persuasion principles

This study did not aim to identify which specific persuasion strategies are more effective for phishing, as has been done previously in other studies. As a result, the current instrument design did not test responses to each of the six persuasion principles. However, as noted by Lawson et al. (2017), phishers tend to use a combination of persuasion types and, thus, in such cases, the instrument in its current form cannot determine which specific persuasion principle a user is more likely to fall victim to. Furthermore, measuring certain persuasion principles such as “liking” makes it difficult to draw conclusions, as individuals each have their own set of preferences that they would “like”. This might explain why Lawson et al. (2020) experienced contradictory findings in their experiment when utilising the liking principle, as

respondents considered both phishing and legitimate emails to be trustworthy. Furthermore, it was not possible to assess the persuasion principles of “commitment” and “reciprocity”, as this would require prior knowledge of the respondents’ past choices and commitments. These specific principles were also identified by Butavicius et al. (2015) as being less suited to their laboratory study.

9.6 Publications emanating from the study

The following publications have resulted directly from the work in this thesis. The following conference paper served as a proof-of-concept and was presented at the 15th International Information Security for South Africa (ISSA) conference in 2016. The paper introduced social network phishing and focused on the behavioural concerns pertaining to the *Big Five personality traits, habit and information processing*. These three constructs contributed to the formation of the main backbone of the contribution and theoretical model of this study. The paper argues that due to the constant information updates users receive on SNSs, users may become habituated to clicking and sharing links, liking posts, and copying and pasting messages. As a result, this behavioural priming might lead to users becoming more susceptible to SE attacks on SNSs, as they do not cognitively process messages with sufficient scrutiny and attention to detail.

Frauenstein, E. D., & Flowerday, S. (2016). <i>Social network phishing: Becoming habituated to clicks and ignorant to threats?</i> In Proceedings of the 2016 Information Security for South Africa (ISSA), Johannesburg, South Africa, 2016 August 17–18, 98–105.
--

Following the previously mentioned proof-of-concept conference publication, a journal article was developed that focused only on two constructs of the study’s final model; namely, the *Big Five personality traits* and *information processing*. Utilising the HSM, the study examined the mediating role that information processing plays with regard to user susceptibility to social network phishing based on their personality traits, thereby identifying user characteristics that may be more susceptible than others to phishing on SNSs. The study revealed that conscientious users were found to have a negative influence on heuristic processing, and are thus less susceptible to phishing on SNSs. This confirmed that heuristic processing increases susceptibility to phishing.

Frauenstein, E. D., & Flowerday, S. (2020). Susceptibility to phishing on social network sites: A personality information processing model. *Computers & Security*, 94, 101862. <https://doi.org/10.1016/j.cose.2020.101862>

The following journal article was written to convey the main findings of the final theoretical model consisting of all the constructs reported in this thesis.

Frauenstein, E. D., Flowerday, S., Mishi, S., & Warkentin, M. (2021). User susceptibility to phishing on social network sites: A personality-habit-information processing model. *Information & Management* (Under review).

9.7 Conclusion

As in previous years, phishing remains an ongoing concern, posing significant financial losses for both organisations and related consumers. Like seasonal changes, phishing techniques evolve, taking advantage of new events, catastrophes and global headlines and incorporating these in the design of persuasive messages. Phishing on SNSs preys to a large extent on the inclination of people to be curious and trusting, thus making it difficult for organisations to address this threat by means of security education, training and awareness programmes. People may serve as a protective measure but only if they pay attention and “recognise” the threat. However, owing to the individual behavioural vulnerabilities that characterise users, any security awareness effort may be limited when users are faced with phishing. Thus, any steps taken to protect users should also include understanding the individual characteristics that may influence user behaviour that makes them vulnerable. In this regard, individual differences or characteristics continue to be an area of exploration for many scholars in the information security domain (Mayhorn et al., 2015; Saleem et al., 2011; Williams et al., 2017). More development and understanding of how these behavioural areas respond to threats will lead to improvement in models to come.

This study identified that the popularity of SNSs and the technical features they offer users have created many opportunities for malicious individuals to exploit the behavioral tendencies of their users via SE tactics. Certain types of social media users are frequently highly engaged with other users and are thus inadvertently priming their behaviour, which can make them less

attentive and lack caution regarding malicious posts on these sites. As a result, this study aimed to develop a theoretical model to identify users who are more likely to be susceptible to phishing on SNSs. Aligned to the individual characteristics alluded to earlier, this study examined the influence of personality traits and habits on information processing. Moreover, it also examined the influence of social norms, computer self-efficacy and perceived risk on phishing susceptibility.

In order to fulfil the main aim of this research, SEM techniques were applied to data derived from a sample of 215 participants. The study yielded several statistically significant findings which contribute to this growing body of research. The results revealed that all of the Big Five traits, apart from openness, showed statistically significant relationships to habit. The conscientiousness trait alone was the trait least at risk of forming a habit. Habit was also shown to have a significant influence on both heuristic and systematic processing, which may influence the outcome of phishing susceptibility. Owing to the direct influence of habit on phishing susceptibility, the study has also shown that habit and heuristic processing may be part of the same process. This is because both habit and heuristic processing involve some aspects of *automaticity*, which advantageously allow individuals to be efficient, using limited cognitive resources when engaging in a task; however, both significantly increase susceptibility to phishing on SNSs. Systematic processing reduces the likelihood of susceptibility to phishing, thus supporting prior studies in this area.

In closing, although technology serves a purpose in detecting phishing, users should ideally also be able to identify phishing for themselves. However, Furnell et al. (2019) use the following metaphor in this regard: technology can offer a “net” of protection to safeguard users from phishing attacks but this net has “big holes”. This emphasises the need for users to play their role in addressing this vulnerability. However, Furnell et al. (2019) add that unfortunately experience has shown that users are not good at this.

REFERENCES

- Aarts, H., Verplanken, B., & Van Knippenberg, A. (1998). Predicting behavior from actions in the past: Repeated decision making or a matter of habit? *Journal of Applied Social Psychology, 28*(15), 1355–1374. <https://doi.org/10.1111/j.1559-1816.1998.tb01681.x>
- Ab Hamid, M. R., Sami, W., & Mohmad Sidek, M. H. (2017). Discriminant validity assessment: Use of Fornell & Larcker criterion versus HTMT Criterion. *Journal of Physics: Conference Series, 890*, 012163. <https://doi.org/10.1088/1742-6596/890/1/012163>
- Abbott, M. L., & McKinney, J. (2013). *Understanding and applying research design*. John Wiley & Sons.
- Abu-Nimeh, S., Nappa, D., Wang, X., & Nair, S. (2007). *A comparison of machine learning techniques for phishing detection*. Paper presented at the anti-phishing working groups 2nd annual eCrime researchers summit, Pittsburgh, PA.
- Acopio, J. R., & Bance, L. (2016). Personality traits as predictors of Facebook use. *International Journal of Psychology and Counselling, 8*, 45–52. <https://doi.org/10.5897/IJPC2015.0311>
- Adewole, K. S., Anuar, N. B., Kamsin, A., Varathan, K. D., & Razak, S. A. (2017). Malicious accounts: Dark of the social networks. *Journal of Network and Computer Applications, 79*(1), 41–67. <https://doi.org/10.1016/j.jnca.2016.11.030>
- Agrawal, A. (2016). *Clickbait detection using deep learning*. Paper presented at the 2nd International Conference on Next Generation Computing Technologies (NGCT) 2016, Dehradun, India.
- Ajzen, I. (1987). Attitudes, traits, and actions: Dispositional prediction of behaviour in personality and social psychology. In L. Berkowitz (Ed.), *Advances in experimental social psychology, 20*, 1–63. Academic Press.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Ajzen, I. (2011). The theory of planned behaviour: Reactions and reflections. *Psychology & Health, 26*(9), 1113–1127. <https://doi.org/10.1080/08870446.2011.613995>
- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Prentice-Hall.
- Akbar, N. (2014). *Analysing persuasion principles in phishing emails* (MSc Computer Science). University of Twente. <http://essay.utwente.nl/66177/>
- Alabdan, R. (2020). Phishing attacks survey: Types, vectors, and technical approaches. *Future Internet, 12*(10), 168. <https://www.mdpi.com/1999-5903/12/10/168>
- Albladi, S. M., & Weir, G. R. S. (2017). *Personality traits and cyber-attack victimisation: multiple mediation analysis*. Paper presented at the Joint 13th CTTE and 10th CMI Conference on Internet of Things – Business Models, Users, and Networks, Piscataway, NJ.
- Albladi, S. M., & Weir, G. R. S. (2018). User characteristics that influence judgment of social engineering attacks in social networks. *Human-centric Computing and Information Sciences, 8*(1). <https://doi.org/10.1186/s13673-018-0128-7>
- Aleroud, A., & Zhou, L. (2017). Phishing environments, techniques, and countermeasures. *Computers & Security, 68*(C), 160–196. <https://doi.org/10.1016/j.cose.2017.04.006>
- Algarni, A., Xu, Y., & Chan, T. (2014). *Social engineering in social networking sites: The art of impersonation*. Paper presented at the 2014 IEEE International Conference on Services Computing, Anchorage, AK.

- Algarni, A., Xu, Y., & Chan, T. (2015). *Susceptibility to social engineering in social networking sites: The case of Facebook*. Paper presented at the International Conference on Information Systems (ICIS 2015), Fort Worth, TX.
- Algarni, A., Xu, Y., & Chan, T. (2016). *Measuring source credibility of social engineering attackers on Facebook*. Paper presented at the 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI.
- Algarni, A., Xu, Y., & Chan, T. (2017). An empirical study on the susceptibility to social engineering in social networking sites: the case of Facebook. *European Journal of Information Systems*, 26(6), 661–687. <https://doi.org/10.1057/s41303-017-0057-y>
- Algarni, A., Xu, Y., Chan, T., & Tian, Y.-C. (2013a). *Social engineering in social networking sites: Affect-based model*. Paper presented at the 8th International Conference for Internet Technology and Secured Transactions (ICITST 2013), London, UK.
- Algarni, A., Xu, Y., Chan, T., & Tian, Y.-C. (2013b). *Toward understanding social engineering*. Paper presented at the 8th International Conference on Legal, Security and Privacy Issues in IT Law, (Critical Analysis and Legal Reasoning), Bangkok, Thailand.
- Algarni, A., Xu, Y., Chan, T., & Tian, Y.-C. (2014). *Social engineering in social networking sites: How good becomes evil*. Paper presented at the 18th Pacific Asia Conference on Information Systems (PACIS 2014), Chengdu, China.
- Alkıř, N., & Temizel, T. T. (2015). The impact of individual differences on influence strategies. *Personality and Individual Differences*, 87, 147–152. <https://doi.org/10.1016/j.paid.2015.07.037>
- Allen, M. (2006). Social engineering: A means to violate a computer system. <https://www.sans.org/reading-room/whitepapers/engineering/social-engineering-means-violate-computer-system-529>
- Alotaibi, M. (2019). *A hypothesised model to examine susceptibility to cyber-social engineering through LinkedIn in the workplace*. Paper presented at the 13th Human Aspects of Information Security & Assurance (HAISA 2019), Nicosia, Cyprus.
- Alqarni, Z., Algarni, A., & Xu, Y. (2016). *Toward predicting susceptibility to phishing victimization on Facebook*. Paper presented at the IEEE International Conference on Services Computing (SCC 2016), San Francisco, CA, USA.
- Alseadoon, I., Othman, M. F. I., & Chan, T. (2015). *What is the influence of users' characteristics on their ability to detect phishing emails?* Paper presented at the 1st International Conference on Communication and Computer Engineering, Malaysia.
- Alutaybi, A., Arden-Close, E., McAlaney, J., Stefanidis, A., Phalp, K., & Ali, R. (2019). *How can social networks design trigger fear of missing out?* Paper presented at the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy.
- Amer, T. S., & Maris, J. M. B. (2007). Signal words and signal icons in application control and information technology exception messages-Hazard matching and habituation effects. *Journal of Information Systems*, 21(2), 1–25. <https://doi.org/10.2308/jis.2007.21.2.1>
- Amichai-Hamburger, Y. (2002). Internet and personality. *Computers in Human Behavior*, 18(1), 1–10. [https://doi.org/10.1016/S0747-5632\(01\)00034-6](https://doi.org/10.1016/S0747-5632(01)00034-6)
- Amichai-Hamburger, Y., & Ben-Artzi, E. (2000). The relationship between extraversion and neuroticism and the different uses of the internet. *Computers in Human Behavior*, 16(4), 441–449. [https://doi.org/10.1016/S0747-5632\(00\)00017-0](https://doi.org/10.1016/S0747-5632(00)00017-0)
- Amichai-Hamburger, Y., & Ben-Artzi, E. (2003). Loneliness and internet use. *Computers in Human Behavior*, 19(1), 71–80. [https://doi.org/10.1016/S0747-5632\(02\)00014-6](https://doi.org/10.1016/S0747-5632(02)00014-6)

- Amichai-Hamburger, Y., & Vinitzky, G. (2010). Social network use and personality. *Computers in Human Behavior*, 26(6), 1289–1295. <https://doi.org/10.1016/j.chb.2010.03.018>
- Amichai-Hamburger, Y., Wainapel, G., & Fox, S. (2002). "On the internet no one knows I'm an introvert": Extroversion, neuroticism, and internet interaction. *CyberPsychology & Behavior*, 5(2), 125–128. <https://doi.org/10.1089/109493102753770507>
- Andreassen, C. S., Griffiths, M. D., Gjertsen, S. R., Krossbakken, E., Kvam, S., & Pallesen, S. (2013). The relationships between behavioral addictions and the five-factor model of personality. *Journal of Behavioral Addictions*, 2(2), 90–99. <https://doi.org/10.1556/jba.2.2013.003>
- APWG. (2021). Phishing activity trends report, 4th quarter 2020. https://docs.apwg.org/reports/apwg_trends_report_q4_2020.pdf
- Arachchilage, N. A. G., & Love, S. (2014). Security awareness of computer users: A phishing threat avoidance perspective. *Computers in Human Behavior*, 38, 304–312. <https://doi.org/10.1016/j.chb.2014.05.046>
- Ardelt, M. (2000). Still stable after all these years? Personality stability theory revisited. *Social Psychology Quarterly*, 63(4), 392–405. <https://doi.org/10.2307/2695848>
- Arntz, P. (2019). Explained: Like-farming. <https://blog.malwarebytes.com/101/2019/04/explained-like-farming/>
- Bada, M., Sasse, A., & Nurse, J. R. C. (2015). *Cyber security awareness campaigns: Why do they fail to change behaviour?* International Conference on Cyber Security for Sustainable Society, Coventry, UK.
- Bagby, R., Vachon, D., Bulmash, E., Toneatto, T., Quilty, L., & Costa, P. (2007). Pathological gambling and the five-factor model of personality. *Personality and Individual Differences*, 43, 873–880. <https://doi.org/10.1016/j.paid.2007.02.011>
- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16(1), 74–94. <https://doi.org/10.1007/BF02723327>
- Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly*, 36(3), 421–458. <https://doi.org/10.2307/2393203>
- Bailey, J. L., Mitchell, R. B., & Jensen, B. K. (2008). *Analysis of student vulnerabilities to phishing*. Paper presented at the 14th Americas Conference on Information Systems (AMCIS 2008), Toronto, Ontario, Canada.
- Balcerowska, J. M., Bereznowski, P., Biernatowska, A., Atroszko, P. A., Pallesen, S., & Andreassen, C. S. (2020). Is it meaningful to distinguish between Facebook addiction and social networking sites addiction? Psychometric analysis of Facebook addiction and social networking sites addiction scales. *Current Psychology*. <https://doi.org/10.1007/s12144-020-00625-3>
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2), 122–147. <https://doi.org/10.1037/0003-066X.37.2.122>
- Bandura, A. (1990). Perceived self-efficacy in the exercise of control over AIDS infection. *Evaluation and Program Planning*, 13(1), 9–17. [https://doi.org/10.1016/0149-7189\(90\)90004-G](https://doi.org/10.1016/0149-7189(90)90004-G)
- Bandura, A. (1999). A social cognitive theory of personality. In L. A. Pervin & O. John (Eds.), *Handbook of personality* (2nd ed., pp. 154–196). Guilford Publications.
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal*

- of *Personality and Social Psychology*, 51(6), 1173–1182.
<https://doi.org/10.1037/0022-3514.51.6.1173>
- Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, 42(5), 815–824. <https://doi.org/10.1016/j.paid.2006.09.018>
- Bashir, I., & Madhavaiah, C. (2015). Trust, social influence, self-efficacy, perceived risk and internet banking acceptance: An extension of technology acceptance model in Indian context. *Metamorphosis*, 14(1), 25–38. <https://doi.org/10.1177/0972622520150105>
- Bayl-Smith, P., Sturman, D., & Wiggins, M. (2020). *Cue utilization, phishing feature and phishing email detection*. Paper presented at the 24th International Conference on Financial Cryptography and Data Security (FC2020), Kota Kinabalu Malaysia.
- Becker, J., Knackstedt, R., Lis, L., & Stein, A. (2010). *Towards a maturity model for research portals*. Paper presented at the 18th European Conference on Information Systems (ECIS 2010), Pretoria, South Africa.
- Berger, C. R., & Calabrese, R. J. (1975). Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication. *Human Communication Research*, 1(2), 99–112. <https://doi.org/10.1111/j.1468-2958.1975.tb00258.x>
- Bhardwaj, A., Sapra, V., Kumar, A., Kumar, N., & Arthi, S. (2020). Why is phishing still successful? *Computer Fraud & Security*, 2020(9), 15–19. [https://doi.org/10.1016/S1361-3723\(20\)30098-1](https://doi.org/10.1016/S1361-3723(20)30098-1)
- Bhaskar, R. (1978). *A realist theory of science* (2nd ed.). Harvester Press.
- Binks, A. (2019). The art of phishing: past, present and future. *Computer Fraud & Security*, 2019(4), 9-11. [https://doi.org/10.1016/S1361-3723\(19\)30040-5](https://doi.org/10.1016/S1361-3723(19)30040-5)
- Blackwell, D., Leaman, C., Trampusch, R., Osborne, C., & Liss, M. (2017). Extraversion, neuroticism, attachment style and fear of missing out as predictors of social media use and addiction. *Personality and Individual Differences*, 116, 69–72. <https://doi.org/10.1016/j.paid.2017.04.039>
- Blythe, M., Petrie, H., & Clark, J. A. (2011). *F for fake: four studies on how we fall for phish*. Paper presented at the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada. <https://doi.org/10.1145/1978942.1979459>
- Bohner, G., Moskowitz, G. B., & Chaiken, S. (1995). The interplay of heuristic and systematic processing of social information. *European Review of Social Psychology*, 6(1), 33–68. <https://doi.org/10.1080/14792779443000003>
- Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic approaches to a successful literature review* (2nd ed.). SAGE Publications.
- Boudreaux, M. J., & Ozer, D. J. (2015). Five factor model of personality, assessment of. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed. 230–235. Elsevier.
- Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications*, 10(1), 7. <https://doi.org/10.1038/s41467-018-07761-2>
- Bowman, N. D. (2017). The Importance of effect size reporting in communication research reports. *Communication Research Reports*, 34(3), 187–190. <https://doi.org/10.1080/08824096.2017.1353338>
- BrainyQuote. (2019). Frank Abagnale quotes. https://www.brainyquote.com/quotes/frank_abagnale_790763
- Branley, D. B., & Covey, J. (2018). Risky behavior via social media: The role of reasoned and social reactive pathways. *Computers in Human Behavior*, 78, 183–191. <https://doi.org/10.1016/j.chb.2017.09.036>
- Brehm, J. W. (1966). *A theory of psychological reactance*. Academic Press.

- Briggs, P., Jeske, D., & Coventry, L. (2017). Chapter 6 Behavior change interventions for cybersecurity. In L. Little, E. Sillence, & A. Joinson (Eds.), *Behavior change research and theory* (pp. 115–136). Academic Press.
- Bryman, A. (2001). *Social research methods* (2nd ed.). Oxford University Press.
- Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2010). Information security policy compliance: An empirical study of rationality-based beliefs and information security awareness. *MIS Quarterly*, *34*(3), 523–548. <https://doi.org/10.2307/25750690>
- Bullée, J.-W., Montoya, L., Pieters, W., Junger, M., & Hartel, P. H. (2015). The persuasion and security awareness experiment: reducing the success of social engineering attacks. *Journal of Experimental Criminology*, *11*(1), 97–115. <https://doi.org/10.1007/s11292-014-9222-7>
- Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, *6*(3), 203–242. <https://doi.org/10.1111/j.1468-2885.1996.tb00127.x>
- Burns, A. J., Johnson, M. E., & Caputo, D. D. (2019). Spear phishing in a barrel: Insights from a targeted phishing campaign. *Journal of Organizational Computing and Electronic Commerce*, *29*(1), 24–39. <https://doi.org/10.1080/10919392.2019.1552745>
- Burns, M., Durcikova, A., & Jenkins, J. L. (2013). *What kind of interventions can help users from falling for phishing attempts: A research proposal for examining stage-appropriate interventions*. Paper presented at the 46th Hawaii International Conference on System Sciences, Hawaii, USA.
- Butavicius, M., Parsons, K., Pattinson, M., & McCormac, A. (2015). *Breaching the human firewall: Social engineering in phishing and spear-phishing emails*. Paper presented at the Australasian Conference on Information Systems, Adelaide, Australia.
- Butt, S., & Phillips, J. G. (2008). Personality and self reported mobile phone use. *Computers in Human Behavior*, *24*(2), 346–360. <https://doi.org/10.1016/j.chb.2007.01.019>
- Cabrera-Nguyen, E. (2010). Author guidelines for reporting scale development and validation results. *Journal of the Society for Social Work and Research*, *1*, 99–103. <https://doi.org/10.5243/jsswr.2010.8>
- Cameron, K. A. (2009). A practitioner’s guide to persuasion: An overview of 15 selected persuasion theories, models and frameworks. *Patient Education and Counseling*, *74*, 309–317. <https://doi.org/10.1016/j.pec.2008.12.003>
- Canfield, C. I., Fischhoff, B., & Davis, A. (2016). Quantifying phishing susceptibility for detection and behavior decisions. *Human Factors*, *58*(8), 1158–1172. <https://doi.org/10.1177/0018720816665025>
- Canham, M., Constantino, M., Hudson, I., Fiore, S. M., Caulkins, B., & Reinerman-Jones, L. (2019). *The Enduring Mystery of the Repeat Clickers*, Paper presented at the 15th Symposium on Usable Privacy and Security (SOUPS 2019), Santa Clara, CA.
- Canova, G., Volkamer, M., Bergmann, C., & Borza, R. (2014). *NoPhish: An anti-phishing education app*. Paper presented at the Security and Trust Management (STM 2014), Wroclaw, Poland.
- Carducci, G., Rizzo, G., Monti, D., Palumbo, E., & Morisio, M. (2018). TwitPersonality: Computing personality traits from tweets using word embeddings and supervised learning. *Information (Switzerland)*, *9*(5). <https://doi.org/10.3390/info9050127>
- Cash, H., Rae, C. D., Steel, A. H., & Winkler, A. (2012). Internet addiction: A brief summary of research and practice. *Current Psychiatry Reviews*, *8*(4), 292–298. <https://doi.org/10.2174/157340012803520513>
- Cellar, D. F., Nelson, Z. C., Yorke, C. M., & Bauer, C. (2001). The five-factor model and safety in the workplace: Investigating the relationships between personality and accident involvement. *Journal of Prevention & Intervention in the Community*, *22*(1), 43–52. <https://doi.org/10.1080/10852350109511210>

- Chan, T. K. H., Cheung, C. M. K., & Lee, Z. W. Y. (2021). Cyberbullying on social networking sites: A literature review and future research directions. *Information & Management*, 58(2), 103411. <https://doi.org/10.1016/j.im.2020.103411>
- Chang, S.-J., van Witteloostuijn, A., & Eden, L. (2010). From the Editors: Common method variance in international business research. *Journal of International Business Studies*, 41(2), 178–184. <https://doi.org/10.1057/jibs.2009.88>
- Chapman, B. P., Duberstein, P. R., Sörensen, S., & Lyness, J. M. (2007). Gender differences in five factor model personality traits in an elderly cohort: Extension of robust and surprising findings to an older generation. *Personality and Individual Differences*, 43(6), 1594–1603. <https://doi.org/10.1016/j.paid.2007.04.028>
- Chen, C., Wen, S., Zhang, J., Xiang, Y., Oliver, J., Alelaiwi, A., & Hassan, M. M. (2017). Investigating the deceptive information in Twitter spam. *Future Generation Computer Systems*, 72, 319–326. <https://doi.org/10.1016/j.future.2016.05.036>
- Chen, J., Mishler, S., Hu, B., Li, N., & Proctor, R. W. (2018). The description-experience gap in the effect of warning reliability on user trust and performance in a phishing-detection context. *International Journal of Human-Computer Studies*, 119, 35–47. <https://doi.org/10.1016/j.ijhcs.2018.05.010>
- Chen, S., Duckworth, K., & Chaiken, S. (1999). Motivated heuristic and systematic processing. *Psychological Inquiry*, 10(1), 44–49. https://doi.org/10.1207/s15327965pli1001_6
- Chen, Y., Conroy, N. J., & Rubin, V. L. (2015). *Misleading online content: Recognizing clickbait as "false news"*. Paper presented at the 2015 ACM on Workshop on Multimodal Deception Detection, Seattle, Washington, USA.
- Chen, Y., Zahedi, F. M., Abbasi, A., & Dobolyi, D. (2021). Trust calibration of automated security IT artifacts: A multi-domain study of phishing-website detection tools. *Information & Management*, 58(1), 103394. <https://doi.org/10.1016/j.im.2020.103394>
- Cheung, C., Lee, Z. W. Y., & Chan, T. K. H. (2015). Self-disclosure in social networking sites: The role of perceived cost, perceived benefits and social influence. *Internet Research*, 25(2), 279–299. <https://doi.org/10.1108/IntR-09-2013-0192>
- Cheung, C. M. K., Chiu, P.-Y., & Lee, M. K. O. (2011). Online social networks: Why do students use facebook? *Computers in Human Behavior*, 27(4), 1337–1343. <https://doi.org/10.1016/j.chb.2010.07.028>
- Chi, M. (2011). Reducing the risks of social media to your organization. <https://www.sans.org/reading-room/whitepapers/policyissues/reducing-risks-social-media-organization-33749>
- Chiew, K. L., Yong, K. S. C., & Tan, C. L. (2018). A survey of phishing attacks: Their types, vectors and technical approaches. *Expert Systems with Applications*, 106, 1–20. <https://doi.org/10.1016/j.eswa.2018.03.050>
- Cho, J.-H., Cam, H., & Oltramari, A. (2016). *Effect of personality traits on trust and risk to phishing vulnerability: Modeling and analysis*. Paper presented at the 2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA). San Diego, CA.
- Choi, D.-H., Yoo, W., Noh, G.-Y., & Park, K. (2017). The impact of social media on risk perceptions during the MERS outbreak in South Korea. *Computers in Human Behavior*, 72, 422–431. <https://doi.org/10.1016/j.chb.2017.03.004>
- Choi, T. R., Sung, Y., Lee, J.-A., & Choi, S. M. (2017). Get behind my selfies: The big five traits and social networking behaviors through selfies. *Personality and Individual Differences*, 109, 98–101. <https://doi.org/10.1016/j.paid.2016.12.057>

- Chowdhury, M. J. M., & Chakraborty, N. R. (2014). Social networking sites: Threat to security. *Journal of Modern Science and Technology*, 2(2), 87–98.
- Cialdini, R. B. (2007). *Influence: The psychology of persuasion*. Harper Collins.
- Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In *The handbook of social psychology, Vol. 1-2* (4th ed.; pp. 151–192). McGraw-Hill.
- Civelek, M. E. (2018). *Essentials of structural equation modeling*. <https://digitalcommons.unl.edu/zeabook/64/>
- Clark, A. M. (1998). The qualitative-quantitative debate: moving from positivism and confrontation to post-positivism and reconciliation. *Journal of Advanced Nursing*, 27(6), 1242–1249. <https://doi.org/10.1046/j.1365-2648.1998.00651.x>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.
- Collis, J., & Hussey, R. (2014). *Business research A practical guide for undergraduate and postgraduate students* (4th ed.). Palgrave.
- Colwill, C. (2009). Human factors in information security: The insider threat – Who can you trust these days? *Information Security Technical Report*, 14(4), 186–196. <https://doi.org/10.1016/j.istr.2010.04.004>
- Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, 19(2), 189–211. <https://doi.org/10.2307/249688>
- Conway, D., Taib, R., Harris, M., Berkovsky, S., Yu, K., & Chen, F. (2017). *A qualitative investigation of bank employee experiences of information security and phishing*. Paper presented at the Thirteenth USENIX Conference on Usable Privacy and Security, Santa Clara, CA.
- Cooper, B., & Naatus, M. (2014). LinkedIn as a learning tool in business education. *American Journal of Business Education*, 7(4), 299–306. <https://doi.org/10.19030/ajbe.v7i4.8815>
- Corner, L. (2012). Teenage kicks: How the 'bling ring' gang used Twitter to burgle Hollywood homes. <https://www.independent.co.uk/news/world/americas/teenage-kicks-how-bling-ring-gang-used-twitter-burgle-hollywood-homes-7804764.html>
- Correa, T., Hinsley, A. W., & de Zúñiga, H. G. (2010). Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in Human Behavior*, 26(2), 247–253. <https://doi.org/10.1016/j.chb.2009.09.003>
- Costa Jr, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, 81(2), 322–331. <https://doi.org/10.1037/0022-3514.81.2.322>
- Costa, P., & McCrae, R. C. (1985). *The NEO personality inventory manual*. Psychological Assessment Resources, Inc.
- Costa, P., & McCrae, R. C. (1992a). *The revised NEO personality inventory (NEO-PI-R)* (Vol. 2). Psychological Assessment Resources.
- Costa, P. T., & McCrae, R. R. (1992b). Four ways five factors are basic. *Personality and Individual Differences*, 13(6), 653–665. [https://doi.org/10.1016/0191-8869\(92\)90236-I](https://doi.org/10.1016/0191-8869(92)90236-I)
- Cox, J. (2012). Information systems user security: A structured model of the knowing–doing gap. *Computers in Human Behavior*, 28(5), 1849–1858. <https://doi.org/10.1016/j.chb.2012.05.003>
- Crano, W. D., & Prislin, R. (2006). Attitudes and persuasion. *Annual Review of Psychology*, 57(1), 345–374. <https://doi.org/10.1146/annurev.psych.57.102904.190034>
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed approaches*. (2nd ed.). Sage Publications.

- Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). Sage Publications.
- Crossler, R., & Bélanger, F. (2014). An extended perspective on individual security behaviors: Protection motivation theory and a unified security practices (USP) instrument. *SIGMIS Database*, 45(4), 51–71. <https://doi.org/10.1145/2691517.2691521>
- Cusack, B., & Adedokun, K. (2018). *The impact of personality traits on user's susceptibility to social engineering attacks*. Paper presented at the 16th Australian Information Security Management Conference, Perth, Australia.
- Da Veiga, A., & Eloff, J. H. P. (2010). A framework and assessment instrument for information security culture. *Computers & Security*, 29(2), 196–207. <https://doi.org/10.1016/j.cose.2009.09.002>
- Dahlberg, T., & Oorni, A. (2007). *Understanding changes in consumer payment habits - Do mobile payments and electronic invoices attract consumers?* Paper presented at the 40th Annual Hawaii International Conference on System Sciences (HICSS 2007), Big Island, Hawaii.
- Damian, R. I., Spengler, M., Sutu, A., & Roberts, B. W. (2019). Sixteen going on sixty-six: A longitudinal study of personality stability and change across 50 years. *Journal of Personality and Social Psychology*, 117(3), 674–695. <https://doi.org/10.1037/pspp0000210>
- Darwish, A., Zarka, A. E., & Aloul, F. (2012). *Towards understanding phishing victims' profile*. Paper presented at the 2012 International Conference on Computer Systems and Industrial Informatics, Sharjah, United Arab Emirates.
- Davis, J. M., & Tuttle, B. M. (2013). A heuristic–systematic model of end-user information processing when encountering IS exceptions. *Information & Management*, 50(2), 125–133. <https://doi.org/10.1016/j.im.2012.09.004>
- de Bruijn, H., & Janssen, M. (2017). Building cybersecurity awareness: The need for evidence-based framing strategies. *Government Information Quarterly*, 34(1), 1–7. <https://doi.org/10.1016/j.giq.2017.02.007>
- de Bruin, T., Freeze, R., Kulkarni, U., & Rosemann, M. (2005). *Understanding the main phases of developing a maturity assessment model*. Paper presented at the Australasian Conference on Information Systems (ACIS 2005), Sydney, Australia.
- de Carvalho, J., & Chima, F. O. (2014). Applications of structural equation modeling in social sciences research. *American International Journal of Contemporary Research*, 4(1), 6–11.
- Dewan, P., & Kumaraguru, P. (2017). Facebook inspector (FbI): Towards automatic real-time detection of malicious content on Facebook. *Social Network Analysis and Mining*, 7(1), 1–25. <https://doi.org/10.1007/s13278-017-0434-5>
- Dhamija, R., Tygar, J. D., & Hearst, M. (2006). *Why phishing works*. Paper presented at the SIGCHI Conference on Human Factors in Computing Systems, Montreal, Quebec, Canada.
- Dincelli, E., & Goel, S. (2017). *Can privacy and security be friends? A cultural framework to differentiate security and privacy behaviors on online social networks*. Paper presented at the 50th Hawaii International Conference on System Sciences (HICSS). Waikoloa Village, Hawaii
- Dixit, R. V., & Prakash, G. (2018). Intentions to use social networking sites (SNS) Using technology acceptance model (TAM): An empirical study. *Paradigm*, 22(1), 65–79. <https://doi.org/10.1177/0971890718758201>
- Dodge, R. C., Carver, C., & Ferguson, A. J. (2007). Phishing for user security awareness. *Computers & Security*, 26(1), 73–80. <https://doi.org/10.1016/j.cose.2006.10.009>

- Dogrue, L., Joeckel, S., & Bowman, N. D. (2015). Choosing the right app: An exploratory perspective on heuristic decision processes for smartphone app selection. *Mobile Media & Communication*, 3(1), 125–144. <https://doi.org/10.1177/2050157914557509>
- Dowd, E. T., & Seibel, C. A. (1990). A cognitive theory of resistance and reactance: Implications for treatment. *Journal of Mental Health Counseling*, 12(4), 458–469.
- Downs, J. S., Holbrook, M. B., & Cranor, L. F. (2006). *Decision strategies and susceptibility to phishing*. Paper presented at the 2nd symposium on Usable privacy and security, Pittsburgh, PA. <https://doi.org/10.1145/1143120.1143131>
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt Brace Jovanovich College Publishers.
- Edward, G. (2020). COVID-19-related scams cost Americans over \$160 million. <https://atlasvpn.com/blog/covid-19-related-scams-cost-americans-over-160-million>
- 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. <https://doi.org/10.1016/j.cose.2016.12.013>
- Egelman, S., Cranor, L. F., & Hong, J. (2008). *You've been warned: An empirical study of the effectiveness of web browser phishing warnings*. Paper presented at the SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy.
- Erder, M., & Pureur, P. (2016). Chapter 8 Role of the architect. In M. Erder & P. Pureur (Eds.), *Continuous architecture* (pp. 187–213). Morgan Kaufmann.
- Erjavec, J., Popovič, A., & Trkman, P. (2019). The effect of personality traits and knowledge on the quality of decisions in supply chains. *Economic Research-Ekonomska Istraživanja*, 32(1), 2269–2292. <https://doi.org/10.1080/1331677X.2019.1642788>
- Facebook. (2021). What steps can I take to protect myself from phishing on Facebook? <https://www.facebook.com/help/166863010078512>
- Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S. R., Park, H., & Shao, C. (2016). Applications of structural equation modeling (SEM) in ecological studies: an updated review. *Ecological Processes*, 5(1), 19. <https://doi.org/10.1186/s13717-016-0063-3>
- Featherman, M. S., & Pavlou, P. A. (2003). Predicting e-services adoption: A perceived risk facets perspective. *International Journal of Human-Computer Studies*, 59(4), 451–474. [https://doi.org/10.1016/S1071-5819\(03\)00111-3](https://doi.org/10.1016/S1071-5819(03)00111-3)
- Feng, Y., & Xie, W. (2014). Teens' concern for privacy when using social networking sites: An analysis of socialization agents and relationships with privacy-protecting behaviors. *Computers in Human Behavior*, 33, 153–162. <https://doi.org/10.1016/j.chb.2014.01.009>
- Ferreira, A., Coventry, L., & Lenzini, G. (2015). *Principles of persuasion in social engineering and their use in phishing*. Paper presented at the 3rd International Conference on Human Aspects of Information Security, Privacy, and Trust, Los Angeles, CA, USA. https://doi.org/10.1007/978-3-319-20376-8_4
- Ferreira, A., & Teles, S. (2019). Persuasion: How phishing emails can influence users and bypass security measures. *International Journal of Human-Computer Studies*, 125, 19–31. <https://doi.org/10.1016/j.ijhcs.2018.12.004>
- Ferrer, R., & Klein, W. M. (2015). Risk perceptions and health behavior. *Current Opinion in Psychology*, 5, 85–89. <https://doi.org/10.1016/j.copsyc.2015.03.012>
- Fette, I., Sadeh, N., & Tomasic, A. (2007). *Learning to detect phishing emails*. Paper presented at the 16th international conference on World Wide Web, Banff, Alberta, Canada. <https://doi.org/10.1145/1242572.1242660>
- Finn, P., & Jakobsson, M. (2007). Designing ethical phishing experiments. *IEEE Technology and Society Magazine*, 26(1), 46–58. <https://doi.org/10.1109/MTAS.2007.335565>

- Fire, M., Goldschmidt, R., & Elovici, Y. (2014). Online social networks: Threats and solutions. *IEEE Communication Surveys & Tutorials*, 16(4).
<https://doi.org/10.1109/COMST.2014.2321628>
- Florencio, D., & Herley, C. (2007). *A large-scale study of web password habits*. Paper presented at the 16th international conference on World Wide Web, Banff, Alberta, Canada.
- Flores, W.R., Holm, H., Svensson, G., & Ericsson, G. (2014). Using phishing experiments and scenario-based surveys to understand security behaviours in practice. *Information Management & Computer Security*, 22(4), 393–406. <https://doi.org/10.1108/IMCS-11-2013-0083>
- Fogel, J., & Nehmad, E. (2009). Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in Human Behavior*, 25(1), 153–160.
<https://doi.org/10.1016/j.chb.2008.08.006>
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50. <https://doi.org/10.2307/3151312>
- Fox, N. (2008). Post-positivism. In L. M. Given (Ed.), *The SAGE encyclopaedia of qualitative research methods*. SAGE Publications.
- Frauenstein, E. D. (2018). *An investigation into students responses to various phishing emails and other phishing-related behaviours*. Paper presented at the 17th Information Security for South Africa conference (ISSA 2018), Pretoria, South Africa.
- Frauenstein, E. D., & Flowerday, S. (2016). *Social network phishing: Becoming habituated to clicks and ignorant to threats?* Paper presented at the 15th Information Security for South Africa conference (ISSA 2016). Johannesburg, South Africa.
- Frauenstein, E. D., & Flowerday, S. (2020). Susceptibility to phishing on social network sites: A personality information processing model. *Computers & Security*, 94, 101862. <https://doi.org/10.1016/j.cose.2020.101862>
- Friendman, B. (2014). *A study of South African computer users' password usage habits and attitude towards password security* (Masters of Science). Rhodes University, Grahamstown, South Africa.
- Fu, Q., Feng, B., Guo, D., & Li, Q. (2018). Combating the evolving spammers in online social networks. *Computers & Security*, 72, 60–73.
<https://doi.org/10.1016/j.cose.2017.08.014>
- Funder, D. C. (2001). Personality. *Annual Review of Psychology*, 52(1), 197–221.
<https://doi.org/10.1146/annurev.psych.52.1.197>
- Furnell, S. (2007). Phishing: can we spot the signs? *Computer Fraud & Security*, 2007(3), 10–15. [https://doi.org/10.1016/S1361-3723\(07\)70035-0](https://doi.org/10.1016/S1361-3723(07)70035-0)
- Furnell, S., Millet, K., & Papadaki, M. (2019). Fifteen years of phishing: can technology save us? *Computer Fraud & Security*, 2019(7), 11–16. [https://doi.org/10.1016/S1361-3723\(19\)30074-0](https://doi.org/10.1016/S1361-3723(19)30074-0)
- Furnham, A. (1996). The big five versus the big four: The relationship between the Myers-Briggs type indicator (MBTI) and NEO-PI five factor model of personality. *Personality and Individual Differences*, 21(2), 303–307.
[https://doi.org/10.1016/0191-8869\(96\)00033-5](https://doi.org/10.1016/0191-8869(96)00033-5)
- Gardikiotis, A., & Crano, W. D. (2015). Persuasion theories. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed.; pp. 941–947). Elsevier.

- Gardner, B. (2015). A review and analysis of the use of 'habit' in understanding, predicting and influencing health-related behaviour. *Health Psychology Review*, 9(3), 277–295. <https://doi.org/10.1080/17437199.2013.876238>
- Garera, S., Provos, N., Chew, M., & Rubin, A. D. (2007). *A framework for detection and measurement of phishing attacks*. Paper presented at the 2007 ACM workshop on Recurring malware, Alexandria, VA. <https://doi.org/10.1145/1314389.1314391>
- Garner, J., & O'Sullivan, H. (2010). Facebook and the professional behaviours of undergraduate medical students. *The Clinical Teacher*, 7(2), 112–115. <https://doi.org/10.1111/j.1743-498X.2010.00356.x>
- Gefen, D., Rigdon, E., & Straub, D. (2011). An update and extension to SEM guidelines for administrative and social science research. Editorial comment. *MIS Quarterly*, 35(2), III–XII. <https://doi.org/10.2307/23044042>
- Gkika, S., Skiada, M., Lekakos, G., & Kourouthanasis, P. E. (2016). *Investigating the role of personality traits and influence strategies on the persuasive effect of personalized recommendations*. Paper presented at the 4th Workshop on Emotions and Personality in Personalized Systems (EMPIRE 2016) co-located with ACM Conference on Recommender Systems (RecSys 2016), Boston, MA.
- Goel, S., Williams, K., & Dincelli, E. (2017). Got phished? Internet security and human vulnerability. *Journal of the Association for Information Systems*, 18, 22–44. <https://doi.org/10.17705/1jais.00447>
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4(1), 26–42. <https://doi.org/10.1037/1040-3590.4.1.26>
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1), 84–96. <https://doi.org/10.1016/j.jrp.2005.08.007>
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6), 504–528. [https://doi.org/10.1016/S0092-6566\(03\)00046-1](https://doi.org/10.1016/S0092-6566(03)00046-1)
- Gou, L., Zhou, M. X., & Yang, H. (2014). *KnowMe and ShareMe: Understanding automatically discovered personality traits from social media and user sharing preferences*. Paper presented at the SIGCHI conference on human factors in computing systems, Toronto, Ontario, Canada. <https://doi.org/10.1145/2556288.2557398>
- Grabner-Kräuter, S. (2009). Web 2.0 Social networks: The role of trust. *Journal of Business Ethics*, 90(4), 505–522. <https://doi.org/10.1007/s10551-010-0603-1>
- Gragg, D. (2003). A multi-level defense against social engineering. In: SANS institute - infosec reading room. <https://www.sans.org/reading-room/whitepapers/engineering/multi-level-defense-social-engineering-920>
- Grant, C., & Osanloo, A. (2015). Understanding, selecting, and integrating a theoretical framework in dissertation research: Developing a 'blueprint' for your "house". *Administrative Issues Journal*, 4. <https://doi.org/10.5929/2014.4.2.9>
- Gratian, M., Bandi, S., Cukier, M., Dykstra, J., & Ginther, A. (2018). Correlating human traits and cyber security behavior intentions. *Computers & Security*, 73, 345–358. <https://doi.org/10.1016/j.cose.2017.11.015>
- Grazioli, S. (2004). Where did they go wrong? An analysis of the failure of knowledgeable internet consumers to detect deception over the internet. *Group Decision and Negotiation*, 13(2), 149–172. <https://doi.org/10.1023/B:GRUP.0000021839.04093.5d>

- Griffin, R. J., Neuwirth, K., Giese, J., & Dunwoody, S. (2002). Linking the heuristic-systematic model and depth of processing. *Communication Research*, 29(6), 705–732. <https://doi.org/10.1177/009365002237833>
- Griffiths, M. (2003). Internet gambling: Issues, concerns, and recommendations. *CyberPsychology & Behavior*, 6(6), 557–568. <https://doi.org/10.1089/109493103322725333>
- Griffiths, M. D. (2012). Internet sex addiction: A review of empirical research. *Addiction Research & Theory*, 20(2), 111–124. <https://doi.org/10.3109/16066359.2011.588351>
- Gross, R., & Acquisti, A. (2005). *Information revelation and privacy in online social networks*. Paper presented at the 2005 ACM workshop on privacy in the electronic society, Alexandria, VA.
- Hadlington, L. (2017). Human factors in cybersecurity; examining the link between internet addiction, impulsivity, attitudes towards cybersecurity, and risky cybersecurity behaviours. *Heliyon*, 3(7). <https://doi.org/10.1016/j.heliyon.2017.e00346>
- Hair Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Prentice Hall.
- Hair Jr, J. F., Hult, G., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)* (2nd Ed.). Sage Publications.
- Halevi, T., Lewis, J., & Memon, N. (2013). *A pilot study of cyber security and privacy related behavior and personality traits*. Paper presented at the 22nd international conference on World Wide Web companion (WWW), Rio de Janeiro, Brazil.
- Ham, M., Jeger, M., & Frajman Ivković, A. (2015). The role of subjective norms in forming the intention to purchase green food. *Economic Research-Ekonomska Istraživanja*, 28(1), 738–748. <https://doi.org/10.1080/1331677X.2015.1083875>
- Hammer, C. (2011). The importance of participant demographics. *American Journal of Speech-Language Pathology*, 20(4), 261. [https://doi.org/10.1044/1058-0360\(2011/ed-04\)](https://doi.org/10.1044/1058-0360(2011/ed-04))
- Harris, A. L., & Yates, D. (2015). *Phishing attacks over time: A longitudinal study*. Paper presented at the 21st Americas Conference on Information Systems, Puerto Rico.
- Harrison, B., Vishwanath, A., Ng, Y. J., & Rao, R. (2015). *Examining the impact of presence on individual phishing victimization*. Paper presented at the 48th Hawaii International Conference on System Sciences (HICSS 2015), Hawaii, USA.
- Harrison, B., Vishwanath, A., & Rao, R. (2016). *A user-centered approach to phishing susceptibility: The role of a suspicious personality in protecting against phishing*. Paper presented at the 49th Hawaii International Conference on System Sciences (HICSS 2016), Hawaii, USA.
- Hassandoust, F., Singh, H., & Williams, J. (2020). The role of contextualization in individuals' vulnerability to phishing attempts. *Australasian Journal of Information Systems*, 24(0). <https://doi.org/10.3127/ajis.v24i0.2693>
- Hatfield, J. M. (2018). Social engineering in cybersecurity: The evolution of a concept. *Computers & Security*, 73, 102–113. <https://doi.org/10.1016/j.cose.2017.10.008>
- Heartfield, R., & Loukas, G. (2015). A taxonomy of attacks and a survey of defence mechanisms for semantic social engineering attacks. *ACM Computing Surveys*, 48(3), 1–39. <https://doi.org/10.1145/2835375>
- Heinström, J. (2003). Five personality dimensions and their influence on information behaviour. *Information Research*, 9(1). <http://InformationR.net/ir/9-1/paper165.html>
- Herath, T., Chen, R., Wang, J., Banjara, K., Wilbur, J., & Rao, H. R. (2014). Security services as coping mechanisms: An investigation into user intention to adopt an email authentication service. *Information Systems Journal*, 24(1), 61–84. <https://doi.org/10.1111/j.1365-2575.2012.00420.x>

- Herath, T., & D'Arcy, J. (2015). *Social networking behaviors: Role of personality, perceived risk, and social influences*. Paper presented at the International Conference on Information Resources Management (Conf-IRM 2015), Ottawa, Ontario, Canada.
- Herath, T., & Rao, H. R. (2009). Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. *Decision Support Systems*, 47(2), 154–165. <https://doi.org/10.1016/j.dss.2009.02.005>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- Hocevar, K. P., Flanagan, A. J., & Metzger, M. J. (2014). Social media self-efficacy and information evaluation online. *Computers in Human Behavior*, 39(C), 254–262. <https://doi.org/10.1016/j.chb.2014.07.020>
- Hofstee, E. (2006). *Constructing a good dissertation: A practical guide to finishing a masters, MBA or PhD on schedule*. EPE.
- 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. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 57(1), 1012–1016. <https://doi.org/10.1177/1541931213571226>
- Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53–60. <https://doi.org/10.21427/D7CF7R>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Huang, D.-L., Rau, P.-L. P., & Salvendy, G. (2007). *A Survey of factors influencing people's perception of information security*. Paper presented at the Human-Computer Interaction. HCI Applications and Services, Beijing, China.
- Huber, M., Kowalski, S., Nohlberg, M., & Tjoa, S. (2009). *Towards automating social engineering using social networking sites*. Paper presented at the 2009 International Conference on Computational Science and Engineering, Vancouver, Canada.
- Hughes, D. J., Rowe, M., Batey, M., & Lee, A. (2012). A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Computers in Human Behavior*, 28(2), 561–569. <https://doi.org/10.1016/j.chb.2011.11.001>
- Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., & Williams, L. J. (1997). Exploratory and Confirmatory Factor Analysis: Guidelines, issues, and alternatives. *Journal of Organizational Behavior*, 18(6), 667–683. [https://doi.org/10.1002/\(SICI\)1099-1379\(199711\)18:6<667::AID-JOB874>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1099-1379(199711)18:6<667::AID-JOB874>3.0.CO;2-T)
- Hussain, Z., Simonovic, B., Stupple, E. J. N., & Austin, M. (2019). Using eye tracking to explore Facebook use and associations with Facebook addiction, mental well-being, and personality. *Behavioral Sciences (Basel)*, 9(2). <https://doi.org/10.3390/bs9020019>
- Hwang, Y., & Kim, D. J. (2007). Customer self-service systems: The effects of perceived web quality with service contents on enjoyment, anxiety, and e-trust. *Decision Support Systems*, 43(3), 746–760. <https://doi.org/10.1016/j.dss.2006.12.008>
- Ifinedo, P. (2012). Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory. *Computers & Security*, 31(1), 83–95. <https://doi.org/10.1016/j.cose.2011.10.007>
- Imhof, M. (2010). Listening to voices and judging people. *International Journal of Listening*, 24(1), 19–33. <https://doi.org/10.1080/10904010903466295>

- Intiful, F. D., Oddam, E. G., Kretchy, I., & Quampah, J. (2019). Exploring the relationship between the big five personality characteristics and dietary habits among students in a Ghanaian university. *BMC Psychology*, 7(1), 10. <https://doi.org/10.1186/s40359-019-0286-z>
- Iuga, C., Nurse, J. R. C., & Erola, A. (2016). Baiting the hook: Factors impacting susceptibility to phishing attacks. *Human-centric Computing and Information Sciences*, 6(1), 8. <https://doi.org/10.1186/s13673-016-0065-2>
- Ivaturi, K., & Janczewski, L. (2011). *A taxonomy for social engineering attacks*. Paper presented at the International Conference on Information Resources Management (Conf-IRM), Seoul, South Korea.
- Jacoby, J., & Kaplan, L. B. (1972). *The components of perceived risk*. Paper presented at the 3rd Annual Conference of the Association for Consumer Research, Chicago, IL.
- Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Communications of the ACM*, 50(10), 94–100. <https://doi.org/10.1145/1290958.1290968>
- Jager, W. (2003). *Breaking 'bad habits': A dynamical perspective on habit formation and change*. Paper presented at the Human Decision Making and Environmental Perception. Understanding and Assisting Human Decision Making in Real-life Settings, University of Groningen, Netherlands.
- Jakobsson, M., & Ratkiewicz, J. (2006). *Designing ethical phishing experiments: A study of (ROT13) rOnl query features*. Paper presented at the 15th International conference on World Wide Web, Edinburgh, Scotland.
- James, T. L., Lowry, P. B., Wallace, L., & Warkentin, M. (2017). The effect of belongingness on obsessive-compulsive disorder in the use of online social networks. *Journal of Management Information Systems*, 34(2), 560–596. <https://doi.org/10.1080/07421222.2017.1334496>
- Jampen, D., Gür, G., Sutter, T., & Tellenbach, B. (2020). Don't click: Towards an effective anti-phishing training. A comparative literature review. *Human-centric Computing and Information Sciences*, 10(1), 1–41. <https://doi.org/10.1186/s13673-020-00237-7>
- Jang-Jaccard, J., & Nepal, S. (2014). A survey of emerging threats in cybersecurity. *Journal of Computer and System Sciences*, 80(5), 973–993. <https://doi.org/10.1016/j.jcss.2014.02.005>
- Jang, K. L., Livesley, W. J., & Vernon, P. A. (1996). Heritability of the big five personality dimensions and their facets: a twin study. *Journal of Personality*, 64(3), 577–591. <https://doi.org/10.1111/j.1467-6494.1996.tb00522.x>
- Jansen, J., & Van Schaik, P. (2018). Persuading end users to act cautiously online: A fear appeals study on phishing. *Information and Computer Security*, 26(3), 264–276. <https://doi.org/10.1108/ICS-03-2018-0038>
- Jansen, J., & Van Schaik, P. (2019). The design and evaluation of a theory-based intervention to promote security behaviour against phishing. *International Journal of Human-Computer Studies*, 123, 40–55. <https://doi.org/10.1016/j.ijhcs.2018.10.004>
- 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. <https://doi.org/10.1080/07421222.2017.1334499>
- Jian, Z., Zhang, W., Tian, L., Fan, W., & Zhong, Y. (2019). Self-deception reduces cognitive load: The role of involuntary conscious memory impairment. *Frontiers in Psychology*, 10, 1718–1718. <https://doi.org/10.3389/fpsyg.2019.01718>
- Jin, L., Joshi, J. B. D., & Anwar, M. (2013). Mutual-friend based attacks in social network systems. *Computers & Security*, 37, 15–30. <https://doi.org/10.1016/j.cose.2013.04.003>

- John, O. P., & Srivastava, S. (1999). The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed.; pp. 102–138). Guilford Press.
- John, S. (2013). *Antecedents and effects of computer self-efficacy on social networking adoption among Asian online users*, Paper presented at the 19th Americas Conference on Information Systems (AMCIS 2013), Chicago, IL.
- Johnson, P. E., Grazioli, S., Jamal, K., & Berryman, R. G. (2001). Detecting deception: Adversarial problem solving in a low base-rate world. *Cognitive Science*, 25(3), 355–392. https://doi.org/10.1207/s15516709cog2503_2
- Johnson, T. J., & Kaye, B. K. (2014). Credibility of social network sites for political information among politically interested internet users. *Journal of Computer-Mediated Communication*, 19(4), 957–974. <https://doi.org/10.1111/jcc4.12084>
- Johnston, A. C., Warkentin, M., McBride, M., & Carter, L. (2016). Dispositional and situational factors: Influences on information security policy violations. *European Journal of Information Systems*, 25(3), 231–251. <https://doi.org/10.1057/ejis.2015.15>
- Joiner, R., Brosnan, M., Duffield, J., Gavin, J., & Maras, P. (2007). The relationship between internet identification, internet anxiety and internet use. *Computers in Human Behavior*, 23(3), 1408–1420. <https://doi.org/10.1016/j.chb.2005.03.002>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 374(2065), <https://doi.org/10.1098/rsta.2015.0202>
- 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. <https://doi.org/10.1016/j.chb.2016.09.012>
- Kajzer, M., D'Arcy, J., Crowell, C. R., Striegel, A., & Van Bruggen, D. (2014). An exploratory investigation of message-person congruence in information security awareness campaigns. *Computers & Security*, 43, 64–76. <https://doi.org/10.1016/j.cose.2014.03.003>
- Kaptein, M., & Eckles, D. (2012). Heterogeneity in the effects of online persuasion. *Journal of Interactive Marketing*, 26(3), 176–188. <https://doi.org/10.1016/j.intmar.2012.02.002>
- Kaptein, M., Markopoulos, P., de Ruyter, B., & Aarts, E. (2009). *Can you be persuaded? Individual differences in susceptibility to persuasion*, Paper presented at the 13th International Conference on Human-Computer Interaction (INTERACT 2009), Uppsala, Sweden.
- Karaiskos, D., Tzavellas, E., Balta, G., & Paparrigopoulos, T. (2010). P02–232 - Social network addiction: A new clinical disorder? *European Psychiatry*, 25, 855. [https://doi.org/10.1016/S0924-9338\(10\)70846-4](https://doi.org/10.1016/S0924-9338(10)70846-4)
- Katz, E., Gurevitch, M., & Haas, H. (1973). On the use of the mass media for important things. *American Sociological Review*, 38(2), 164–181. <https://doi.org/10.2307/2094393>
- Kayış, A. R., Satici, S. A., Yilmaz, M. F., Şimşek, D., Ceyhan, E., & Bakioğlu, F. (2016). Big five-personality trait and internet addiction: A meta-analytic review. *Computers in Human Behavior*, 63, 35–40. <https://doi.org/10.1016/j.chb.2016.05.012>
- Kearney, W. D., & Kruger, H. A. (2016). Theorising on risk homeostasis in the context of information security behaviour. *Information and Computer Security*, 24(5), 496–513. <https://doi.org/10.1108/ICS-04-2016-0029>
- Kesan, J. P., & Hayes, C. (2016). Bugs in the market: Creating a legitimate, transparent, and vendor-focused market for software vulnerabilities. *The Arizona Law Review*, 58(3), 753–830. <https://doi.org/10.2139/ssrn.2739894>

- Khamis, S., Ang, L., & Welling, R. (2017). Self-branding, 'micro-celebrity' and the rise of social media influencers. *Celebrity Studies*, 8(2), 191–208.
<https://doi.org/10.1080/19392397.2016.1218292>
- Ki-Aries, D., & Faily, S. (2017). Persona-centred information security awareness. *Computers & Security*, 70, 663–674. <https://doi.org/10.1016/j.cose.2017.08.001>
- Kim, D., & Kim, H. (2015). *Performing clickjacking attacks in the wild: 99% are still vulnerable!* Paper presented at the 1st International Conference on Software Security and Assurance (ICSSA 2015), Suwon, South Korea.
- Kim, H.-S. (2016). What drives you to check in on Facebook? Motivations, privacy concerns, and mobile phone involvement for location-based information sharing. *Computers in Human Behavior*, 54, 397–406.
<https://doi.org/10.1016/j.chb.2015.08.016>
- Kim, S. J., & Hancock, J. T. (2015). Optimistic bias and Facebook use: Self–other discrepancies about potential risks and benefits of Facebook use. *Cyberpsychology, Behavior, and Social Networking*, 18(4), 214–220.
<https://doi.org/10.1089/cyber.2014.0656>
- Kircaburun, K., & Griffiths, M. D. (2018). Instagram addiction and the big five of personality: The mediating role of self-liking. *Journal of Behavioral Addictions*, 7(1), 158–170. <https://doi.org/10.1556/2006.7.2018.15>
- Kirlappos, I., & Sasse, M. A. (2012). Security education against phishing: A modest proposal for a major re-think. *IEEE Security and Privacy*, 10(2), 24–32.
<https://doi.org/10.1109/MSP.2011.179>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.
- Korukonda, A. R. (2007). Differences that do matter: A dialectic analysis of individual characteristics and personality dimensions contributing to computer anxiety. *Computers in Human Behavior*, 23(4), 1921–1942.
<https://doi.org/10.1016/j.chb.2006.02.003>
- Kothari, C. R. (2008). *Research methodology methods and techniques* (2nd rev. ed.). New Age International.
- Krasnova, H., Kolesnikova, E., & Günther, O. (2009). "It won't happen to me!" *Self-disclosure in online social networks*. Paper presented at the 15th Americas Conference on Information Systems (AMCIS 2009), Atlanta, Georgia.
- Krombholz, K., Hobel, H., Huber, M., & Weippl, E. (2015). Advanced social engineering attacks. *Journal of Information Security Applications*, 22(C), 113–122.
<https://doi.org/10.1016/j.jisa.2014.09.005>
- Krueger, R. F., South, S., Johnson, W., & Iacono, W. (2008). The heritability of personality is not always 50%: gene-environment interactions and correlations between personality and parenting. *Journal of Personality*, 76(6), 1485–1522.
<https://doi.org/10.1111/j.1467-6494.2008.00529.x>
- Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). University of Chicago Press.
- Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2010). Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology*, 10(2), 1–31.
<https://doi.org/10.1145/1754393.1754396>
- Kumaran, N., & Lugani, S. (2020). *Protecting businesses against cyber threats during COVID-19 and beyond*. <https://cloud.google.com/blog/products/identity-security/protecting-against-cyber-threats-during-covid-19-and-beyond>

- Kuss, D. J., & Griffiths, M. D. (2011). Online social networking and addiction: A review of the psychological literature. *International Journal of Environmental Research and Public Health*, 8(9), 3528–3552. <https://doi.org/10.3390/ijerph8093528>
- Kuss, D. J., & Griffiths, M. D. (2012). Internet gaming addiction: A systematic review of empirical research. *International Journal of Mental Health and Addiction*, 10(2), 278–296. <https://doi.org/10.1007/s11469-011-9318-5>
- Kuss, D. J., & Griffiths, M. D. (2017). Social networking sites and addiction: Ten lessons learned. *International Journal of Environmental Research and Public Health*, 14(3), 311. <https://doi.org/10.3390/ijerph14030311>
- La Barbera, D., La Paglia, F., & Valsavoia, R. (2009). Social network and addiction. *Studies in Health Technology and Informatics*, 144, 33–36. <https://doi.org/10.3233/978-1-60750-017-9-33>
- Landau, S., & Everitt, B. S. (2004). *A handbook of statistical analyses using SPSS*. Chapman & Hall/CRC Press.
- Langheinrich, M., & Karjoth, G. (2010). Social networking and the risk to companies and institutions. *Information Security Technical Report*, 15(2), 51–56. <https://doi.org/10.1016/j.istr.2010.09.001>
- LaRose, R. (2015). The psychology of interactive media habits. In S. S. Sundar (Ed.), *The handbook of the psychology of communication technology*. Wiley Online Library.
- LaRose, R., Kim, J.-H., & Peng, W. (2011). Social networking: Addictive, compulsive, problematic or just another media habit? In Z. Papacharissi (Ed.), *A networked self: Identity, community, and culture on social network sites* (pp. 59–81). Routledge.
- LaRose, R., Lin, C. A., & Eastin, M. S. (2003). Unregulated internet usage: Addiction, habit, or deficient self-regulation? *Media Psychology*, 5(3), 225–253. https://doi.org/10.1207/S1532785XMEP0503_01
- Lastdrager, E. E. (2014). Achieving a consensual definition of phishing based on a systematic review of the literature. *Crime Science*, 3(1), 1–10. <https://doi.org/10.1186/s40163-014-0009-y>
- 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, 103084. <https://doi.org/10.1016/j.apergo.2020.103084>
- Lawson, P., Zielinska, O., Pearson, C., & Mayhorn, C. B. (2017). Interaction of personality and persuasion tactics in email phishing attacks. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61(1), 1331–1333. <https://doi.org/10.1177/1541931213601815>
- Lawson, P. A., Crowson, A. D., & Mayhorn, C. B. (2018). *Baiting the hook: Exploring the interaction of personality and persuasion tactics in email phishing attacks*. Paper presented at the 20th Congress of the International Ergonomics Association (IEA 2018), Florence, Italy.
- Lee, A. R., Son, S.-M., & Kim, K. K. (2016). Information and communication technology overload and social networking service fatigue: A stress perspective. *Computers in Human Behavior*, 55, 51–61. <https://doi.org/10.1016/j.chb.2015.08.011>
- Lee, M.-C. (2009). Factors influencing the adoption of internet banking: An integration of TAM and TPB with perceived risk and perceived benefit. *Electronic Commerce Research and Applications*, 8(3), 130–141. <https://doi.org/10.1016/j.elerap.2008.11.006>
- Leedy, P., & Ormrod, J. (2016). *Practical research: Planning and design* (11th ed.). Pearson.

- Leow, S., & Wang, Z. (2019). You don't know me but can I be your friend? Accepting strangers as friends in Facebook. *Social Networking*, 8(1), 52–73. <https://doi.org/10.4236/sn.2019.81004>
- Levers, M.-J. D. (2013). Philosophical paradigms, grounded theory, and perspectives on emergence. *SAGE Open*, 3(4), 2158244013517243. <https://doi.org/10.1177/2158244013517243>
- Levine, T. R. (2014). Truth-default theory (TDT): A theory of human deception and deception detection. *Journal of Language and Social Psychology*, 33(4), 378–392. <https://doi.org/10.1177/0261927x14535916>
- Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science: The International Journal of an Emerging Transdiscipline*, 9(1), 181–212. <https://doi.org/10.28945/479>
- Li, Y., Huang, Z., Wu, Y. J., & Wang, Z. (2019). Exploring how personality affects privacy control behavior on social networking sites. *Frontiers in Psychology*, 10(1771). <https://doi.org/10.3389/fpsyg.2019.01771>
- Li, Y., & Siponen, M. (2011). *A call for research on home users' information security behaviour*. Paper presented at the Pacific Asia Conference on Information Systems (PACIS 2011), Brisbane, Queensland, Australia.
- Liao, C., Palvia, P., & Lin, H.-N. (2006). The roles of habit and web site quality in e-commerce. *International Journal of Information Management*, 26(6), 469–483. <https://doi.org/10.1016/j.ijinfomgt.2006.09.001>
- Lim, S., Saldanha, T., Malladi, S., & Melville, N. (2009). *Theories used in information systems research: Identifying theory networks in leading IS journals*. Paper presented at the International Conference on Information Systems (ICIS 2009), Phoenix, AZ.
- Limayem, M., Hirt, S. G., & Cheung, C. M. K. (2007). How habits limit the predictive power of intention: The case of information systems continuance. *MIS Quarterly*, 31(4), 705–737. <https://doi.org/10.2307/25148817>
- Lin, E., Greenberg, S., Trotter, E., Ma, D., & Aycocock, J. (2011). *Does domain highlighting help people identify phishing sites?* Paper presented at the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada.
- Lin, X., Spence, P. R., & Lachlan, K. A. (2016). Social media and credibility indicators: The effect of influence cues. *Computers in Human Behavior*, 63, 264–271. <https://doi.org/10.1016/j.chb.2016.05.002>
- Lönnqvist, J.-E., & Itkonen, J. V. A. (2016). Homogeneity of personal values and personality traits in Facebook social networks. *Journal of research in personality*, 60, 24–35. <https://doi.org/10.1016/j.jrp.2015.11.001>
- Lopes, B., & Yu, H. (2017). Who do you troll and why: An investigation into the relationship between the dark triad personalities and online trolling behaviours towards popular and less popular Facebook profiles. *Computers in Human Behavior*, 77, 69–76. <https://doi.org/10.1016/j.chb.2017.08.036>
- Loughlin, N. (2012). The benefits and disadvantages of post-positivism in international theory. https://www.e-ir.info/2012/01/20/what-are-the-benefits-and-disadvantages-of-post-positivism-for-international-theory/#_ftn9
- Lowry, P. B., & Gaskin, J. (2014). Partial least squares (PLS) structural equation modeling (SEM) for building and testing behavioral causal theory: When to choose it and how to use it. *IEEE Transactions on Professional Communication*, 57(2), 123–146. <https://doi.org/10.1109/TPC.2014.2312452>
- Luo, X., Brody, R., Seazzu, A., & Burd, S. (2011). Social engineering: The neglected human factor for information security management. *Information Resources Management Journal (IRMJ)*, 24(3), 1–8. <https://doi.org/10.4018/irmj.2011070101>

- Luo, X., 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. <https://doi.org/10.1016/j.cose.2012.12.003>
- Mancinelli, E., Bassi, G., & Salcuni, S. (2019). Predisposing and motivational factors related to social network sites use: Systematic review. *JMIR Formative Research*, 3(2), <https://doi.org/10.2196/12248>
- Mann, I. (2008). *Hacking the human - social engineering techniques and security countermeasures*, Routledge. <https://doi.org/10.4324/9781351156882>
- Mansfield-Devine, S. (2018). The ever-changing face of phishing. *Computer Fraud & Security*, 2018(11), 17–19. [https://doi.org/10.1016/S1361-3723\(18\)30111-8](https://doi.org/10.1016/S1361-3723(18)30111-8)
- Marakas, G., Johnson, R., & Clay, P. (2007). The evolving nature of the computer self-efficacy construct: An empirical investigation of measurement construction, validity, reliability and stability over time. *Journal of the Association for Information Systems*, 8(1), 16–46. <https://doi.org/10.17705/1jais.00112>
- March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, 15(4), 251–266. [https://doi.org/10.1016/0167-9236\(94\)00041-2](https://doi.org/10.1016/0167-9236(94)00041-2)
- Marchal, S., Armano, G., Gröndahl, T., Saari, K., Singh, N., & Asokan, N. (2017). Off-the-hook: An efficient and usable client-side phishing prevention application. *IEEE Transactions on Computers*, 66(10), 1717–1733. <https://doi.org/10.1109/TC.2017.2703808>
- Marino, C., Vieno, A., Pastore, M., Albery, I. P., Frings, D., & Spada, M. M. (2016). Modeling the contribution of personality, social identity and social norms to problematic Facebook use in adolescents. *Addictive Behaviors*, 63, 51–56. <https://doi.org/10.1016/j.addbeh.2016.07.001>
- Mashapa, J. (2013). *A model for managing user experience*. (Philosophiae Doctor: Information Technology). Nelson Mandela University, Port Elizabeth, South Africa.
- Masip, J., Alonso, H., Herrero, C., & Garrido, E. (2016). Experienced and novice officers' generalized communication suspicion and veracity judgments. *Law and Human Behavior*, 40(2), 169–181. <https://doi.org/10.1037/lhb0000169>
- Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50(4), 370–396. <https://doi.org/10.1037/h0054346>
- Mattke, J., Maier, C., Reis, L., & Weitzel, T. (2020). Herd behavior in social media: The role of Facebook likes, strength of ties, and expertise. *Information & Management*, 57(8), 103370. <https://doi.org/10.1016/j.im.2020.103370>
- Mayhorn, C. B., & Nyeste, P. G. (2012). Training users to counteract phishing. *Work*, 41 Suppl 1, 3549–3552. <https://doi.org/10.3233/wor-2012-1054-3549>
- Mayhorn, C. B., Welka, A. K., Zielinska, O. A., & Murphy-Hill, E. (2015). *Assessing individual differences in a phishing detection Task*. Paper presented at the 19th Triennial Congress of the IEA, Melbourne, Australia.
- McBride, M., Carter, L., & Warkentin, M. (2012). *Exploring the role of individual employee characteristics and personality on employee compliance with cyber security policies*. RTI International–Institute of Homeland Security Solutions, North Carolina.
- McCormac, A., Zwaans, T., Parsons, K., Calic, D., Butavicius, M., & Pattinson, M. (2017). Individual differences and information security awareness. *Computers in Human Behavior*, 69, 151–156. <https://doi.org/10.1016/j.chb.2016.11.065>
- McCrae, R. R., & Costa Jr, P. T. (1990). *Personality in adulthood*. Guilford Press.
- McCrae, R. R., & Costa Jr, P. T. (1999). A five-factor theory of personality. In *Handbook of personality: Theory and research*, 2nd ed. (pp. 139–153). Guilford Press.

- McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60(2), 175–215. <https://doi.org/10.1111/j.1467-6494.1992.tb00970.x>
- McCrae, R. R., & Terracciano, A. (2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology*, 88(3), 547–561. <https://doi.org/10.1037/0022-3514.88.3.547>
- McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7(1), 64–82. <https://doi.org/10.1037/1082-989X.7.1.64>
- McElroy, J. C., Hendrickson, A. R., Townsend, A. M., & DeMarie, S. M. (2007). Dispositional factors in internet use: Personality versus cognitive style. *MIS Quarterly*, 31(4), 809–820. <https://doi.org/10.2307/25148821>
- McGaghie, W., Bordage, G., Crandall, S., & Pangaro, L. (2001). Research design. *Academic Medicine*, 76, 929–930. <https://doi.org/10.1097/00001888-200109000-00024>
- Mehdizadeh, S. (2010). Self-presentation 2.0: Narcissism and self-esteem on Facebook. *Cyberpsychology, Behavior, and Social Networking*, 13(4), 357–364. <https://doi.org/10.1089/cyber.2009.0257>
- Mertens, D. M. (2010). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods* (3rd ed.). SAGE Publications.
- Mesch, G. S. (2012). Is online trust and trust in social institutions associated with online disclosure of identifiable information online? *Computers in Human Behavior*, 28(4), 1471–1477. <https://doi.org/10.1016/j.chb.2012.03.010>
- Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, 59, 210–220. <https://doi.org/10.1016/j.pragma.2013.07.012>
- Miltenberger, R. G. (2016). *Behavior modification: Principles and procedures* (6th ed.). Cengage Learning.
- Mischel, W. (2009). From Personality and Assessment (1968) to Personality Science, 2009. *Journal of Research in Personality*, 43(2), 282–290. <https://doi.org/10.1016/j.jrp.2008.12.037>
- Mitchell, P. (2000). Internet addiction: Genuine diagnosis or not? *The Lancet*, 355(9204), 632. [https://doi.org/10.1016/S0140-6736\(05\)72500-9](https://doi.org/10.1016/S0140-6736(05)72500-9)
- Mitnick, K. D., & Simon, W. L. (2002). *The art of deception: Controlling the human element of security*. Wiley.
- Modic, D., & Lea, S. E. G. (2012). How neurotic are scam victims, really? The big five and internet scams. *Law & Humanities eJournal*. <https://doi.org/10.2139/ssrn.2448130>
- Mohebzada, J. G., Zarka, A. E., Bhojani, A. H., & Darwish, A. (2012). *Phishing in a university community: Two large scale phishing experiments*. Paper presented at the International Conference on Innovations in Information Technology (IIT 2012), Abu Dhabi, UAE.
- Montag, C., Sindermann, C., Lester, D., & Davis, K. L. (2020). Linking individual differences in satisfaction with each of Maslow's needs to the big five personality traits and Panksepp's primary emotional systems. *Heliyon*, 6(7), e04325. <https://doi.org/10.1016/j.heliyon.2020.e04325>
- 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. <https://doi.org/10.1057/s41303-017-0058-x>

- Moore, K., & McElroy, J. C. (2012). The influence of personality on Facebook usage, wall postings, and regret. *Computers in Human Behavior*, 28(1), 267–274.
<https://doi.org/10.1016/j.chb.2011.09.009>
- 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.
<https://doi.org/10.1016/j.chb.2016.12.044>
- Morgan, D. L. (2014). Pragmatism as a paradigm for social research. *Qualitative Inquiry*, 20(8), 1045–1053. <https://doi.org/10.1177/1077800413513733>
- Morgan, G., & Smircich, L. (1980). The case for qualitative research. *Academy of Management Review*, 5(4), 491–500. <https://doi.org/10.2307/257453>
- Mouakket, S. (2015). Factors influencing continuance intention to use social network sites: The Facebook case. *Computers in Human Behavior*, 53, 102–110.
<https://doi.org/10.1016/j.chb.2015.06.045>
- Mouton, F., Leenen, L., Malan, M. M., & Venter, H. S. (2014). *Towards an ontological model defining the social engineering domain*. Paper presented at the 11th Human Choice and Computers International Conference (HC11), Turku, Finland.
- Mouton, F., Leenen, L., & Venter, H. S. (2016). Social engineering attack examples, templates and scenarios. *Computers & Security*, 59, 186–209.
<https://doi.org/10.1016/j.cose.2016.03.004>
- Mouton, F., Malan, M. M., Kimppa, K. K., & Venter, H. S. (2015). Necessity for ethics in social engineering research. *Computers & Security*, 55, 114–127.
<https://doi.org/10.1016/j.cose.2015.09.001>
- Mouton, F., Malan, M. M., & Venter, H. S. (2013). *Social engineering from a normative ethics perspective*. Paper presented at the 12th Information Security for South Africa conference (ISSA 2013), Johannesburg, South Africa.
- Muncaster, P. (2020). #COVID19 Drives phishing emails up 667% in under a month.
<https://www.infosecurity-magazine.com/news/covid19-drive-phishing-emails-667/>
- Myers, I. B., McCaulley, M. H., Quenk, N. L., & Hammer, A. L. (1998). *MBTI manual: A guide to the development and use of the Myers-Briggs type indicator* (3rd ed.). Consulting Psychologists.
- Nilsson, J. (2009). Segmenting socially responsible mutual fund investors: The influence of financial return and social responsibility. *International Journal of Bank Marketing*, 27, 5–31. <https://doi.org/10.1108/02652320910928218>
- Nord, J. H., Koochang, A., Floyd, K., & Paliszkievicz, J. (2020). Impact of habits on information security policy compliance. *Issues in Information Systems*, 21(3), 217–226.
- Nurse, J. R. C., Buckley, O., Legg, P. A., Goldsmith, M., Creese, S., Wright, G. R. T., & Whitty, M. (2014). *Understanding insider threat: A framework for characterising attacks*. Paper presented at the 2014 IEEE Security and Privacy Workshops, San Jose, CA, USA.
- Nyoni, P., & Velepini, M. (2018). Privacy and user awareness on Facebook. *South African Journal of Science*, 114(5-6). <https://doi.org/10.17159/sajs.2018/20170103>
- Oeldorf-Hirsch, A., & Sundar, S. S. (2015). Posting, commenting, and tagging: Effects of sharing news stories on Facebook. *Computers in Human Behavior*, 44, 240–249.
<https://doi.org/10.1016/j.chb.2014.11.024>
- Oest, A., Safei, Y., Doupe, A., Ahn, G., Wardman, B., & Warner, G. (2018). *Inside a phisher's mind: Understanding the anti-phishing ecosystem through phishing kit analysis*, Paper presented at the 13th Symposium on Electronic Crime Research (eCrime 2018), San Diego, CA.

- 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*. Paper presented at the 2017 CHI Conference on Human Factors in Computing Systems, Denver, Colorado, USA.
- Ollmann, G. (2002). *The phishing guide understanding & preventing phishing attacks*. IBM Internet Security Systems. https://www.nccgroup.com/globalassets/our-research/uk/whitepapers/the_phishing_guide_understanding_preventing_phishing_attacks.pdf
- Ophoff, J., & Robinson, M. (2014). *Exploring end-user smartphone security awareness within a South African context*. Paper presented at the 13th Information Security for South Africa conference (ISSA 2014), Johannesburg, South Africa.
- Oyibo, K., Orji, R., & Vassileva, J. (2017). *Investigation of the influence of personality traits on Cialdini's persuasive strategies*. Paper presented at the Personalization in Persuasive Technology Workshop, Persuasive Technology 2017, Amsterdam, Netherlands.
- Oyibo, K., & Vassileva, J. (2019). The relationship between personality traits and susceptibility to social influence. *Computers in Human Behavior*, 98, 174–188. <https://doi.org/10.1016/j.chb.2019.01.032>
- Özbek, V., Alnaçık, Ü., Koc, F., Akkılıç, M. E., & Kaş, E. (2014). The Impact of personality on technology acceptance: A study on smart phone users. *Procedia - Social and Behavioral Sciences*, 150, 541–551. <https://doi.org/10.1016/j.sbspro.2014.09.073>
- Panhwar, A. H., Ansari, S., & Shah, A. (2017). Post-positivism: An effective paradigm for social and educational research. *International Research Journal of Art & Humanities*, 45(45), 253–259.
- Pansiri, J. (2009). Evolution of a doctoral thesis research topic and methodology: A personal experience. *Tourism Management*, 30(1), 83–89. <https://doi.org/10.1016/j.tourman.2008.04.001>
- Park, S. (2018). Effects of heuristic-systematic information processing about the flu and the flu vaccination. *Social Sciences*, 7(6), 260–267. <https://doi.org/10.11648/j.ss.20180706.13>
- Parker, H. J., & Flowerday, S. V. (2020). Contributing factors to increased susceptibility to social media phishing attacks. *South African Journal of Information Management*, 22(1), 1–10. <https://doi.org/10.4102/sajim.v22i1.1176>
- Parrish Jr, J. L., Bailey, J. L., & Courtney, J. F. (2009). A Personality Based Model for Determining Susceptibility to Phishing Attacks. *Little Rock: University of Arkansas*, 285–296.
- 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. <https://doi.org/10.1016/j.ijhcs.2019.02.007>
- Parsons, K., McCormac, A., Butavicius, M., & Ferguson, L. (2011). *Human factors and information security: Individual, culture and security environment executive summary*, Command Control Communications and Intelligence Division (C3ID) Defence Science and Technology Organization (DSTO), Edinburgh, Australia.
- Patomäki, H., & Wight, C. (2000). After postpositivism? The promises of critical realism. *International Studies Quarterly*, 44(2), 213–237. <https://doi.org/10.1111/0020-8833.00156>
- Pattinson, M., & Anderson, G. (2005). Risk communication, risk perception and information security. In P. Dowland, S. Furnell, B. Thuraisingham, & X. S. Wang (Eds.), *Security management, integrity, and internal control in information systems*. IICIS 2004. IFIP

- International Federation for Information Processing, vol 193. Springer.
https://doi.org/10.1007/0-387-31167-X_11
- Pattinson, M., Jerram, C., Parsons, K., McCormac, A., & Butavicius, M. (2012). Why do some people manage phishing e-mails better than others? *Information Management & Computer Security*, 20(1), 18–28. <https://doi.org/10.1108/09685221211219173>
- Pavlou, P. A. (2003). Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. *International Journal of Electronic Commerce*, 7(3), 101–134. <https://doi.org/10.1080/10864415.2003.11044275>
- Pavlou, P. A., Liang, H., & Xue, Y. (2007). Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective. *MIS Quarterly*, 31(1), 105–136. <https://doi.org/10.2307/25148783>
- Peluchette, J. V., Karl, K., Wood, C., & Williams, J. (2015). Cyberbullying victimization: Do victims' personality and risky social network behaviors contribute to the problem? *Computers in Human Behavior*, 52, 424–435.
<https://doi.org/10.1016/j.chb.2015.06.028>
- Petelka, J., Zou, Y., & Schaub, F. (2019). *Put your warning where your link is: Improving and evaluating email phishing warnings*. Paper presented at the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland.
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion central and peripheral routes to attitude change. In *Communication and Persuasion* (pp. 1–24). Springer Verlag.
- Pfeiffer, T., Kauer, M., & Röth, J. (2014). “A bank would never write that!” A qualitative study on e-mail trust decisions. Paper presented at the annual conference of the Gesellschaft für Informatik (GI), Stuttgart, Germany.
- Pfleger, S. L., Sasse, M. A., & Furnham, A. (2014). From weakest link to security hero: Transforming staff security behavior. *Journal of Homeland Security and Emergency Management*, 11(4), 489–510. <https://doi.org/10.1515/jhsem-2014-0035>
- PhishLabs (2019). *2019 Phishing trends and intelligence report*.
<https://info.phishlabs.com/hubfs/2019%20PTI%20Report/2019%20Phishing%20Trends%20and%20Intelligence%20Report.pdf>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903.
<https://doi.org/10.1037/0021-9010.88.5.879>
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63(1), 539–569. <https://doi.org/10.1146/annurev-psych-120710-100452>
- Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, 12(4), 531–544.
<https://doi.org/10.1177/014920638601200408>
- Possemato, A., Lanzi, A., Chung, S. P. H., Lee, W., & Fratantonio, Y. (2018). *ClickShield: Are you hiding something? Towards eradicating clickjacking on Android*. Paper presented at the 2018 ACM SIGSAC Conference on Computer and Communications Security, Toronto, Canada.
- Potosky, D. (2002). A field study of computer efficacy beliefs as an outcome of training: the role of computer playfulness, computer knowledge, and performance during training. *Computers in Human Behavior*, 18(3), 241–255. [https://doi.org/10.1016/S0747-5632\(01\)00050-4](https://doi.org/10.1016/S0747-5632(01)00050-4)

- Purkait, S. (2015). Examining the effectiveness of phishing filters against DNS based phishing attacks. *Information & Computer Security*, 23(3), 333–346. <https://doi.org/10.1108/ICS-02-2013-0009>
- Quan-Haase, A., & Young, A. L. (2010). Uses and gratifications of social media: A comparison of Facebook and instant messaging. *Bulletin of Science, Technology & Society*, 30(5), 350–361. <https://doi.org/10.1177/0270467610380009>
- Quercia, D., Lambiotte, R., Stillwell, D., Kosinski, M., & Crowcroft, J. (2012). *The personality of popular facebook users*. Paper presented at the ACM 2012 conference on Computer Supported Cooperative Work, Seattle, WA.
- Raffetseder, T., Kirda, E., & Kruegel, C. (2007). *Building anti-phishing browser plug-ins: An experience report*. Paper presented at the 3rd International Workshop on Software Engineering for Secure Systems (SESS'07: ICSE Workshops 2007), Minneapolis, MN, USA.
- Rajesh, T., & Rangaiah, D. B. (2020). Facebook addiction and personality. *Heliyon*, 6(1), e03184. <https://doi.org/10.1016/j.heliyon.2020.e03184>
- Rathore, S., Sharma, P. K., Loia, V., Jeong, Y.-S., & Park, J. H. (2017). Social network security: Issues, challenges, threats, and solutions. *Information Sciences*, 421, 43–69. <https://doi.org/10.1016/j.ins.2017.08.063>
- Rehman, Z. U., Baharun, R., & Salleh, N. Z. M. (2020). Antecedents, consequences, and reducers of perceived risk in social media: A systematic literature review and directions for further research. *Psychology & Marketing*, 37(1), 74–86. <https://doi.org/10.1002/mar.21281>
- Reznik, M. (2013). Identity theft on social networking sites: Developing issues of internet impersonation. *Touro Law Review*, 29(2), 455–483. <https://digitalcommons.tourolaw.edu/lawreview/vol29/iss2/12>
- Rigdon, E. E., Sarstedt, M., & Ringle, C. M. (2017). On comparing results from CB-SEM and PLS-SEM: Five perspectives and five recommendations. *Marketing ZFP*, 39(3), 4–16. <https://doi.org/10.15358/0344-1369-2017-3-4>
- Robbins, T. W., & Costa, R. M. (2017). Habits. *Current Biology*, 27(22), R1200–R1206. <https://doi.org/10.1016/j.cub.2017.09.060>
- Robinson, J. (2010). *Triandis' theory of interpersonal behaviour in understanding software piracy behavior in the South African context*. (Masters degree in Industrial Psychology). University of the Witwatersrand, Johannesburg. <http://hdl.handle.net/10539/8377>
- Roby, K. (2019). Famous con man Frank Abagnale: Crime is 4,000 times easier today. <https://www.techrepublic.com/article/famous-con-man-frank-abagnale-crime-is-4000-times-easier-today/?ftag=TRE20d3f17&bhid=64978360>
- Rogers, R. W. (1975). A protection motivation theory of fear appeals and attitude change. *The Journal of Psychology*, 91(1), 93–114. <https://doi.org/10.1080/00223980.1975.9915803>
- Rolland, J. P. (2002). The cross-cultural generalizability of the five factor model of personality. In R. R. McCrae & J. Allik (Eds.), *The five factor model of personality across cultures* (pp. 7–28). Kluwer Academic/Plenum Publishers.
- Ross, C., Orr, E. S., Sisic, M., Arseneault, J. M., Simmering, M. G., & Orr, R. R. (2009). Personality and motivations associated with Facebook use. *Computers in Human Behavior*, 25(2), 578–586. <https://doi.org/10.1016/j.chb.2008.12.024>
- Rößling, G., & Müller, M. (2009). *Social engineering: A serious underestimated problem*. Paper presented at the 14th annual ACM SIGCSE conference on Innovation and technology in computer science education, Paris, France.

- Russell, D. (2014). Aristotelian Virtue Theory: After the Person-Situation Debate. *Revue internationale de philosophie*, 1(1), 37-63. <https://doi.org/10.3917/rip.267.0037>.
- Ryan, A. B. (2006). Post-positivist approaches to research. In M. Antonesa, H. Fallon, A. B. Ryan, A. Ryan, T. Walsh, & L. Borys (Eds.), *Researching and writing your thesis: A guide for postgraduate students* (pp. 12–26). Maynooth Adult and Community Education (MACE).
- Ryan, T., & Xenos, S. (2011). Who uses Facebook? An investigation into the relationship between the big five, shyness, narcissism, loneliness, and Facebook usage. *Computers in Human Behavior*, 27(5), 1658–1664. <https://doi.org/10.1016/j.chb.2011.02.004>
- Safa, N. S., Sookhak, M., Von Solms, R., Furnell, S., Ghani, N. A., & Herawan, T. (2015). Information security conscious care behaviour formation in organizations. *Computers & Security*, 53, 65–78. <https://doi.org/10.1016/j.cose.2015.05.012>
- Sagarin, B. J., & Cialdini, R. B. (2004). Creating critical consumers: Motivating receptivity by teaching resistance. In *Resistance and persuasion*. (pp. 259–282). Lawrence Erlbaum Associates.
- Saleem, H., Beaudry, A., & Croteau, A.-M. (2011). Antecedents of computer self-efficacy: A study of the role of personality traits and gender. *Computers in Human Behavior*, 27(5), 1922–1936. <https://doi.org/10.1016/j.chb.2011.04.017>
- Salgado, J. F. (2002). The big five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment*, 10(1-2), 117–125. <https://doi.org/10.1111/1468-2389.00198>
- Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar big-five markers. *Journal of Personality Assessment*, 63(3), 506–516. https://doi.org/10.1207/s15327752jpa6303_8
- Saunders, M., Lewis, P., & Thornhill, A. (2016). *Research methods for business students* (7th ed.). Pearson Education.
- Savalei, V., & Bentler, P. (2006). Structural equation modeling. In R. Grover & M. Vriens (Eds.), *The handbook of marketing research: Uses, misuses, and future advances*. <https://doi.org/10.4135/9781412973380>
- Schneier, B. (2016). Stop trying to fix the user. *IEEE Security & Privacy*, 14(5), 96–96. <https://doi.org/10.1109/MSP.2016.101>
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323–338. <https://doi.org/10.3200/JOER.99.6.323-338>
- Schuetz, S., Lowry, P., Pienta, D., & Thatcher, J. (2020). The effectiveness of abstract versus concrete fear appeals in information security. *Journal of Management Information Systems*, 37, 723–757. <https://doi.org/10.1080/07421222.2020.1790187>
- Schuetz, S. W., Lowry, P. B., & Thatcher, J. B. (2016). *Defending against spear-phishing : Motivating users through fear appeal manipulations*. Paper presented at the 20th Pacific Asia Conference on Information Systems (PACIS 2016), Chiayi, Taiwan.
- Shan, T. L., Samy, G. N., Shanmugam, B., Azam, S., Yeo, K. C., & Kannoorpatti, K. (2016). *Heuristic systematic model based guidelines for phishing victims*. Paper presented at the 2016 IEEE Annual India Conference (INDICON). Bangalore, India.
- Sharot, T. (2011). The optimism bias. *Current Biology*, 21(23), R941–R945. <https://doi.org/10.1016/j.cub.2011.10.030>
- Sheeran, P. (2002). Intention-behavior relations: A conceptual and empirical review. *European Review of Social Psychology*, 12(1), 1–36. <https://doi.org/10.1080/14792772143000003>

- 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*. Paper presented at the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, GA.
- Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007). *Anti-phishing phil: The design and evaluation of a game that teaches people not to fall for phish*. Paper presented at the 3rd symposium on Usable privacy and security, Pittsburgh, PA.
- Shropshire, J., Warkentin, M., Johnston, A., & Schmidt, M. B. (2006). *Personality and IT security: An application of the five-factor model*. Paper presented at the 12th Americas Conference on Information Systems (AMCIS 2006), Acapulco, México.
- Shropshire, J., Warkentin, M., & Sharma, S. (2015). Personality, attitudes, and intentions: Predicting initial adoption of information security behavior. *Computers & Security*, *49*, 177–191. <https://doi.org/10.1016/j.cose.2015.01.002>
- Schumacker, R., & Lomax, R. (2016). *A beginner's guide to structural equation modeling* (4th ed.). Routledge.
- Shappie, A. T., Dawson, C. A., & Debb, S. M. (2020). Personality as a predictor of cybersecurity behavior. *Psychology of Popular Media*, *9*(4), 475–480. <https://doi.org/10.1037/ppm0000247>
- Siadati, H., Nguyen, T., Gupta, P., Jakobsson, M., & Memon, N. (2017). Mind your SMSes: Mitigating social engineering in second factor authentication. *Computers & Security*, *65*, 14–28. <https://doi.org/10.1016/j.cose.2016.09.009>
- Siegle, D. (2010). Cyberbullying and sexting: Technology abuses of the 21st Century. *Gifted Child Today*, *33*(2), 14–65. <https://doi.org/10.1177/107621751003300206>
- Silic, M., & Back, A. (2016). The dark side of social networking sites: Understanding phishing risks. *Computers in Human Behavior*, *60*, 35–43. <https://doi.org/10.1016/j.chb.2016.02.050>
- Song, Q., Wang, Y., Chen, Y., Benitez, J., & Hu, J. (2019). Impact of the usage of social media in the workplace on team and employee performance. *Information & Management*, *56*(8), 103160. <https://doi.org/10.1016/j.im.2019.04.003>
- Sophos. (2021). *Threat report*. <https://www.sophos.com/en-us/medialibrary/PDFs/technical-papers/sophos-2021-threat-report.pdf>
- Soror, A., Steelman, Z. R., & Turel, O. (2021). Exhaustion and dependency: A habituation–sensitization perspective on the duality of habit in social media use. *Information Technology & People, ahead-of-print* (ahead-of-print). <https://doi.org/10.1108/ITP-11-2019-0603>
- Soto, C. J., & John, O. P. (2017). The next big five inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, *113*(1), 117–143. <https://doi.org/10.1037/pspp0000096>
- Spottswood, E. L., & Hancock, J. T. (2017). Should I share that? Prompting social norms that influence privacy behaviors on a social networking site. *Journal of Computer-Mediated Communication*, *22*(2), 55–70. <https://doi.org/10.1111/jcc4.12182>
- Srivastava, S., John, O. P., Gosling, S., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology*, *84*(5), 1041–1053. <https://doi.org/10.1037/0022-3514.84.5.1041>
- Stajano, F., & Wilson, P. (2011). Understanding scam victims: Seven principles for systems security. *Communications of the ACM*, *54*(3), 70–75. <https://doi.org/10.1145/1897852.1897872>

- Statista. (2020). Number of monthly active Facebook users worldwide as of 4th quarter 2020 (in millions). <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>
- Steinberg, J. (2017). *Why do Facebook hoaxes often ask you to copy, paste, and modify, rather than to share?* <https://www.inc.com/joseph-steinberg/why-do-facebook-hoaxes-often-ask-you-to-copy-paste-modify-rather-than-to-share.html>
- Sterrett, D., Malato, D., Benz, J., Kantor, L., Tompson, T., Rosenstiel, T., Sonderman, J., & Loker, K. (2019). Who shared it? Deciding what news to trust on social media. *Digital Journalism*, 7(6), 783–801. <https://doi.org/10.1080/21670811.2019.1623702>
- Stobert, E., & Biddle, R. (2016). *Expert password management*, Paper presented at 9th International Conference on Passwords, Cambridge, UK. https://doi.org/10.1007/978-3-319-29938-9_1
- Sudzina, F., & Pavlicek, A. (2017). *Propensity to click on suspicious links: Impact of gender, of age, and of personality traits*, Paper presented at 30th Bled eConference Digital Transformation, Bled, Slovenia.
- Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of Graduate Medical Education*, 4(3), 279–282. <https://doi.org/10.4300/jgme-d-12-00156.1>
- Sumner, C., Byers, A., & Shearing, M. (2011). *Determining personality traits & privacy concerns from Facebook activity*. Paper presented at the Black Hat Briefings, Abu Dhabi, UAE.
- Sun, J. C.-Y., Yu, S.-J., Lin, S. S. J., & Tseng, S.-S. (2016). The mediating effect of anti-phishing self-efficacy between college students' internet self-efficacy and anti-phishing behavior and gender difference. *Computers in Human Behavior*, 59, 249–257. <https://doi.org/10.1016/j.chb.2016.02.004>
- Sushama, C., Sunil Kumar, M., & Neelima, P. (2021). Privacy and security issues in the future: A social media. *Materials Today: Proceedings*. <https://doi.org/10.1016/j.matpr.2020.11.105>
- Swickert, R. J., Hittner, J. B., Harris, J. L., & Herring, J. A. (2002). Relationships among internet use, personality, and social support. *Computers in Human Behavior*, 18(4), 437–451. [https://doi.org/10.1016/S0747-5632\(01\)00054-1](https://doi.org/10.1016/S0747-5632(01)00054-1)
- Tacchini, E., Ballarin, G., Vedova, M. L. D., Moret, S., & de Alfaro, L. (2017). *Some like it hoax: Automated fake news detection in social networks*. Paper presented at the 2nd Workshop on Data Science for Social Good (SoGood 2017), Skopje, Macedonia.
- Tanaka, J. S. (1987). “How big is big enough?”: Sample size and goodness of fit in structural equation models with latent variables. *Child Development*, 58(1), 134–146. <https://doi.org/10.2307/1130296>
- Tang, J.-H., Chen, M.-C., Yang, C.-Y., Chung, T.-Y., & Lee, Y.-A. (2016). Personality traits, interpersonal relationships, online social support, and Facebook addiction. *Telematics and Informatics*, 33(1), 102–108. <https://doi.org/10.1016/j.tele.2015.06.003>
- Tanhovska, H. (2021). *Facebook access penetration 2021, by device*. <https://www.statista.com/statistics/377808/distribution-of-facebook-users-by-device/>
- Terracciano, A., Costa, P. T., & McCrae, R. R. (2006). Personality plasticity after age 30. *Personality & social psychology bulletin*, 32(8), 999–1009. <https://doi.org/10.1177/0146167206288599>
- Thadani, D., & Cheung, C. (2011). *Exploring the role of online social network dependency in habit formation*. Paper presented at the 32nd International Conference on Information Systems (ICIS), Shanghai, China.

- Thomas-Jones, A. (2010). 6 - You've been poked: bullying, harassment and everyday undercurrents. In A. Thomas-Jones (Ed.), *The Host in the Machine* (pp. 99–121): Chandos Publishing.
- Triandis, H. C. (1977). *Interpersonal behavior*. Brooks/Cole Publishers.
- Triandis, H. C. (1980). Values, attitudes, and interpersonal behavior. *Nebraska Symposium on Motivation*, 27, 195–259.
- Trivedi, S. D., Dave, D., & Sridaran, R. (2016). *Analysis and impact of cyber threats on online social networks*. Paper presented at the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India.
- Trochim, W. H. (2006). *The research methods knowledge base*.
<http://www.socialresearchmethods.net/kb>
- Trumbo, C. W. (2006). Information processing and risk perception: An adaptation of the heuristic-systematic model. *Journal of Communication*, 52(2), 367–382.
<https://doi.org/10.1111/j.1460-2466.2002.tb02550.x>
- Tsohou, A., Karyda, M., Kokolakis, S., & Kiountouzis, E. (2015). Managing the introduction of information security awareness programmes in organisations. *European Journal of Information Systems*, 24(1), 38–58. <https://doi.org/10.1057/ejis.2013.27>
- Turel, O., & Serenko, A. (2011). *Developing a (bad) habit: Antecedents and adverse consequences of social networking website use habit*. Paper presented at the 17th Americas Conference on Information Systems (AMCIS 2011). Detroit, Michigan, USA.
- Turel, O., & Serenko, A. (2012). The benefits and dangers of enjoyment with social networking websites. *European Journal of Information Systems*, 21(5), 512–528.
<https://doi.org/10.1057/ejis.2012.1>
- Uebelacker, S., & Quiel, S. (2014). *The social engineering personality framework*. Paper presented at the 2014 Workshop on Socio-Technical Aspects in Security and Trust, Vienna, Austria.
- Vaishnavi, V. K., & Kuechler, W. (2015). *Design science research methods and patterns innovating information and communication technology* (2nd ed.). CRC Press.
- Valecha, R., Chen, R., Herath, T., Vishwanath, A., Wang, J., & Rao, H. R. (2015). *An exploration of phishing information sharing: A heuristic-systematic approach*. Paper presented at the 2015 IEEE 9th International Symposium on Intelligent Signal Processing (WISP) Proceedings, Siena, Italy.
- van Bavel, R., Rodríguez-Priego, N., Vila, J., & Briggs, P. (2019). Using protection motivation theory in the design of nudges to improve online security behavior. *International Journal of Human-Computer Studies*, 123, 29–39.
<https://doi.org/10.1016/j.ijhcs.2018.11.003>
- van der Schyff, K., & Flowerday, S. (2019). Social media surveillance: A personality-driven behaviour model. *South African Journal of Information Management*, 21(1), 1–9.
<https://doi.org/10.4102/sajim.v21i1.1034>
- van der Schyff, K., Flowerday, S., Kruger, H., & Patel, N. (2020). Intensity of Facebook use: A personality-based perspective on dependency formation. *Behaviour & Information Technology*, 1–17. <https://doi.org/10.1080/0144929X.2020.1800095>
- van der Schyff, K., Flowerday, S., & Lowry, P. B. (2020). Information privacy behavior in the use of Facebook apps: A personality-based vulnerability assessment. *Heliyon*, 6(8), e04714. <https://doi.org/10.1016/j.heliyon.2020.e04714>
- van Schaik, P., Jeske, D., Onibokun, J., Coventry, L., Jansen, J., & Kusev, P. (2017). Risk perceptions of cyber-security and precautionary behaviour. *Computers in Human Behavior*, 75, 547–559. <https://doi.org/10.1016/j.chb.2017.05.038>

- van Schaik, P., Jansen, J., Onibokun, J., Camp, J., & Kusev, P. (2018). Security and privacy in online social networking: Risk perceptions and precautionary behaviour. *Computers in Human Behavior*, 78, 283–297. <https://doi.org/10.1016/j.chb.2017.10.007>
- Vance, A., Lowry, P. B., & Eggett, D. (2013). Using accountability to reduce access policy violations in information systems. *Journal of Management Information Systems*, 29(4), 263–290. <https://doi.org/10.2753/MIS0742-1222290410>
- Vance, A., Siponen, M., & Pahlila, S. (2012). Motivating IS security compliance: Insights from habit and protection motivation theory. *Information & Management*, 49(3), 190–198. <https://doi.org/10.1016/j.im.2012.04.002>
- Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2), 186–204. <https://doi.org/10.1287/mnsc.46.2.186.11926>
- Venkatesh, V., Thong, J., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1), 157–178. <https://doi.org/10.2307/41410412>
- Verizon. (2019). *2019 Data breach investigations report*. <https://enterprise.verizon.com/resources/reports/2019-data-breach-investigations-report.pdf>
- Verizon. (2020). *2020 Data breach investigations report*. <https://enterprise.verizon.com/resources/reports/2020-msi-report.pdf>
- Verkijika, S. F. (2019). “If you know what to do, will you take action to avoid mobile phishing attacks”: Self-efficacy, anticipated regret, and gender. *Computers in Human Behavior*, 101, 286–296. <https://doi.org/10.1016/j.chb.2019.07.034>
- Verplanken, B., & Aarts, H. (1999). Habit, attitude, and planned behaviour: Is habit an empty construct or an interesting case of goal-directed automaticity? *European Review of Social Psychology*, 10(1), 101–134. <https://doi.org/10.1080/14792779943000035>
- Verplanken, B., & Orbell, S. (2003). Reflections on past behavior: A self-report index of habit strength. *Journal of Applied Social Psychology*, 33(6), 1313–1330. <https://doi.org/10.1111/j.1559-1816.2003.tb01951.x>
- Verplanken, B., & Wood, W. (2006). Interventions to break and create consumer habits. *Journal of Public Policy & Marketing*, 25(1), 90–103. <https://doi.org/10.1509/jppm.25.1.90>
- Visentin, L. (2014). *Facebook wages war on click-bait*. <https://www.smh.com.au/technology/facebook-wages-war-on-clickbait-20140826-108dd8.html>
- Vishwanath, A. (2014). Diffusion of deception in social media: Social contagion effects and its antecedents. *Information Systems Frontiers*, 17(6), 1353–1367. <https://doi.org/10.1007/s10796-014-9509-2>
- Vishwanath, A. (2015a). Examining the distinct antecedents of e-mail habits and its influence on the outcomes of a phishing attack. *Journal of Computer-Mediated Communication*, 20(5), 570–584. <https://doi.org/10.1111/jcc4.12126>
- Vishwanath, A. (2015b). Habitual Facebook use and its impact on getting deceived on social media. *Journal of Computer-Mediated Communication*, 20(1), 83–98. <https://doi.org/10.1111/jcc4.12100>
- Vishwanath, A. (2016). Mobile device affordance: Explicating how smartphones influence the outcome of phishing attacks. *Computers in Human Behavior*, 63, 198–207. <https://doi.org/10.1016/j.chb.2016.05.035>
- Vishwanath, A. (2017). Getting phished on social media. *Decision Support Systems*, 103(C), 70–81. <https://doi.org/10.1016/j.dss.2017.09.004>

- Vishwanath, A., Harrison, B., & Ng, Y. J. (2018). Suspicion, cognition, and automaticity model of phishing susceptibility. *Communication Research*, 45(8), 1146–1166. <https://doi.org/10.1177/0093650215627483>
- 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. <https://doi.org/10.1016/j.dss.2011.03.002>
- Volkamer, M., Renaud, K., & Reinheimer, B. (2016). TORPEDO: TOoltip-poweRed phishing email detectiOn. In J. H. Hoepman & S. Katzenbeisser (Eds.), *ICT Systems Security and Privacy Protection* (pp. 161–175): Springer.
- Volkman, E. (2019a). *The rise in mobile phishing attacks*. <https://info.phishlabs.com/blog/rise-mobile-social-engineering-phishing-attacks>
- Volkman, E. (2019b). *Why social media is increasingly abused for phishing attacks*. <https://info.phishlabs.com/blog/how-social-media-is-abused-for-phishing-attacks>
- Waheed, H., Anjum, M., Rehman, M., & Khawaja, A. (2017). Investigation of user behavior on social networking sites. *PLoS ONE*, 12(2): e0169693. <https://doi.org/10.1371/journal.pone.0169693>
- Wall, H. J., Campbell, C. C., Kaye, L. K., Levy, A., & Bhullar, N. (2019). Personality profiles and persuasion: An exploratory study investigating the role of the big-5, type D personality and the dark triad on susceptibility to persuasion. *Personality and Individual Differences*, 139, 69–76. <https://doi.org/10.1016/j.paid.2018.11.003>
- Wall, J. D., & Warkentin, M. (2019). Perceived argument quality's effect on threat and coping appraisals in fear appeals: An experiment and exploration of realism check heuristics. *Information & Management*, 56(8), 103157. <https://doi.org/10.1016/j.im.2019.03.002>
- Wang, D., Xu, L., & Chan, H. C. (2015). Understanding the continuance use of social network sites: a computer self-efficacy perspective. *Behaviour & Information Technology*, 34(2), 204–216. <https://doi.org/10.1080/0144929X.2014.952778>
- Wang, J., Chen, R., & Rao, H. R. (2012). An exploration of the design features of phishing attacks. In H. R. Rao & S. Upadhyaya (Eds.), *Handbook in information systems* (Vol. 4). Emerald Group Publishing.
- Wang, J., Li, Y., & Rao, H. R. (2017). Coping responses in phishing detection: An investigation of antecedents and consequences. *Information Systems Research*, 28(2), 378–396. <https://doi.org/10.1287/isre.2016.0680>
- Wang, Y., Norcie, G., Komanduri, S., Acquisti, A., Leon, P. G., & Cranor, L. F. (2011). “I regretted the minute I pressed share”: A qualitative study of regrets on Facebook. Paper presented at the 7th Symposium on Usable Privacy and Security, Pittsburgh, Pennsylvania.
- Warkentin, M., McBride, M. E., Carter, L., & Johnston, A. C. (2012). *The role of individual characteristics on insider abuse intentions*. Paper presented at the 18th Americas Conference on Information Systems (AMCIS 2012), Seattle, Washington.
- Watkins, M. W. (2018). Exploratory Factor Analysis: A guide to best practice. *Journal of Black Psychology*, 44(3), 219–246. <https://doi.org/10.1177/0095798418771807>
- Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26(2), xiii–xxiii.
- Wehrli, S. (2008). Personality on social network sites: An Application of the five factor model. *Eth Zurich Sociology Working Papers*, 0.37–33.54.
- Wenyin, L., Huang, G., Xiaoyue, L., Min, Z., & Deng, X. (2005). *Detection of phishing webpages based on visual similarity*. Paper presented at the special interest tracks and posters of the 14th International conference on World Wide Web, Chiba, Japan.

- West, R. (2008). The psychology of security. *Communications of the ACM*, 51(4), 34–40. <https://doi.org/10.1145/1330311.1330320>
- West, R., Mayhorn, C. B., Hardee, J., & Mendel, J. (2009). The weakest link: a psychological perspective on why users make poor security decisions. In *Social and Human Elements of Information Security: Emerging Trends and Countermeasures* (pp. 43–60).
- Weston, R., & Gore, P. A. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist*, 34(5), 719–751. <https://doi.org/10.1177/0011000006286345>
- Wilcox, H., & Bhattacharya, M. (2015). *Countering social engineering through social media: An enterprise security perspective*. Paper presented at the 7th International Conference on Computational Collective Intelligence Technologies and Applications (ICCCI 2015), Madrid, Spain.
- Wilde, G. J. S. (1998). Risk homeostasis theory: an overview. *Injury Prevention*, 4(2), 89–91. <https://doi.org/10.1136/ip.4.2.89>
- Williams, C. (2007). Research methods. *Journal of Business & Economic Research (JBER)*, 5(3), 65–71. <https://doi.org/10.19030/jber.v5i3.2532>
- 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. <https://doi.org/10.1016/j.chb.2017.03.002>
- 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. <https://doi.org/10.1016/j.ijhcs.2018.06.004>
- Wilson, K., Fornasier, S., & White, K. M. (2010). Psychological predictors of young adults' use of social networking sites. *Cyberpsychology, Behavior, and Social Networking*, 13(2), 173–177. <https://doi.org/10.1089/cyber.2009.0094>
- Withers, K. L., Parrish, J. L., Terrell, S. R., & Ellis, T. J. (2017). *The relationship between the “dark triad” personality traits and deviant behavior on social networking sites*. Paper presented at the 23rd Americas Conference on Information Systems (AMCIS 2017), Boston, MA.
- Wood, W. (2017). Habit in personality and social psychology. *Personality and Social Psychology Review*, 21(4), 389–403. <https://doi.org/10.1177/1088868317720362>
- Wood, W., & Neal, D. (2007). A new look at habits and the habit-goal interface. *Psychological Review*, 114, 843–863. <https://doi.org/10.1037/0033-295X.114.4.843>
- Wood, W., Quinn, J. M., & Kashy, D. A. (2002). Habits in everyday life: Thought, emotion, and action. *Journal of Personality and Social Psychology*, 83(6), 1281–1297. <https://doi.org/10.1037/0022-3514.83.6.1281>
- Wood, W., & Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology*, 67, 289–314. <https://doi.org/10.1146/annurev-psych-122414-033417>
- Workman, M. (2007). Gaining access with social engineering: An empirical study of the threat. *Information Systems Security*, 16(6), 315–331. <https://doi.org/10.1080/10658980701788165>
- 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. <https://doi.org/10.1002/asi.20779>
- Woszczyński, A. B., & Whitman, M. E. (2004). The problem of common method variance in IS research. In W. Michael & W. Amy (Eds.), *The handbook of information systems research* (pp. 66–78). IGI Global.

- 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. <https://doi.org/10.2753/MIS0742-1222270111>
- Wu, L., Brandt, B., Du, X., & Bo, J. (2016). *Analysis of clickjacking attacks and an effective defense scheme for Android devices*. Paper presented at the 2016 IEEE Conference on Communications and Network Security (CNS), Philadelphia, PA.
- Wu, M., Miller, R. C., & Garfinkel, S. L. (2006). *Do security toolbars actually prevent phishing attacks?* Paper presented at the SIGCHI Conference on Human Factors in Computing Systems, Montreal, Quebec, Canada.
- Xu, F., Xue, B., & Warkentin, M. (2020). *How do Leader emotional displays influence employees' information security effort?* Paper presented at the 26th Americas Conference on Information Systems (AMCIS 2020).
- Xu, Z., & Zhang, W. (2012). Victimized by phishing: A heuristic-systematic perspective. *Journal of Internet Banking and Commerce*, 17(3).
- Yamane, T. (1967). *Statistics: An introductory analysis* (2nd ed.). Harper and Row.
- Yan, Y., Nie, J., Huang, L., Li, Z., Cao, Q., & Wei, Z. (2015). *Is your first impression reliable? Trustworthy analysis using facial traits in portraits*, Paper presented at the 21st International Conference on Multimedia Modeling, Sydney, Australia.
- Yang, J., Du, F., Qu, W., Gong, Z., & Sun, X. (2013). Effects of personality on risky driving behavior and accident involvement for Chinese drivers. *Traffic Injury Prevention*, 14(6), 565–571. <https://doi.org/10.1080/15389588.2012.748903>
- Yang, S., Wang, B., & Lu, Y. (2016). Exploring the dual outcomes of mobile social networking service enjoyment: The roles of social self-efficacy and habit. *Computers in Human Behavior*, 64, 486–496. <https://doi.org/10.1016/j.chb.2016.07.010>
- Yao, M. Z., Rice, R. E., & Wallis, K. (2007). Predicting user concerns about online privacy. *Journal of the American Society for Information Science and Technology*, 58(5), 710–722. <https://doi.org/10.1002/asi.20530>
- Yong, A., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology*, 9, 79–94. <https://doi.org/10.20982/tqmp.09.2.p079>
- Young, K. S. (1999). *Internet addiction: Symptoms, evaluation, and treatment innovations in clinical practice* (L. VandeCreek & T. L. Jackson Eds. Vol. 17). Professional Resource Press.
- Zajenkowski, M., & Fronczyk, K. (2020). How do narcissists perceive personality items? Measurement invariance of a big five scale across low and high narcissism groups. *Personality and Individual Differences*, 152, 109595. <https://doi.org/10.1016/j.paid.2019.109595>
- Zhang, L.-F. (2006). Thinking styles and the big five personality traits revisited. *Personality and Individual Differences*, 40(6), 1177–1187. <https://doi.org/10.1016/j.paid.2005.10.011>
- Zhang, W., Burd, S. D., Luo, X., & Seazzu, A. F. (2012). *How could I fall for that? Exploring phishing victimization with the heuristic-systematic model*. Paper presented at the 45th Hawaii International Conference on System Sciences, Hawaii, USA.
- Zhang, Y., Egelman, S., Cranor, L., & Hong, J. (2007). *Phinding phish: Evaluating anti-phishing tools*. Paper presented at the 14th Annual Network and Distributed System Security Symposium, San Diego, CA.
- Zhang, Y., Hong, J. I., & Cranor, L. F. (2007). *Cantina: A content-based approach to detecting phishing web sites*. Paper presented at the 16th International conference on World Wide Web, Banff, Alberta, Canada.

- Zhu, W. (2016). $p < 0.05, < 0.01, < 0.001, < 0.0001, < 0.00001, < 0.000001, \text{ or } < 0.0000001$
.... *Journal of sport and health science*, 5(1), 77–79.
<https://doi.org/10.1016/j.jshs.2016.01.019>
- Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2013). *Business research methods*
(9th ed.). South-Western Cengage Learning.

APPENDIX A: CONSTRUCT DESCRIPTIVE STATISTICS

Constructs	Item no.	Mean	SD	Loading	CR	AVE
Extraversion	1	3.76	1.18	0.522	0.70	0.502
	6	3.56	1.17	0.828		
	11	4.15	0.94	0.566		
	16	3.74	0.94	0.751		
	21	2.66	1.39	0.660		
	26	4.10	0.96	0.716		
	31	2.43	1.31	0.743		
	36	3.56	1.28	0.681		
Agreeableness	2	3.53	1.24	0.882	0.57	0.518
	7	4.48	1.02	0.793		
	12	4.31	0.91	0.835		
	17	4.35	0.99	0.697		
	22	3.84	1.11	0.806		
	27	3.04	1.34	0.796		
	32	4.39	0.89	0.671		
	37	3.95	1.38	0.749		
	42	4.25	0.91	0.632		
Conscientiousness	3	3.95	1.01	0.727	0.70	0.694
	8	2.95	1.30	0.659		
	13	4.31	0.91	0.597		
	18	3.59	1.29	0.562		
	23	3.33	1.39	0.629		
	28	4.17	0.99	0.754		
	33	4.15	0.79	0.677		
	38	3.84	1.07	0.740		
	43	2.83	1.43	0.796		
Neuroticism	4	2.02	1.20	0.865	0.73	0.572
	9	2.04	1.15	0.673		
	14	3.41	1.26	0.656		
	19	3.44	1.43	0.674		
	24	2.27	1.34	0.674		
	29	2.82	1.47	0.735		
	34	2.12	1.14	0.707		
	39	3.08	1.46	0.556		
Openness	5	4.09	0.96	0.599	0.72	0.904
	10	4.43	0.78	0.749		
	15	3.79	0.99	0.650		
	20	4.34	0.82	0.619		
	25	3.51	1.02	0.662		
	30	3.91	1.10	0.739		
	35	2.04	1.08	0.875		
	40	3.96	1.00	0.479*		

	41	2.59	1.25	0.779		
	44	3.38	1.35	0.733		
Systematic processing	S1_5	3.37	1.42	0.530	0.90	0.578
	S1_6	3.33	1.58	0.465*		
	S7_7	3.28	1.51	0.547		
	S8_5	3.36	1.45	0.675		
	S8_6	3.40	1.57	0.675		
	S8_7	3.20	1.49	0.720		
	S9_5	3.21	1.39	0.624		
	S9_6	3.13	1.55	0.575		
	S9_7	3.07	1.53	0.568		
	S10_5	3.11	1.60	0.590		
	S10_6	3.00	1.66	0.639		
	S10_7	2.90	1.62	0.610		
	S11_5	3.02	1.57	0.614		
	S11_6	2.64	1.56	0.624		
	S11_7	2.69	1.54	0.662		
	S12_5	2.99	1.55	0.633		
S12_6	2.83	1.56	0.638			
S12_7	2.84	1.56	0.698			
Heuristic processing	S7_1	1.78	1.41	-0.013*	0.84	0.482
	S7_2	3.86	1.27	0.484*		
	S7_3	3.01	1.58	0.527		
	S7_4	2.27	1.51	0.296*		
	S8_1	1.87	1.53	0.022*		
	S8_2	3.76	1.38	0.492*		
	S8_3	3.30	1.57	0.557		
	S8_4	2.44	1.51	0.284*		
	S9_1	1.84	1.47	0.045*		
	S9_2	3.65	1.38	0.527		
	S9_3	2.70	1.54	0.465*		
	S9_4	2.48	1.44	0.356*		
	S10_1	1.72	1.61	0.103*		
	S10_2	3.35	1.56	0.563		
	S10_3	2.73	1.62	0.577		
	S10_4	2.13	1.60	0.462*		
S11_1	1.81	1.58	0.032*			
S11_2	3.22	1.60	0.629			
S11_3	2.27	1.48	0.552			
S11_4	2.15	1.65	0.402*			
S12_1	1.73	1.58	-0.013*			
S12_2	3.27	1.56	0.638			
S12_3	2.36	1.52	0.584			
S12_4	2.33	1.54	0.397*			
Habit	1	3.43	1.28	0.443*	0.88	0.504
	2	2.74	1.33	0.622		
	3	2.78	1.31	0.652		
	4	2.27	1.26	0.561		
	5	2.33	1.36	0.663		
	6	2.36	1.21	0.361*		

	7	2.58	1.33	0.623		
	8	2.30	1.28	0.703		
	9	2.47	1.34	0.742		
	10	2.74	1.36	0.501		
	11	2.56	1.34	0.791		
	12	2.65	1.42	0.770		
Social norms	1	3.40	1.43	0.500	0.70	0.640
	2	4.45	1.04	0.021*		
	3	4.21	1.07	0.151*		
	4	3.60	1.34	0.632		
	5	2.89	1.47	0.843		
	6	2.61	1.46	0.651		
Perceived risk	1	3.05	1.62	0.723	0.73	0.613
	2	3.26	1.64	0.801		
	3	3.15	1.72	0.784		
	4	3.83	1.30	0.224*		
Computer self- efficacy	1	4.17	0.84	0.470*	0.79	0.671
	2	4.37	0.70	0.431*		
	3	4.07	0.90	0.544		
	4	3.61	0.96	0.580		
	5	4.37	0.79	0.301*		
	6	3.13	0.97	0.823		
	7	2.97	0.97	0.771		
	8	3.47	1.23	0.500		

*Note: *indicates the items with less than 0.5 factor loading which were dropped from the model*

APPENDIX B: MEASUREMENT ITEMS OF THE STUDY

BFI Personality Trait (John and Srivastava, 2009).		
Items measured (1= disagree strongly – 5 = agree strongly)		
Construct	Item no:	Description
Extraversion	1	Is talkative
	6	Is reserved (R)
	11	Is full of energy
	16	Generates a lot of enthusiasm
	21	Tends to be quiet (R)
	26	Has an assertive (i.e. confident) personality
	31	Is sometimes shy, inhibited (R)
	36	Is outgoing, sociable
Agreeableness	2	Tends to find fault with others (R)
	7	Is helpful and unselfish with others
	12	Starts quarrels (i.e. arguments) with others (R)
	17	Has a forgiving nature
	22	Is generally trusting
	27	Can be cold and aloof (i.e. distant) (R)
	32	Is considerate and kind to almost everyone
	37	Is sometimes rude to others (R)
Conscientiousness	42	Likes to cooperate with others
	3	Does a thorough job
	8	Can be somewhat careless (R)
	13	Is a reliable worker
	18	Tends to be disorganized (R)
	23	Tends to be lazy (R)
	28	Perseveres until the task is finished
	33	Does things efficiently
Neuroticism	38	Makes plans and follows through with them
	43	Is easily distracted (R)
	4	Is depressed, blue
	9	Is relaxed, handles stress well (R)
	14	Can be tense (i.e. nervous, anxious)
	19	Worries a lot
	24	Is emotionally stable, not easily upset (R)
	29	Can be moody
Openness	34	Remains calm in tense situations (R)
	39	Gets nervous easily
	5	Is original, comes up with new ideas
	10	Is curious about many different things
	15	Is ingenious (i.e. clever), a deep thinker
	20	Has an active imagination
	25	Is inventive

	30	Values artistic (i.e. beauty), aesthetic experiences
	35	Prefers work that is routine (i.e. procedure) (R)
	40	Likes to reflect, play with ideas
	41	Has few artistic interests (R)
	44	Is sophisticated in art, music, or literature

Note: (R) = denotes reverse scaled items

Habits (Verplanken & Orbell, 2003).	
Items measured (1= disagree strongly – 5 = agree strongly)	
When a friend posts a link (e.g. could contain a video, image, news article) on a social network site and requests me to share it, I click/open/share it as this is something:	
1.	I do frequently/often.
2.	I do automatically.
3.	I do without having to consciously remember.
4.	That makes me feel weird if I do not do it.
5.	I do without thinking.
6.	That would require effort not to do it.
7.	That belongs to my (daily, weekly, monthly) routine.
8.	I start doing before I realize I'm doing it.
9.	I would find hard not to do it.
10.	I have no need to think about doing.
11.	That's typically 'me'.
12.	I have been doing for a long time.

Information Processing (Griffin et al., 2002; Vishwanath et al., 2011).**Items measured (1= disagree strongly – 5 = agree strongly)****Your Facebook friend posts the image seen above on their/your timeline, select the action YOU would most likely take:**

Construct	Items
1. Heuristic	I skimmed (i.e. moved quickly) through the Facebook message.
2. Heuristic	I briefly looked at the sender/source of the message.
3. Heuristic	The message is attractive to me as I am interested in the benefits it has to offer.
4. Heuristic	I ignored the content in the message.
5. Systematic	I thought about the action I took based on what I saw in the Facebook message.
6. Systematic	I spent some time thinking about the request before I made my decision.
7. Systematic	I found myself making connections between the message request and what I have heard about on social networks requesting such information.

Social norms.**Items measured (1= disagree strongly – 5 = agree strongly)****When a friend posts a status update and asks me to also “share” it, I will consider sharing it based on:**

1. If it is my friend, then it is the friendly thing to do.
2. It depends on what it is that I must share.
3. If it is a topic of interest to me personally.
4. The post is very popular (i.e. trending) and I know others will also find it interesting.
5. If I can see many of my friends or others have also already liked it.
6. If it could get me noticed with some likes from my friends.

Computer Self-Efficacy.

Items measured (very poor, poor, average, good, excellent)

Please evaluate your abilities based on the following:

1. Using a computer to type a document (e.g. assignment, CV, report).
2. Using a web browser (e.g. Chrome, Explorer, Firefox) to search for information on the internet.
3. Using the features of an email client app (e.g. Gmail, Yahoo) to send/receive messages and download/upload attachments.
4. Identifying different file extensions (e.g. .docx, .xlsx, .pdf, .rar, .zip).
5. Using social network websites (e.g. Facebook, Twitter, Instagram) to post and interact with other users.
6. Checking the security settings of a website to determine if it can be trusted as safe/original.
7. Identifying safe web links/URLs.
8. Installing software on a desktop/laptop computer.

Perceived risk.

Items measured (1= disagree strongly – 5 = agree strongly)

To what extent do you agree/disagree with the following statements:

1. There is little risk in sharing posts which instruct you to share to your profile (e.g. share this post and R100 will be donated to a charity).
2. There is little risk in accepting friend requests from strangers, as I can remove them later if I want to.
3. There is little risk that I can be personally affected on social networking websites (e.g. losing money, damaged reputation, and identity theft).
4. I am able to protect myself against threats on social network websites as I have control of my account.

APPENDIX C: INFORMED CONSENT FORM

You are invited to participate in a research study that concerns investigating the behaviours of social network users. This study is conducted by Mr. Edwin Donald Frauenstein, a PhD student at Rhodes University. The purpose of this questionnaire is to collect information used to determine particular behaviours that might make users vulnerable to social engineering threats (i.e. phishing) on social networking sites. The main behavioural aspects that this study focuses on is the Big Five personality traits, habit, information processing, social norms, computer self-efficacy and perceived risk. The objective of this research study is to develop a theoretical model that can help identify users who might be susceptible to phishing on social networking sites.

Participation in this study is voluntary. Your identity is treated as **anonymous** and the answers that you provide, if you choose to participate in this study, will be kept completely **confidential** to the full extent of the law. There is no means to identify you as this questionnaire does not capture any personal identifiable information (outside answering the questions stated). Moreover, as this survey is conducted online, information pertaining to your computer's IP address and email address will not be collected from the survey. Ethical approval for this study was obtained from the university's Ethics committee with certificate reference number: FLO071SFRA01.

In order to participate in this survey, it is a requirement that you must be an active social media user (i.e. Facebook, Twitter, LinkedIn). Kindly answer the questions in this questionnaire as carefully and as honestly as possible based on your own experiences. Participating in this study may not benefit you directly, but will greatly benefit social networking users and organisations at large in order to develop better interventions to protect social media users from information security threats. Any possible future publications emanating from this study would most likely be available online for you to access.

If you have any questions about this study or the survey, please contact me via email at edwin.frauenstein@gmail.com. This survey will take approximately 30-40 minutes to complete. Completion and submission of the survey constitute consent for the data to be used in the study. You may stop at any point should you wish to no longer continue and that this decision will not in any way affect you negatively. By proceeding (clicking on I AGREE), you

give your **full consent** to participate in this survey freely and without being forced in any way to do so. If you do not agree, you may exit the survey now by clicking on I DISAGREE.

I AGREE

I DISAGREE