Human Factors in a Computable Cybersecurity Risk Model

S. Williams[1] and D.A. Marriott[2]

[1]Defence Science and Technology Laboratory, United Kingdom
[2]Defence Science and Technology, Edinburgh, South Australia
e-mail: swilliams2@dstl.gov.uk, damian.marriott@dst.defence.gov.au

Abstract

Computable risk models are used for risk management in organisations to assess possible cybersecurity threats to the system and consider appropriate response options. These models might include humans, but usually do not contain information about how human factors have a role in information security. We describe the necessary aspects to consider when designing a framework for including human factors in a risk model, based on the example of a spear-phishing attack. Some key elements of this framework were implemented as a proof of concept using Chimera, a computable cybersecurity risk model.

Keywords

Human factors, spear-phishing, computable risk model, semantics, cybersecurity

1 Introduction

It is well understood that human factors (HF) have an important role in the information security and risk management of socio-technical systems (Schultz, 2005), although cybersecurity risk models often group humans as homologous objects in the system. This study has considered the question of how do we include (these) HF in computable risk models to ensure that appropriate security controls are used effectively to reduce cybersecurity risk and to minimise utility risk, which is disruption to the user's duties and the broader business operations. For the purposes of this study the National Institute of Standards and Technology (NIST) risk management terminology has been used where risks are a function of likelihood and impact of a threat occurring (NIST, 2012).

Previous work has considered how to dynamically calculate and balance security risk with utility risk (Billard, 2015). This study is cognisant of both security and utility risks, and is interested in how HF can affect the cybersecurity risks to a socio-technical system by affecting the likelihood and impact of threats (Aleroud et al., 2017). Similarly, security controls can be adapted for HF to ensure that all users of the system are better protected against cybersecurity threats. We aimed to integrate key aspects of the body of literature around HF in cybersecurity into a model and to demonstrate how HF can add value to a computable tool.

This paper will outline the concept and challenges faced with this task, introduce the concept of a cyber hazard and then show a proof of concept implementation of key elements in a semantic risk management tool. The implementation used to test this approach is based on Chimera, a semantic modelling and inference tool which itself is based on earlier risk management tools and methods (Chakravarthy et al., 2015).

2 Framework and challenges developing a computable model with HF

Our proposed HF risk model is a semantic ontology, which models a set of concepts and the semantic associations among them. The model includes static HF such as job role and relevant training undertaken, as well as dynamic factors (email load, for example), to improve tailoring of security controls for a system user or situation. The model tailors controls to individual needs so system users are more responsive to them, and utility threats only occur when necessary. Utility threats are threats that could disrupt the user's duties and broader business operations; these should be considered and balanced against often competing security-driven controls. A semantic ontology allows a flexible approach to the way that information is added to the model, depending on the amount of detail available in the data (i.e. the level of abstraction). New humans who interact with components already in the cyber risk model can be added in easily.

HF related to a user may influence how they respond to a cybersecurity threat. For example certain job roles involve frequent email use to initiate business relationships, so humans in these roles might be more responsive to spear-phishing emails posing as new potential clients. Our model has been developed based on the targeted phishing example as there was relevant literature available to support a good understanding of the HF involved in this threat (Sheng et al., 2010, Vishwanath et al., 2011, Pattinson et al., 2012). These include individual factors (for example visibility, suggestibility, responsiveness, culture, technical knowledge or authority), workplace-specific factors (such as workload, job role, confidence in IT security, morale, information and trust in procedures), or threat-specific factors (in the spear-phishing use case this would be email content, sender familiarity or credibility and use of persuasion tools) (Ovelgönne et al., 2017). Some factors, including workload, affect the likelihood of being exposed to a risk; whereas others, such as cybersecurity training or job role, will limit the impact of a particular threat once exposed.

We developed a framework by reviewing published literature on HF involvement in spear-phishing, then applying the measurable and relevant factors, described above, into our risk model with the appropriate level of detail.

2.1 Defining the effect size

There is a wide range of HF potentially relevant to a cybersecurity threat, which must be captured in the model. The weightings of each of the HF vary between system users, and are dynamic to situational changes. Only HF with a correlation to risk (from empirical data) are included in the model, and these are matched to a security control. Both factors which greatly influence threat likelihood or impact, and those with smaller effects, must be included in the computable HF model. Since the computable

HF model uses correlational data, it will not model a human's motivation or the underlying reasons for association of HF with susceptibility to specific cybersecurity threats. The effect size is based on empirical data, so confidence of these correlations should be communicated clearly.

HF are associated with either the whole organisation, applied to a group who work in the same organisational unit, all users of a piece of equipment, or to individual humans. The interactions between HF for different scopes must be considered for the model. However it is not possible, for many HF, to measure the effect size of each factor and compute how different combinations affect a system user's susceptibility to cybersecurity threats.

2.2 Human factor dynamic profiles

Including additional human instances and HF in the model will ensure that the organisation's diversity is captured, but it increases the model complexity. This could be managed by considering how some HF are often seen together, for example those related to a certain job role. Clustering system users into profiles, based on their HF, makes it simpler to estimate susceptibility and match controls.

A dynamic profile could be used to incorporate dynamic HF at run-time, and adapt controls to a human and the current situation. The dynamic aspect comes in by adding information about the environment, time of day or current workload. The dynamic profile is used to set a threshold for action, either to protect against general cyber risks or block a specific vulnerability to a threat. This is proportional; the threshold can change for different contexts. These profiles could be designed based on experimental data with weightings assigned based on statistical models such as linear regression, Principle Component Analysis (PCA) or clustering on real-world data sets. It is important that privacy is respected when used in an organisation, so that data which is included is limited to that which has prior consent to be collected and used in the model.

The use of dynamic profiles in our model will provide a better outcome for an organisation with a diverse workforce, by reducing users' susceptibility to cybersecurity threats through tailored controls. This simultaneously has the effect of reducing utility risks by virtue of reducing the application of blanket controls (which generally inhibit utility such as through reductions in efficiency) where the risk assessment, taking into account the HF profile, concludes they are not warranted.

2.3 Pairing with controls

Different security controls can be implemented depending on the dynamic HF profile of users involved and the environment. Controls are split into people-based and technology-based actions which are linked to design-time (modelled at a different time to the risk exposure) and run-time (at a time concurrent with the risk exposure) actions respectively. People-based controls promote enduring behaviour changes by increasing security knowledge (e.g. training relevant to identifying spear-phishing emails). Technology-based controls reduce immediate exposure to cybersecurity risks by limiting threats (e.g. increasing the sensitivity of the junk filter on incoming emails). Both types of control are deployed taking into account HF.

At present in the model, a threat is considered fully mitigated following deployment of an appropriate control; although in the future it would be better to quantify the effectiveness of a control. The model must compare the impact of combinations of controls on risk, and the impact of applying a control to a human compared to application across the entire organisation. A user's HF profile will affect how well they respond to a particular control. Further, utility risks, cost and security workarounds should be considered when choosing controls in the model.

Risk may change following a change in the dynamic HF profile, or based on feedback from a control (e.g. feedback from an online training game). Table 1 shows examples of how the default control action for a cyber hazard (see next section) can be tailored to individual needs based on their HF profile.

Cyber hazard	Default action	Adaptation to HF model
Untrustworthy attachment	Obstructive security warning for all system users	Nudging techniques (Voyer, 2015): a passive, or less obstructive, reminder for those trained
External URL	Allow click through to the hyperlink	Replace hyperlink with text for those less security aware; remove altogether for those who have failed an online security training game
Email from a known suspicious source	Move email to junk folder	Delay user interaction with the suspicious email until the next morning
Email (undetected spear-phishing)	No analysis of the email content	Flag use of Cialdini's persuasion factors for susceptible HF profiles (Cialdini, 2007)

Table 1: Control actions and adaptation to situation

2.4 Cyber hazards

In dynamic risk modelling, we differentiate between cyber hazards and cyber threats whereby a realised hazard is a circumstance or event that increases the likelihood or potential impact of a risk occurrence, but does not in itself constitute a risk occurrence. Conversely, a realised threat produces an adverse impact, i.e. risk occurrence. We do not believe this distinction between cyber hazard and threat has been made previously. Cyber hazards can indicate (and give advance warning of) a potential cyber threat. Here, hazards are steps towards or precursors of a risk occurrence and signal opportunities to apply extra caution. This application of extra caution can be regulated by HF.

Where there are hazards detected at run-time, we may only want this information to be acted upon in certain situations, for example with users that have not had cybersecurity training and are not in a technical role. The controls for hazards are less critical than for risks due to the nature of hazards being precursors to (but not actual) risk occurrences. Therefore they are open to controls which are persuasive actions, such as nudging, rather than absolute actions, such as blocking. The selection of such persuasive controls can more readily take into account HF, which has the potential to reduce security risks by HF-tailored pre-emptive controls whilst also reducing utility risks by being less absolute or even absent where users have sufficient cybersecurity awareness.

2.5 Spear-phishing example

Spear-phishing emails are a problem because actions resultant from them can lead to disruption, loss of trust and financial consequences, and these cyber threats are getting increasingly sophisticated and difficult to identify (Hong, 2012). We modelled a spear-phishing attack in the risk model framework and included HF where they fit the model requirements of being measurable and quantifiable. The attack is divided into stages which have technical control point failures and cyber hazards identified (see Figure 1). Individual and situational HF can affect a system user's susceptibility to phishing, and the persuasive measures in a phishing email may be more effective on some users depending on their HF.

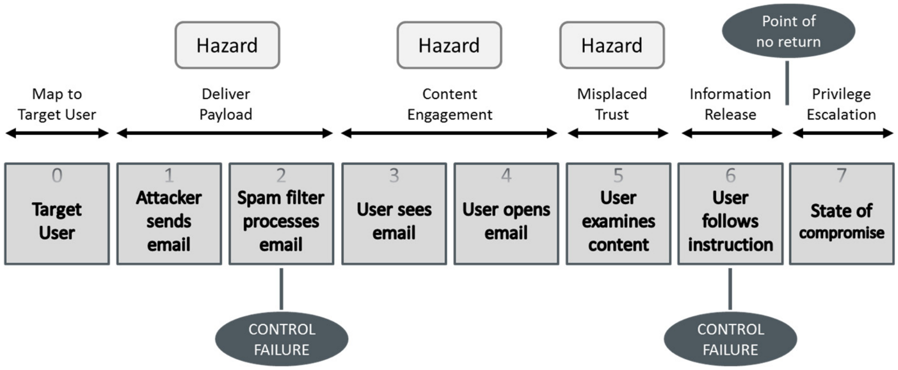


Figure 1: The stages of a spear-phishing attack

People are usually selected to receive spear-phishing emails based on being known to have access to something of interest to the attacker. Target attributes (Stage 0) can be described by visibility (online and offline), responsiveness and culture (individual, workplace, and even national) (Rocha Flores et al., 2015, Vroom and Von Solms, 2004, Butavicius et al., 2017).

After arriving into the user's inbox, different HF weightings shape whether a spear-phishing email is actually read (Stages 3-4). Some important HF include cultural factors, workload, technical knowledge, job role and confidence in IT security. The initial email properties relevant are subject, credibility, expectation and sender familiarity or authenticity, which could be computed in our risk model.

When considering email content (Stage 5), a user's responsiveness and suggestibility contribute to their perception of the motivation of the sender (Williams et al., 2017). Use of persuasion factors will change susceptibility, depending on HF such as workplace morale, emotional plea, reciprocity, social influence, urgency, credibility, training, and job role (Cialdini, 2007).

The impact of trusting a spear-phishing email, and acting on its instructions is determined by the information released (or other action such as deleting or encrypting data; Stages 6-7). HF related to the trait of compliance may influence this. It is at this stage, when information is released or other unauthorised actions are carried out on the system, that a threat is realised and there is a risk occurrence. Before that, there are a number of hazards present, e.g. when the email is processed and spooled ready for the user (Stage 2), when the email is seen and opened by the user (Stages 3-4), and when the email content is examined by the user (Stage 5). Controls may be deployed, informed by HF, at the hazard and threat stages (e.g. see Table 1).

3 Proof of concept implementation using Chimera

The Chimera tool from the UK's Defence Science & Technology Laboratory (Dstl) builds upon existing risk management tools and methods (Chakravarthy et al., 2015), security standards (e.g. RFC 4949), and uses semantic modelling and inference for automated threat determination and mitigation strategies for Information & Communications Technology (ICT) systems. Chimera takes a system design, composed of assets and their relationships, and identifies threats to the assets and suggests mitigation strategies from expert knowledge encoded in the semantic model. Here we show how Chimera can be adapted to include HF in its design-time risk management approach using a simple use case. TopBraid Composer™ (TopQuadrant™, 2018) was used to model the ontology; screen grabs below are taken directly from the tool to demonstrate SPIN rules (see Figures 3-5).

The assets and their relationships of interest in the use case are depicted in Figure 2, comprising organisations (MOD, DSTL), humans (DSTLStaff_1, DSTLStaff_2, CyberSpecialist_1, UKMilitary_1), logical and physical assets (ApacheServer_T3, IntensePC_1, TerminalService_1, PythonFlaskWebServer_S1, Snorby_1), and the relationships between them. Organisations control humans that in turn control assets. For illustrative purposes the screen grab from Chimera in Figure 2 has been manually overlaid with HF properties that have been asserted or inferred by SPIN rules as described in the following.

In order to demonstrate the implementation of HF at an organisational level, and then at an individual level, we consider cybersecurity training. The organisations and humans in the model may have a Boolean property, *hasTraining*. In our fictitious scenario, we have one organisation (MOD) which has the *hasTraining* property set to *true*, whereas for the other (DSTL) it is set to *false*. This indicates that MOD have a good cybersecurity training regime in place. These settings would be determined and set perhaps on an annual basis, i.e. essentially statically or in design time. One human (CyberSpecialist_1) has the *hasTraining* property set to *true* directly, and this would be set when the training was completed and perhaps reset if the training proficiency lapses. We have a SPIN rule in place to add at run time the *hasTraining* property from an organisation to humans it controls unless that property is set directly on a human (see Figure 3). The result of the rule for *hasTraining* on our model is that DSTLStaff 1&2 gain their organisational setting of *false*, CyberSpecialist_1 retains its existing setting of *true*, and UKMilitary_1 gains its organisational setting of *true*. That is, the more general setting is adopted in the absence of the more specific.

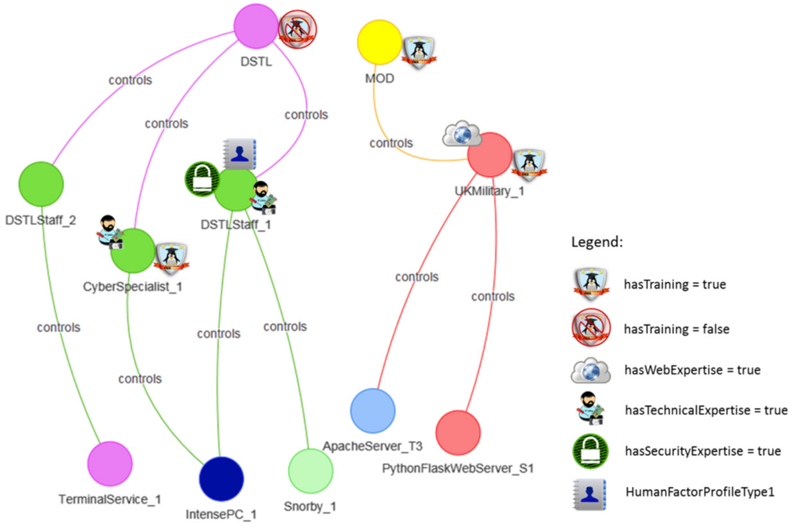
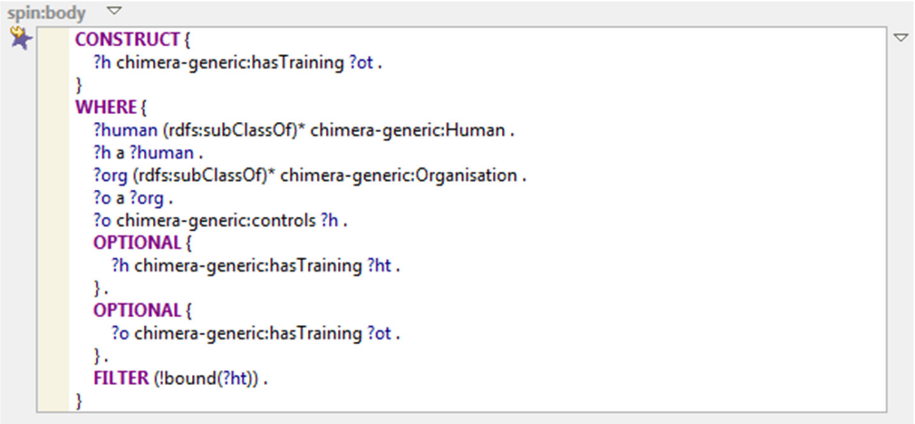
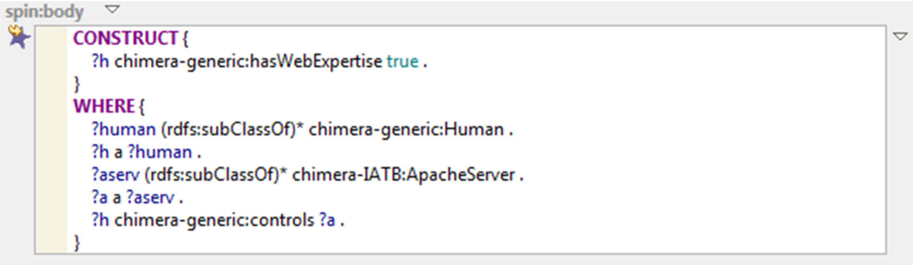


Figure 2: Chimera representation of objects and their relationships, overlaid with HF properties



```
CONSTRUCT {
   ?h chimera-generic:hasTraining ?ot .
}
WHERE {
   ?human (rdfs:subClassOf)* chimera-generic:Human .
   ?h a ?human .
   ?org (rdfs:subClassOf)* chimera-generic:Organisation .
   ?o a ?org .
   ?o chimera-generic:controls ?h .
   OPTIONAL {
      ?h chimera-generic:hasTraining ?ht .
   } .
   OPTIONAL {
      ?o chimera-generic:hasTraining ?ot .
   } .
   FILTER (!bound(?ht)) .
}
```

Figure 3: *hasTraining* SPIN rule



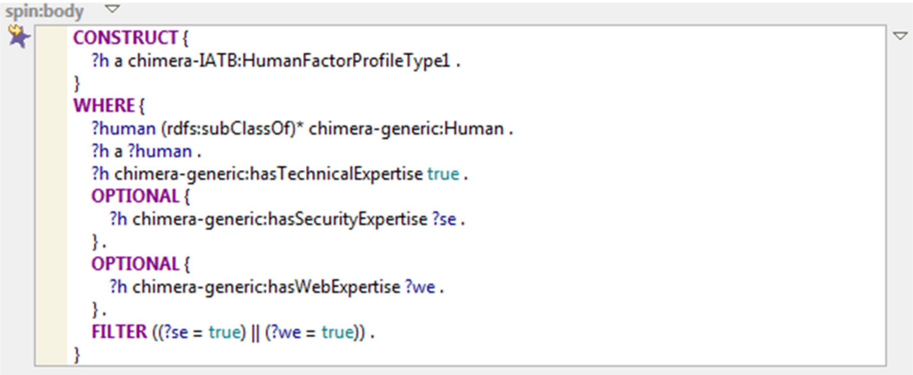
```
CONSTRUCT {
  ?h chimera-generic:hasWebExpertise true .
}
WHERE {
  ?human (rdfs:subClassOf)* chimera-generic:Human .
  ?h a ?human .
  ?aserv (rdfs:subClassOf)* chimera-IATB:ApacheServer .
  ?a a ?aserv .
  ?h chimera-generic:controls ?a .
}
```

Figure 4: *hasWebExpertise* SPIN rule

The relationships between humans and assets in the model connect system users to logical and physical assets. We use the control relationships to indicate job roles from which we infer HF. For instance, there is a rule (see Figure 4) that sets the *hasWebExpertise* property to *true* for anyone who controls an Apache server (UKMilitary_1). Similarly for anyone who controls an IntensePC we set *hasTechnicalExpertise* (DSTL_Staff_1, CyberSpecialist_1), and for anyone who controls Snort or Snorby we set *hasSecurityExpertise* (DSTL_Staff_1).

A more refined model might aim to capture this kind of expertise information for individuals, however, in the absence of that data we argue that the approach above is an improvement over not taking into account these HF when considering risk analysis.

The SPIN rule in Figure 5 assigns humans to a particular HF profile if they have technical expertise and at least either security or web expertise.



```
CONSTRUCT {
  ?h a chimera-IATB:HumanFactorProfileType1 .
}
WHERE {
  ?human (rdfs:subClassOf)* chimera-generic:Human .
  ?h a ?human .
  ?h chimera-generic:hasTechnicalExpertise true .
  OPTIONAL {
    ?h chimera-generic:hasSecurityExpertise ?se .
  } .
  OPTIONAL {
    ?h chimera-generic:hasWebExpertise ?we .
  } .
  FILTER ((?se = true) || (?we = true)) .
}
```

Figure 5: Human Factor Profile SPIN rule

4 Conclusions and future work

We present the framework and an implementation for including key human factor elements in an ontology-based risk model at design time.

Others have used data-driven approaches to look at how susceptibility to cyber-attack is related to features of human behaviour, including job role (Ovelgönne et al., 2017). We are using a similar approach to tailor security controls to the user based on their

HF, to improve response to the controls and reduce utility risk. We describe the benefit of including even a small number of HF in the model to acknowledge the diversity of system users in an organisation. The HF weightings and control pairings in our model must be validated, possibly from published experimental data sets.

We have also covered concepts which are necessary for implementation in a run-time model and introduced the concept of a cyber hazard. Future work will include HF in run-time risk management using this tool. The algorithm (e.g. SPIN rules) can learn and add knowledge from previous scenarios, because a new data point is added for each change to a HF variable. There is no action until the defined threshold is reached, when an appropriate control is chosen. Our model could be used to capture the complexity of cybersecurity risk management in socio-technical systems, and current application of this model will be limited by data regulations and by information that is available about humans in the organisation.

5 Acknowledgements

We thank Sam R. & Andrew F. of the NCSC and A. Chakravarthy (Dstl) for useful discussions and help generating Figure 1.

6 References

Aleroud, A. and Zhou, L. (2017). “Phishing environments, techniques, and countermeasures: A survey”, Computers & Security, 68, pp. 160-196.

Billard, A. (2015), “Security and Utility Risk Evaluation (SURE) Framework for Dynamic Cyber Systems”, DSTO-TR-3080, Defence Science and Technology, Australia.

Butavicius, M.A., Parsons, K., Pattinson, M.R., McCormac, A., Calic, D. and Lillie, M. (2017), “Understanding susceptibility to phishing emails: Assessing the impact of individual differences and culture”, Proceedings of the 11th International Symposium on Human Aspects of Information Security and Assurance, pp. 12-23.

Chakravarthy, A., Chen, X., Nasser, B. and Surridge, M. (2015), “Trustworthy systems design using semantic risk modelling”, 1st International Conference on Cyber Security for Sustainable Society, pp. 49-81

Cialdini, R.B. (2007), “Influence: The psychology of persuasion”, New York: Collins, pp. 173-174.

Hong, J. (2012), “The state of phishing attacks”, Communications of the ACM, 55(1), pp. 74-81.

NIST (2012), “Guide for conducting risk assessments”, NIST SP800-30Rev1, National Institute of Standards and Technology, U.S.

Ovelgönne, M., Dumitras, T., Prakash, B.A., Subrahmanian, V.S. and Wang, B. (2017), “Understanding the Relationship between Human Behavior and Susceptibility to Cyber Attacks: A Data-Driven Approach”, ACM Transactions on Intelligent Systems and Technology (TIST), 8(4), p.51.

Pattinson, M., Jerram, C., Parsons, K., McCormac, A. and Butavicius, M. (2012), “Why do some people manage phishing e-mails better than others?”, Information Management & Computer Security, 20(1), pp. 18-28.

Rocha Flores, W., Holm, H., Nohlberg, M. and Ekstedt, M. (2015), “Investigating personal determinants of phishing and the effect of national culture”, Information & Computer Security, 23(2), pp. 178-199.

Schultz, E. (2005), “The human factor in security”, Computers & Security, 24, pp. 425-426.

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F. and Downs, J. (2010), “Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions”, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, pp. 373-382.

TopQuadrant™ (2018), TopBraid Composer™, https://www.topquadrant.com/tools/modeling-topbraid-composer-standard-edition/. (Accessed 24 May 2018)

Vishwanath, A., Herath, T., Chen, R., Wang, J. and 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), pp. 576-586.

Voyer, B.G. (2015), “‘Nudging’ behaviours in healthcare: Insights from behavioural economics”, British Journal of Healthcare Management, 21(3), pp. 130-135.

Vroom, C. and Von Solms, R. (2004), "Towards information security behavioural compliance", Computers & Security, 23(3), pp. 191-198.

Williams, E.J., Beardmore, A. and Joinson, A.N. (2017), "Individual differences in susceptibility to online influence: A theoretical review", Computers in Human Behavior, 72, pp.412-421.