TECHNICAL ANNEX: EVIDENCE REVIEWS

Review of Behavioural Sciences Research in the Field of Cybersecurity

DECEMBER 2018

www.enisa.europa.eu European Union Agency For Network and Information Security

About ENISA

The European Union Agency for Network and Information Security (ENISA) is a centre of network and information security expertise for the EU, its member states, the private sector and EU citizens. ENISA works with these groups to develop advice and recommendations on good practice in information security. It assists member states in implementing relevant EU legislation and works to improve the resilience of Europe's critical information infrastructure and networks. ENISA seeks to enhance existing expertise in member states by supporting the development of cross-border communities committed to improving network and information security throughout the EU. More information about ENISA and its work can be found at www.enisa.europa.eu.

Contact

For media enquires about this paper, please use press@enisa.europa.eu.

ISBN 978-92-9204-326-1 DOI 10.2824/717888

Table of Contents

1. Scope of the reviews

This technical annex contains the four reviews that supported the writing of the report Review of “Behavioural Sciences Research in the Field of Cybersecurity”. The reviews are:

1. Measurement of cyber security attitudes and behaviours
2. Interventions to change cybersecurity behaviour
3. Beyond surveys - qualitative and mixed-method studies
4. Current Practice

Where possible (and appropriate) the reviews followed systematic reviewing protocol, but in order to survey the field as widely as possible this was not always rigorously adhered to. The evidence reviews were compiled independently, with shared conclusions and insights used for the main report.

Given the limitations of the review, some specific topics and instances are under-reported. For instance, ‘cybersecurity’ is often not used in the title / abstract of papers when the publication outlet is security / technology based. The reviews may therefore have missed some work that deals with security without explicitly using the terms ‘cybersecurity’ or ‘information security’ in the title, abstract or keywords. Similarly, some papers deal with specific threats (e.g. phishing) or solutions (backups) without mention of cybersecurity in the same fields, so again may have been missed.

2. Evidence review: Measurement of cyber security attitudes and behaviours

A recent analysis of papers that claimed to use behavioural science constructs to study and/or influence human behaviour in cyber security was conducted by Becker & Sasse at UCL with funding from the UK National Cyber Security Centre (NCSC). The authors reviewed 688 publications that study human behaviour in information security via survey methods to describe how a population (typically the employees of an organisation) score in terms of those constructs. Between them, these publications used 984 constructs from 92 categories to measure various aspects of human behaviour. The majority (695/984) of constructs were derived from existing social science theories, and relevant instruments for measuring these constructs were used. In addition there were 217 newly created `security constructs'. A list of the publications, the categorisation, detailed analysis and conclusions can be found here https://verdi.cs.ucl.ac.uk/constructDB/ but we describe the methodology and conclusions here to feed into our discussion and conclusions.

1. The authors searched google scholar for combinations of the terms `information security', `security', `survey', `questionnaire' and `construct'. Over 3 million relevant articles were returned by the search engine. The first 30 pages of search results (i.e. 3000 articles) were analysed against the three criteria above and 124 relevant publications were identified.

2. For every article analysed, the articles that were cited for constructs were added to the analysis queue (going backwards in time).

3. For every article analysed, Google Scholar was used to identify citing publications (i.e.going forward in time). The 30 most cited publications (some psychology publications had tens of thousands of citations) were added to the analysis queue if they conformed to the selection criteria above.

Steps 2 and 3 initially increased the size of the analysis queue exponentially, every paper analysed would add 10 papers to the queue. However after analysing 400 publications, the queue started to become of x-ed length, i.e. for every paper analysed one more paper was added to the queue. At over 1000 constructs identified (before merging), the authors say “we are reasonably condent that a comprehensive view of constructs in security has been achieved.”

For each publication, the following data was extracted:

1. a short description of the research (usually a snippet of the abstract)
2. research type and sample size (for example user study with 180 students / meta review /construct validation study)
3. the source PDF file
4. the constructs used or discussed, where for each construct we collected:

    - the exact sources referenced for the construct, or any comment by the authors if they created the construct themselves
    - the type of the construct (usually the theory on which the construct is based)
    - whether the article provides the exact questions used
    - whether the article gives the answer options to the questions (and if so, what type they are)
    - Whether two measures of validation were used.

We incorporated the results from a review carried out earlier this year with funding from the UK National Cyber Security Centre (NCSC). The review analysed 688 publications that claimed to use behavioural science constructs - variables that are not directly observable, such as attitudes or personality traits, and are assumed to influence human behaviour in cyber security - often towards compliance or non-compliance with security policies. Examples of how the results from such surveys in organisations might then be used include:

1. The organisation screens prospective employees to identify those who score highly on constructs associated with compliance.
2. The organisation screens existing employees to identify those who score highly on constructs associated with non-compliance, and targeting them with security awareness and/or behaviour modification activities.
3. The organisation assess the effectiveness of security awareness and/or behaviour modification activities by how selected or all employees score on such constructs.

Most of these publications claim to use well-established constructs and associated instruments from social sciences, the looked at whether a) the original (validated) constructs and instruments had been used, and b) the studies and information provided met scientific quality standards. The review revealed that there were 92 categories and 984 constructs that have been investigated in relation to cyber security behaviour.

Most studies claim to have found a link between constructs and behaviour - and generally assert that some factor (or pattern of factors) within the employees correlates with undesirable security behaviour. For instance, Sohrabi Safa, Von Solms & Furnell (2016) measured responses to the constructs *information security knowledge sharing, collaboration, intervention and experience*, plus *attachment, commitment, personal norms* and *attitude to information security policy compliance* and *intention to comply with information security policies* with 462 employees in 4 companies, and conclude that *'the lack of information security awareness, ignorance, negligence, apathy, mischief, and resistance are the root of users' mistakes'* - i.e. attributing undesirable security behaviour to entirely to failure on behalf of the employees. From a scientific point of view, this and conclusions from similar studies is not tenable without some form of validation - triangulation and/or repeated measurements (see Section Evidence Review 4).

2.1 Conclusion

Our conclusions for researchers and practitioners is that from a scientific point of view, the vast majority of the studies reviewed does not provide solid evidence of specific psychological traits driving security behaviour.

1. Some of the studies find correlations between some constructs and (almost entirely) self-reported behaviour. But this ignores the difference between *correlation* and *causality* - there may be other underlying factors influence both the constructs as measured by the instrument and the security behaviour. For instance, that daily experience of unworkable security policies shapes the attitude to security, as well as driving non-compliance behaviour. Adams & Sasse (1999), for instance, observed how unworkable password policies had led employees to conclude that cyber security measure were put in place to make their life difficult, rather than offer protection.

2. The studies all assume that compliance with security policies is sensible behaviour. It is assumed - rather uncritically - that the security policies and measures that employees are supposed to comply with are effective, and that following them improves security. Herley (2009) made a powerful argument that ignoring most common security advice is a rational choice once effort and effectiveness are compared. Bonneau et al. (2015) demonstrated that strong passwords as advised mean significant effort for users, but offer little

protection against current attacks against passwords - and thus not bothering is rational. The review of studies grounded in organisational contexts in (Evidence Review) finds that in most cases, when employees do not comply it is because doing so would reduce their productivity, and that of the organisation, to unacceptable levels.

3. The 9 constructs used by Sorhabi Safa et al. are among 789 unique constructs identified in the review that have been used to try and explain security behaviour https://verdi.cs.ucl.ac.uk/constructDB/constructs/ and range from personality traits measured through the widely used Big Five (*Openness to Experience, Neuroticism, Conscientiousness, Extraversion, Agreeableness*) over ethical stances (*Utilitarianism vs. Formalism*) to high security-specific *intention of comply with information security policies*. The top investigated behaviour (60 studies) is Ajzen's (1991) generic *Theory of Planned Behaviour (TPB)* - discussed in more detail below. The other top concepts are *Compliance* (40) and *Intention* (29). The large number of constructs itself is an indication that there is no agreement in the research community - apart from TBP - on which theories are likely to be applicable. The picture that emerges is one of security researchers with engineering backgrounds 'grasping' plausible constructs that can be measured and explain non-compliant behaviour.

4. This latter related to the fourth point - lack of reliable results. The review found that most results are not reliable - only a quarter of the studies met basic criteria for scientific survey research. Even where previously validated constructs have been used, the security surveys often made 'tweaks' to adjust the original, validated instruments - and then used them without further validation. The conclusion from the review is that most of these surveys are an exercise in trying to find something in employees that can be blamed for their non-compliant security behaviour, and used by organisations to 'fix' it (see points 1-3 above). But the results of three quarters of the studies cannot be regarded as reliable - a conclusion that is reinforced by largely divergent results.

5. The top investigated behaviour (60 studies) is Ajzen's (1991) generic *Theory of Planned Behaviour (TPB)*, which posits that - if people evaluate the suggested behavior as positive (*attitude*), and if they think their significant others want them to perform the behavior (*subjective norm*), this results in a higher intention (*motivations*) and they are more likely to perform that behaviour - so far, so rational. TPB then adds the construct of *self-efficacy* - whether a person believes that she can successfully execute the behavior required to produce the desired outcomes. This is a concept adapted from Bandura (1977), who stated that self-efficacy is the most important precondition for behavioral change because it determines the initiation of coping behaviour - a conclusion which is supported by Evidence Review 2.

3. Evidence review: Interventions to change cybersecurity behaviour

For this review, we look to models of behaviour change generally (and guidance for the application of models), and then conduct a systematic literature review to examine the effectiveness of human-level interventions to improve cyber security behaviour amongst users. As noted above, one of the challenges faced is that rather than study actual behaviour most studies of human aspects of cyber-security measure either awareness, concerns or intentions to behave.

3.1 Models of behaviour in cyber-security

Within the field of behaviour change there are multiple models and theories used for behaviour change which in turn underpin a vast array of behaviour change techniques. In a study by Abraham and Michie (2008), 26 different behaviour change techniques were identified and categorized (this expanded to 93 in a later 2013 study). Michie and colleagues (2011) have subsequently developed the 'COM-B' model that seeks to identify why a behaviour is occuring (or not happening), and then apply the appropriate intervention based on that analysis.

The 'COM-B' model (Michie et al., 2011) argues that whether or not a behaviour is enacted (e.g. locking a screen when leaving for lunch) is dependant upon three interrelated factors: 1) capability (can they do it? Do they know how to?); 2) opportunity (do they have the chance to do the action?); and motivation (are they motivated to lock the screen?). The type of intervention is dependent upon the cause of the (non)behaviour - so for instance, if users are able to lock a screen, have the opportunity to do it, but are not motivated to, then interventions based around creating a motivation (e.g. education, awareness, reward/punishment) are most appropriate. However, if initial analysis found that users were motivated, but did not know how to lock screens (capability), then an intervention should be based on training and education.

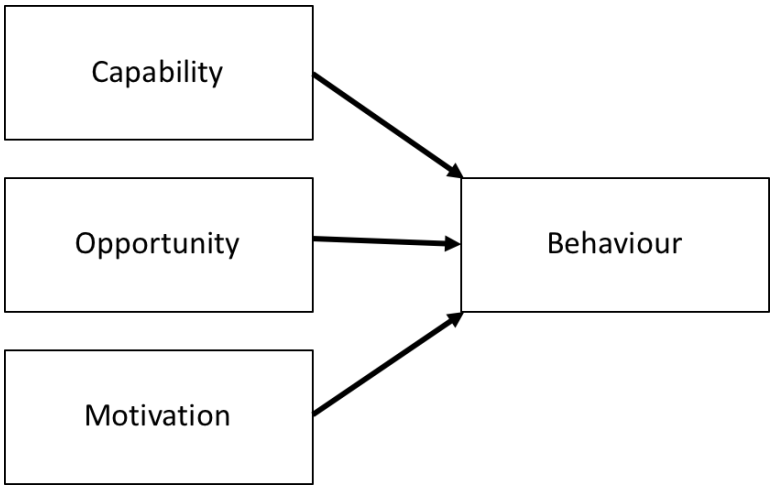


Figure 1: COM-B model (adapted from Michie et al., 2011)

A related model is that developed by Fogg (2009) that seeks to identify the type of cue needed to encourage the appropriate action, dependent on an individual's motivation and ability to perform the act. According to the B=MAT model, the likelihood of a behaviour occurring is a product of motivation (M), Ability (A), and the appropriate trigger (T).

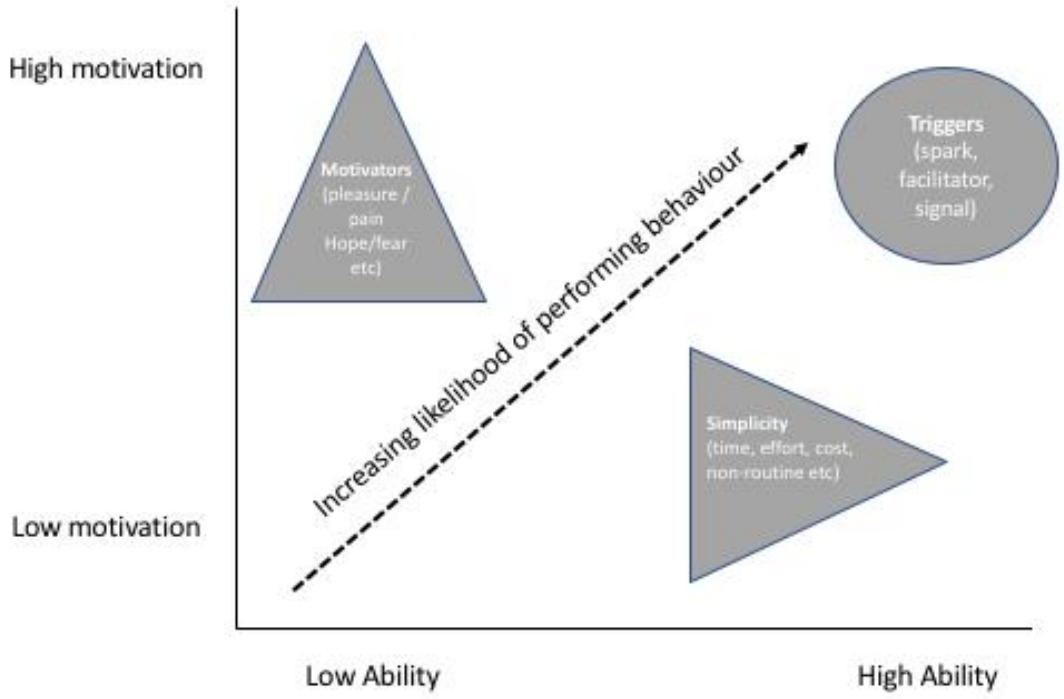


Figure 2: B=MAT model (adapted from Fogg, 2009)

According to the B=MAT model, the type of persuasion required to bring about a behaviour depends on where it lies in the motivation / ability dimensions, with different interventions needed to increase either motivation or ability. For instance, if people are motivated to undertake a task (e.g. updating software), then addressing their ability (e.g. by reducing the cost or effort) should increase the likelihood of carrying out the behaviour. Similarly, if an action is simple and the person is able to complete it, then addressing motivation (e.g. fear of outcome, hopes, pain) should also increase the likelihood. Once motivation and ability is addressed, according to Fogg's model, we should then look to triggers that signal to people that a behaviour is required. These triggers can take the form of: 1) signals (e.g. a message saying that updates are ready to be installed), best used when someone has motivation and ability,; 2) sparks that seek to motivate as well as trigger a behaviour (e.g. warning that the computer will be at risk if the update isn't installed); or 3) facilitators, that seek to both trigger a behaviour and make it easier (e.g. "just click here to download and install the update").

This behavioural approach stands in contrast to the attitude-behaviour approaches more popular within health and social psychology, and applied in many cyber-security studies. According to the majority of these approaches, a behavioural intention is the consequence of multiple factors such as the individuals' attitudes towards the behaviour, the amount of control they have over the behaviour and the behaviour of others around them. One such

example is the theory of planned behaviour (TPB; see Figure 3). According to the TPB, an intention to behave is determined by a combination of the person's attitudes towards the behaviour, what others are doing ('subjective norms') and the amount about control they have over whether or not to enact the behaviour ('perceived behavioural control').

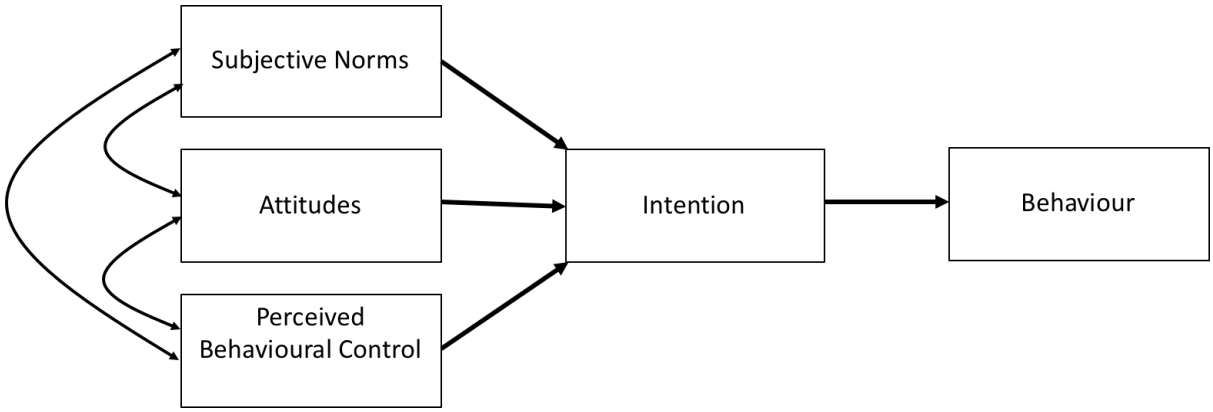


Figure 3: Theory of Planned Behaviour

In health behaviour, this is often translated into models that take into account not only people's attitudes towards a health behaviour (e.g. "exercise is good"), but also factors that might influence whether or not people will act on that belief to actually take action. For instance, the health belief model (Figure 4) places the perceived threat of an illness or disease at the centre of people's decisions. This perceived threat is made up of two factors - the perceived seriousness of an illness / health problem, and the person's perceived susceptibility to the illness. Assuming that a threat is seen as serious, an individual's likelihood of taking preventative action is also determined by their individual characteristics such as their belief in being able to bring about change ('self-efficacy'), cues in the environment to act (e.g. mass marketing campaigns, medical leaflets), and the weighing up of the benefits of taking the action vs. costs / barriers to taking the action.

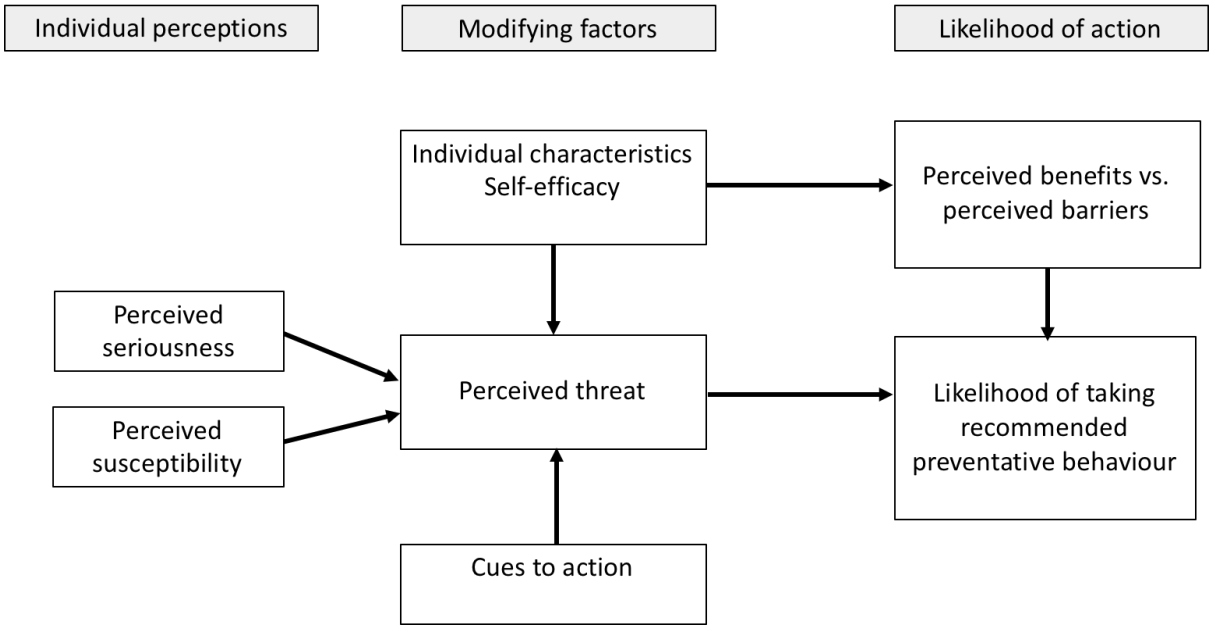


Figure 4: Health belief model (adapted)

The core ideas of the health belief model have gained some traction in studying people's information security behaviour. The central role of 'perceived threat' for a health problem has, in these approaches, been treated as analogous to the perceived threat posed by information security breaches, with seriousness and severity used to predict the likelihood of users' taking preventative action.

3.2 Protection Motivation Theory

Protection motivation theory (PMT; Rogers, 1975) has proved to be one of the most resilient models for studying how humans make decisions about security risks, in particular their likelihood of taking protective action against a perceived threat (Mayer, Kunz, & Volkamer, 2017). According to PMT, the likelihood of someone being motivated to take protective action is as a consequence of the balance between people's appraisal of the threat (how severe the threat is, their likely vulnerability to it, and the rewards felt by continuing the unsafe behaviour), and their appraisal of the ways to cope with the threat (how effective a response will be, their ability to take effective action and the costs of the response).

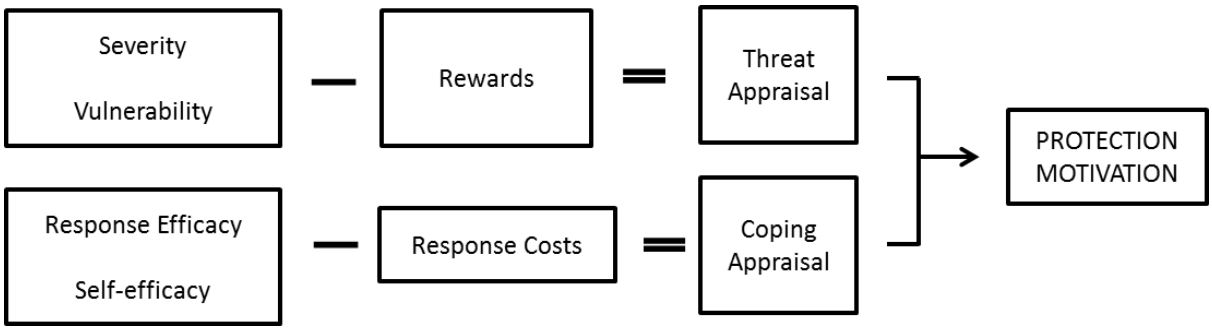


Figure 5: Protection motivation theory (from: https://commons.wikimedia.org/wiki/File:Protection_motivation_theory.png)

In the following section, we conduct a systematic review of studies of information security behaviour, in particular attempts to change security attitudes or behaviours. While not specifically searching for articles that utilized the models above, they are used to guide the write up.

3.3 Systematic review: information security attitudes and behaviours

3.3.1 Methodology

The current review follows the guidelines that are set out by the Preferred Reporting Items for Systematic reviews and Meta-Analysis PRIMSA (Moher, Liberati, Tetzlaff, & Altman, 2009; Shamseer et al., 2015). The search terms (Table 1) were run in PubMed, Association for Computing Machinery (ACM), PsychNet and Embase on the 6th August 2018.

Table 1. Search terms used for the different databases.

Database	Search terms	NOT	Results
PsychNet	**Title**: Cybersecurity OR **Title**: "Cyber security" OR **Title**: cyber-security OR **Title**: "Information security" NOT Title: "cyber crime" NOT Title: "cyber bullying" AND **Abstract**:	Cyber crime Cyber bullying	68

	behav* OR **Abstract**: attitude*		
ACM	acmdlTitle:("cyber-security" "cybersecurity" "cyber security" "information security") AND recordAbstract:(behav* attitud* -cyberbullying -cybercrime) [new search] [edit/save query]		138
Embase	(('cybersecurity':ab,ti OR 'cyber-security':ab,ti OR 'information security':ab,ti) AND 'attitud*':ab,ti) OR (('cybersecurity':ab,ti OR 'cyber-security':ab,ti OR 'information security':ab,ti) AND 'perceiv*':ab,ti) OR (('cybersecurity':ab,ti OR 'cyber-security':ab,ti OR 'information security':ab,ti) AND 'human*':ab,ti) OR (('cybersecurity':ab,ti OR 'cyber-security':ab,ti OR 'information security':ab,ti) AND 'behav*':ab,ti)		80
PubMed	(((((cyber-security[Title/Abstract]) OR cybersecurity[Title/Abstract]) OR cyber security[Title/Abstract]) OR information security[Title/Abstract]) AND (behav*[Title/Abstract] OR	Cybercrime Cyberbullying	59

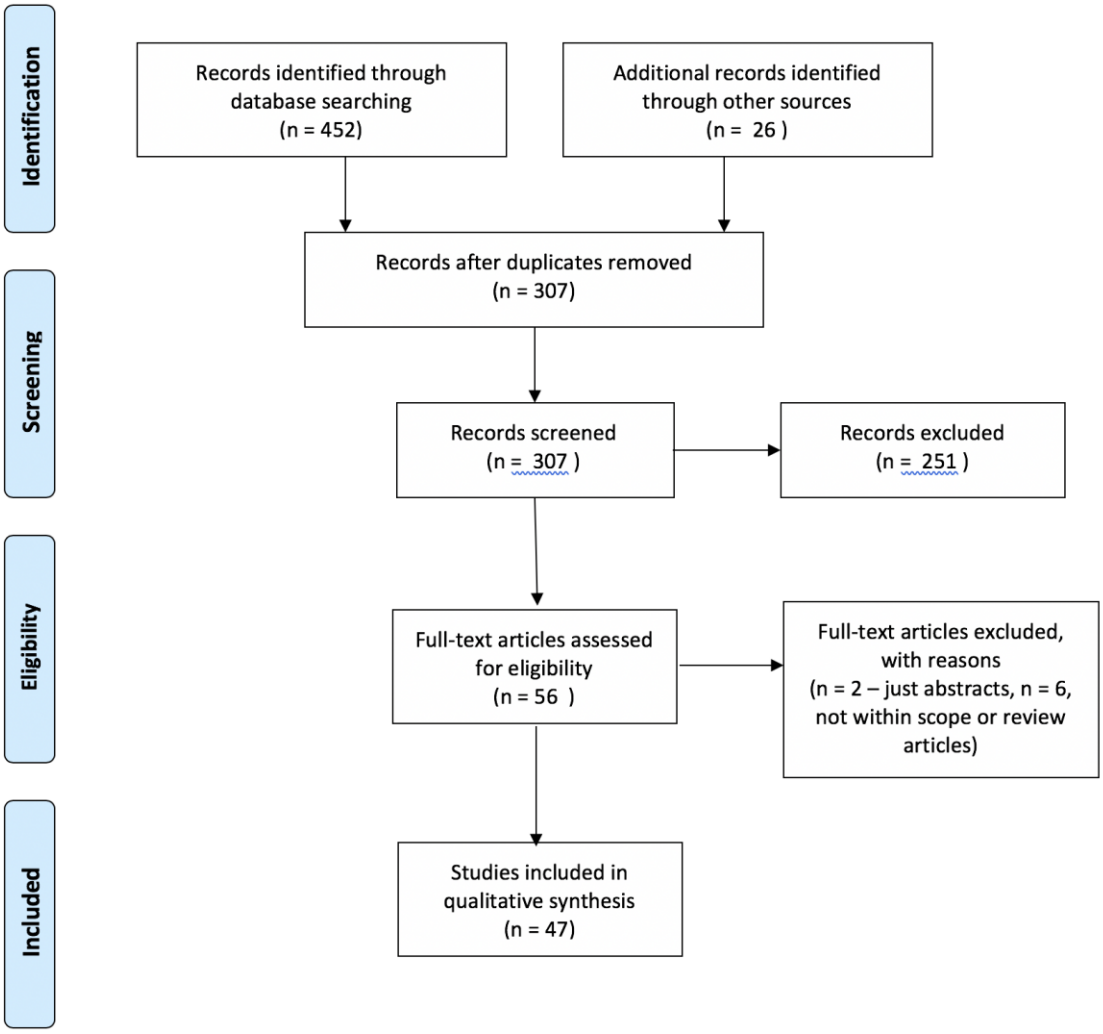
	attitud*[Title/Abstract])) NOT (cyberbullying[Title/Abstract] OR cybercrime[Title/Abstract])		

3.3.2 Inclusion/Exclusion Criteria

All articles were assessed for inclusion according to the following criteria: (1) the behaviour or attitudes needed to be aimed at information security and (2) journal or peer reviewed proceedings articles. Papers were excluded if; (1) they were not peer reviewed, (2) they were not available in English and (3) there was no full text available.

3.3.3 Search Process

The initial searches returned 452 articles. Additional articles were sourced by looking for papers published by authors who appeared multiple times in the initial database searches. In addition, the references used in a (relatively) review paper (Mayer et al.) were included in the first pass. Once duplicates were removed, this led to 307 articles. The titles and abstracts of all articles were screened for inclusion and 54 articles remained for full text review. Of the remaining records a further three were excluded, all of which were conference or poster abstracts or plans for studies reported at doctoral colloquium. This left a total of 47 full text papers to be included within the review.


Identification
Screening
Eligibility
Included
Records identified through database searching (n = 452)
Additional records identified through other sources (n = 26)
Records after duplicates removed (n = 307)
Records screened (n = 307)
Records excluded (n = 251)
Full-text articles assessed for eligibility (n = 56)
Full-text articles excluded, with reasons (n = 2 – just abstracts, n = 6, not within scope or review articles)
Studies included in qualitative synthesis (n = 47)

3.4 Results

3.4.1 Measurements used to study cyber-security

The studies within the review utilized a variety of metrics to measure information security behaviour or attitudes. A small number used measures of actual behaviour, including: actions with a game (Ben-Asher & Meyer, 2018); reactions to warning pop-ups (Boss, Galletta, Lowry, Moody, & Polak, 2015; Williams, Morgan, & Joinson, 2017); identification of phishing emails in a web survey / mocked up page or inbox (Canfield, Fischhoff, & Davis, 2016; Egelman, Harbach, & Peer, 2016; Martin, Dube, & Coovert, 2018) or via phishing simulations where a phishing email is sent to participants' actual inbox (Caputo, Pfleeger, Freeman, & Johnson, 2014; Halevi, Lewis, & Memon, 2013; Oliveira et al., 2017; Williams, Hinds, & Joinson, 2018); collecting data from users' actual computer (e.g. operating system version for updating, or apps on phones to collect unlock patterns: (Egelman et al., 2016); mouse movements when navigating spoof and non-spoof sites (Kelley, Amon, & Bertenthal, 2018); malware attacks detected by anti-virus software (Ovelgönne et al., 2017); and having participants enter passwords that are then checked for strength (Egelman et al., 2016; Mamonov & Benbunan-Fich, 2018). One study (Boss et al., 2015) edged towards experience sampling methodology by asking participants to record when they made backups over eight weeks (Study 1). An additional study examined cyber security behaviour and risk taking in a capture the flag event (Shoshitaishvili, Invernizzi, Doupe, & Vigna, 2014).

A substantial number of studies used self-reported information security behaviours (e.g. (Anwar et al., 2017; Bauer & Bernroider, 2017; Bulgurcu, Cavusoglu, Benbasat, & Ieee, 2010; Chou & Chou, 2016; Egelman & Peer, 2015; Hadlington & Murphy, 2018; Hadlington & Parsons, 2017; Halevi et al., 2016; Pahnila, Siponen, & Mahmood, 2007; Shahri, Ismail, & Mohanna, 2016; Siponen, Adam Mahmood, & Pahnila, 2014; Siponen, Pahnila, & Mahmood, 2007; Siponen, Pahnila, & Mahmood, 2010; van Schaik et al., 2017; Whitty, Doodson, Creese, & Hodges, 2015; Zhang-Kennedy, Chiasson, & Biddle, 2016).

This reliance on self-report extended to studies of compliance with organisational security policies, with many studies focussing on users' intention to behave securely or to comply with a security policy or requirement (Bulgurcu, Cavusoglu, & Benbasat, 2010; Djajadikerta, Roni, & Trireksani, 2015; Tejaswini Herath & H. R. Rao, 2009; Tejaswini Herath & H Raghav Rao, 2009; Kajtazi, Bulgurcu, Cavusoglu, & Benbasat, 2014; Kim, Yang, & Park, 2014; Siponen et al., 2014; Siponen et al., 2007; Yazdanmehr & Wang, 2016). A small number of studies examined users' intention to adopt a specific security technology (Herath et al., 2014; Johnston & Warkentin, 2010) or likely future behaviour using scenarios where participants predicted their own likely responses to security-breach scenarios (Hu, Xu, Dinev, & Ling, 2011).

Finally, a small number of studies included in the review looked at the seeking of security information (Dang-Pham, Pittayachawan, & Bruno, 2017a, 2017b; Wang, Xiao, & Rao, 2010) or intention to share security information (Safa & Von Solms, 2016).

3.4.2 Protection motivation theory and theory of planned behaviour studies

Of the 47 papers reviewed, 30 drew on either protection motivation theory (PMT) or the theory of planned behaviour (TPB) in some form or other - for instance, by manipulating perceived threat (Boss et al., 2015), measuring various PMT factors and correlating with a security behaviour (Chou & Chou, 2016; Siponen et al., 2014; Siponen et al., 2007), or using attitudes, social norms and self-efficacy to predict an intention to enact a security-related behaviour (e.g. (Bauer & Bernroider, 2017). A number of studies combined elements of PMT with other models such as TPB (e.g. Hu et al., 2011; Herath & Rao, 2009). In the following sections, we examine the key findings according to the core constructs of each theory.

3.4.3 Protection Motivation Theory: Threat appraisal

According to protection motivation theory, threat appraisal is construct of two elements: a person's perceived vulnerability to a threat, and the perceived severity of that threat. Some studies (e.g. Siponen et al., 2007) combine the two elements into a single 'threat appraisal' construct. A substantial proportion of the studies use intention (or actual) compliance to organisational security policies as a proxy for information security behaviour (e.g. Herath & Rao, 2009a). Often these studies also incorporate aspects of general deterrence theory (GDT), with the potential penalties for policy violation divided into the severity of the penalty and the probability of detection, echoing severity and vulnerability from PMT. For instance, Herath & Rao (2009) found that while the likelihood of detection was positively linked to policy compliance intention (ß = 0.26), the opposite was true for severity of penalty, which had a negative association with compliance intention (ß = -0.21). The study by Chou & Chou (2016) on PMT and security also found a boomerang effect, but for perceived vulnerability, with increased vulnerability associated with more problematic Internet behaviours (ß = -.19). Herath & Rao (2009b) divided 'threats' into those posed by the potential security breach (severity and vulnerability as per PMT), which would in turn lead to a 'security policy attitude', which then predicts compliance intention (considering other factors such as self-efficacy and social norms, and punishment severity and detection likelihood). They found weak support for their model - although threat appraisal did predict the level of concern about a security breach, it only predicted 6% of the variance in that variable, and the links were either weak (severity: ß = 0.19) or not significant (vulnerability). The authors conclude that, "response cost and security concern did not significantly contribute to predicting compliance intentions" (p. 118). In line with their other work (2009a), they find again that a more severe penalty for non-compliance backfired, and led to lower compliance intentions. In a later study (2012), Herath and colleagues found a significant link between threat appraisal (in the form of risk perception of email) and intention to adopt an email authentication system (ß = .30).

Four additional studies in the review (Siponen et al., 2007; Siponen et al., 2010) examined security policy compliance and protection motivation theory alongside general deterrence theory. It should be noted that it is possible that these papers are analyses of the same dataset because they report similar data / constructs, and the description of the sample sourcing is similar across the studies. The authors of the papers were contacted for clarification, but were not able to recall if the data came from the same dataset. For this reason, the results should be treated with some caution.

Pahnila et al., (2007) studied employees of a single company (n = 245) in Finland. They found a weak but significant relationship between threat appraisal (not sub-divided) and attitudes towards security policy compliance (ß = 0.27). Siponen et al. (2007) found a reliable but weak positive relationship between threat appraisal (severity and vulnerability again combined into a single construct) and intention to comply (ß= 0.24) studying 917 employees of 4 Finnish companies. Siponen et al. (2010) report results of 917 employees from 4 Finnish companies, and report a weak (ß= 0.12) relationship between threat appraisal and intention to comply with IS policies. Finally, Siponen et al., (2014)report significant but very weak associations between both perceived severity (ß= 0.07) and vulnerability (ß= 0.06) and intention to comply in employees (n = 669) from four Finnish companies.

Some studies have found an somewhat more reliable association between **perceived severity** of a threat and either problematic / risky behaviour (Chou & Chou, 2016) and the taking of protective action (Boss et al., 2015), although the relationship is relatively weak. In the case of Chou and Chou, the relationship was (ß = .34, n = 505). Boss and colleagues reported a direct effect of perceived severity of ß = .27 (study 1) and a non-significant relationship (Study 2) unless in their 'high fear' condition.

3.4.4 Indirect and associated measures for threat appraisal

There have been a number of studies that address PMT threat appraisal in an indirect way. One study (Wang et al., 2010) found that searching for information security knowledge (measured using AOL search queries) was higher after a network attack. They suggest that press coverage of network attacks increases users' threat perception (presumably both severity and vulnerability). In a qualitative analysis of employees discussing phishing simulations, Williams et al. (2018) noted perceptions of reduced vulnerability in the workplace compared to the home, and faith in technical protection, led to a reduced perception of risk.

While van Schaik et al., (2017) did not explicitly measure PMT constructs, their study examined perceptions of risks (called 'hazards') online and precautionary behaviour. Their measure of risks included not only hazards encountered online (e.g. malware), but also the consequences of those hazards (e.g. severity, dread, catastrophic potential). They found that both hazards and the consequences (particularly severity) predicted perceived risk, but they did not reliably predict self-reported precautionary behaviour.

Martin et al. (2018) conceive of threat appraisal as a signal detection task, where recognition of the appropriate threat signals increases the chance of taking protective action. In their study using mTurk workers, they found that generic phishing emails were easier to accurately spot than spear phishing emails. They did not measure threat appraisal per se, but did not that *recognizing* a threat is a key precursor to deciding on the appropriate response.

3.4.5 Manipulations of threat appraisal and fear

The most common method to manipulate threat appraisal is to provide users with information about cybersecurity threats versus a control condition, and thence to measure either intention to take protective behaviour, attitudes, or actual behaviour. Although they did not compare against a control condition, Johnston & Warkentin (2010) did provide information designed to increase fear of spyware. However, they found that by increasing perceived severity this way, they reduced people's perception of the effectiveness of acting against the threat, and their own belief about the likelihood of being able to successfully protect themselves (note: perceived vulnerability did not have any effect on these variables). However, they argue that if fear messages are combined with measures to improve people's sense of control over threats, this will lead to greater intention to take preventative action.

Similarly, Mamonov & Benbunan-Fich (2018) manipulated (but did not measure) 'awareness of information security threats' using either cyber-security or control stories to successfully predict both password strength and willingness to disclose sensitive personal information (increasing the former, and reducing the latter).

Boss et al., (2015) conducted two studies to examine the ways in which threat appraisal links to both fear and protection motivation. In the first study, MBA students were randomly allocated to either a high or low fear messages group based around the importance of (and impact of not) back-ups. They were also asked to keep a record of their backing up over a period of eight weeks. In this first study, the researchers found that while the high fear message did not influence perceived severity or vulnerability, they did increase reported fear in the expected direction, which in turn predicted both intentions to back up, and actual backup behaviour. In their second study, Boss and colleagues manipulated the severity of the fear message in a pop-up window designed to inform participants about a catastrophic or non-serious malware infection on their computer, and measured either clicking 'OK' (to remove the threat] or 'x' to close the window. They found that in the high fear condition, both threat severity and fear were increased, as was acceptance of the message (ie. Clicking 'ok' rather than closing the dialog box). This effect was not, however, strong enough to be maintained across all conditions.

This section of the review therefore concurs with the conclusion of Mayer et al. (2017) that **threat appraisal** represents a non-reliable (at worst) or weak (at best) predictor of information security behaviour. However, there may be some mileage in the provision of

information that both increases awareness and the effectiveness of a response. In keeping with the same review, we also find that increased severity of penalty has the potential to 'boomerang' on the IS community, leading to reduced compliance.

3.4.6 Protection Motivation Theory: Coping appraisal

The second element of protection motivation theory is the individual's appraisal of their likely response to a threat, both in terms of the likely efficacy of the response and their own ability to bring complete the required response (also called 'self-efficacy'). These two factors are commonly referred to as 'response efficacy' and 'self-efficacy'. Many of the same studies outlined above also measure (and model) these two variables in predicting people's behaviour, attitudes or intentions. In later PMT models, the cost of completing the response was also factored into the model. Johnston and Warkentin (2010), in their study of fear messages and intention, found significant relationships between response efficacy (ß = .21) and self-efficacy (ß = .19) on behavioural intention. Chou and Chou (2016) found no significant effect for response efficacy, a small effect for self-efficacy (ß = .15) and a stronger effect for response costs (ß = .32) on problematic internet security behaviour. Boss et al. (2015) found an effect for response costs on security behaviour (ß = -.67 and -.14 for study 1 and 2 respectively), and mixed results (mostly non-significant) for response- and self-efficacy. Herath et al. (2012) studied the adoption of an email authentication service as a coping mechanism against cybercrime. They synthesized a model of technology adoption and PMT to conceptualize response efficacy as 'usefulness' and 'usability', and response cost as 'privacy concern'. They found that that perceived usefulness, ease of use (ß = .20, and ß = .27) and privacy concern (ß = -.21) predicted overall coping appraisal, which in turn predicted adoption intention (ß = .49). Herath & Rao (2009b) report significant associations between response efficacy (ß = 0.29), self-efficacy (ß = 0.15), and response cost (ß = -0.195) and security policy attitudes, and a direct relationship between self-efficacy and compliance intention (ß = 0.17). The various papers by Siponen and colleagues (2007, 2010, 2014) also measured self-efficacy and response efficacy, and reported a weak relationship between self-efficacy (ßs ranging from 0.09 to 0.31) and policy compliance intention, and mostly non-significant relationships between response efficacy (ß range -.002 > 0.06) and policy compliance intention. In a study of 263 users of a health information system in Iran, Shahri et al. (2016) report a positive relationship (ß = .19) between security self-efficacy and security effectiveness (although they do fully explain how 'effectiveness' was measured). While not a study of PMT, Bulgurcu, Cavusoglu, & Benbasat (2010) included 'perceived cost of compliance' and 'work impediment of compliance' in their study, and although they did not connect either to intention, they did find a relationship between the impediment to work through compliance and the perceived cost of compliance, which in turn then predicted negative attitudes towards security policy compliance (ß = -.15).

This section of the review therefore concurs with the conclusion of Mayer et al. (2017) that **coping appraisal** represents a reliable but weak predictor of information security behaviour, with self-efficacy more reliable a predictor than response efficacy beliefs, and response costs being a promising addition to predicting people's intention to take

precautionary information security behaviour. This suggests that a combination of training and more effective, usable security systems would be the most effective way to improve human aspects of cyber-security.

3.4.7 Theory of Planned Behaviour and Security Behaviour

A typical study of information security that utilises the theory of planned behaviour would measure people's attitudes towards a specific security behaviour or behaviours, the social norms about those same behaviours, people's control over whether or not they completed the behaviour, and their intention towards carrying out the behaviour. Some will also measure actual behaviour (often through self report).

Seven papers in the review dataset explicitly tested constructs from the theory of planned behaviour and information security, and a further four took aspects of TPB but combined / repurposed into different models. The majority studied security policy compliance intention rather than other security behaviours.

Attitudes are typically found to be a reliable predictor of security compliance intention, with beta weights ranging from 0.25 - 0.3 (Bulgurcu, Cavusoglu, & Benbasat, 2010; Kim et al., 2014) to between .41 and .49 (Bauer & Bernroider, 2017; Siponen et al., 2014). There was one non-significant relationship between attitudes and policy compliance intention reported in the papers studied (Herath and Rao, 2009b). Alshboul & Streff (2017) didn't measure attitudes directly, but rather measured the perceived usefulness of the security policy, which strongly predicted satisfaction with the security policy (ß = 0.66). Attitudes also predicted intention to share security information, with beta weights ranging from 0.41 (Dang-Pham et al., 2017b) to 0.7 (Safa & Von Solms, 2016), and intention to adopt an email authentication service (ß = 0.49, Herath et al., 2014).

Djajadikerta et al. (2015) presented users with four vignettes of destructive insider acts, and asked their participants to score their attitude towards the behaviour, and intention of carrying out a similar behaviour in the future. Their overall model found that attitudes predicted intention well (ß = 0.45), but the strength of this relationship differed substantially between the scenarios, with attitude being the weakest when predicting a naive mistake, and strongest predicting intentional destruction (ßs 0.17 to 0.63).

Perceived behavioural control (PBC) was measured less in the studies within this review. In part this may be because there is overlap between PBC and self-efficacy, and many of the studies included self- and response-efficacy as variables (see above for a reporting of these results). For instance, Bulgurcu, Cavusoglu, & Benbasat (2010) included a variable called 'self-efficacy to comply' in their model, and found that it predicted (ß = 0.22) intention to comply. Safa & Von Solms (2016) included PBC in their model to predict security knowledge sharing intention (ß = 0.69). Djajadikerta et al. (2015) divided PBC into two aspects - control over outcomes and control over resources / skills needed to conduct a behaviour, but neither were consistent in predicting intention to conduct destructive insider

behaviour, although perceived control over an outcome did predict intention in the naive mistake vignette (ß = 0.37), and perceived control over resources did predict intention in the 'dangerous tinkering' vignette (ß = 0.32). Herath et al. (2014) did not measure PBC, but did use 'perceived ease of use' which could act as a proxy for control (at least partially). They found a relationship towards the weak side between perceived ease of use and attitude towards use of a security technology (ß = - 0.27).

The '**subjective norms**' element of TPB has been relatively well studied (in comparison to PBC). In their studies of security policy compliance, Siponen et al. (2014) found that social norms (also termed 'normative beliefs) predicted intention to comply (ß = 0.33). This link was reported as 0.45 in their 2010 paper, the strongest link to compliance intention of any variables measured in their study. Bauer & Bernroider (2017) and Bulgurcu, Cavusoglu, & Benbasat (2010) also studied intention to comply with security policies, and reported smaller, but significant, relationships between normative beliefs and policy compliance (ßs = 0.23 and 0.29 respectively).

This link is also found in other information security contexts - including intention to share information (Safa & Von Solms, 2016), although Dang-Pham et al. (2017b) did not find a significant link between subjective norms and security information sharing, but their measure of subjective norms was only two items long. Djajadikerta et al. (2015) report that subjective norms predict people's intention to undertake a range of destructive insider actions (ß = 0.24), but not 'dangerous tinkering'.

Finally, Yazdanmehr & Wang (2016) studied three types of norms to predict security policy compliance: 1) descriptive norms (what employees think most people actually do); 2) injunctive norms (what employees think should be done); and 3) subjective norms (what employees think *important others expect* them to do). They found that both injunctive and subjective norms (but not descriptive) predicted 'personal norms', which in turn predicted compliance behaviour (ß = 0.36).

In conclusion then, there is sufficient evidence to suggest that all three elements of the theory of planned behaviour - attitudes, control and social norms - are indicators of cyber-security behaviour (or more precisely, intention to behave). In particular, attitudes were the strongest predictor, with the self-efficacy element of perceived behavioural control being weaker, but also reliably linked to (security behaviour) intention. This conclusion also mirrors that made by Mayer, Kunz and Volkamer (2017) in an earlier review.

3.4.8 Other human-aspects of cyber-security

A number of additional studies appeared in the review, but did not explicitly address either PMT or TPB variables. In some cases these studies were focussed on development of methodologies (e.g. validation of a scale (Egelman et al., 2016); mouse tracking (Kelley et al., 2018)), intervention and awareness raising (Zhang-Kennedy et al., 2016), studies of

demographics in some form (e.g. Anwar et al., 2017; Oliveira et al., 2017) or other approaches to analysing risk (Ovelg et al., 2017). These studies will be now summarized.

3.4.9 Individual differences

A number of studies (Bauer et al., 2017; Egelman & Peer., 2015; Herath & Rao, 2009a, Herath & Rao, 2009b) reported that gender was included as a control variable in their studies. In the case of Herath and Rao (2009a), females had higher security compliance intention (the direction is not reported in their second study). Bauer et al. (2017) included both age and gender as control variables, and found no significant link to policy compliance intention. Egelman and Peer (2015) included income level, education and gender as control variables in their study predicting privacy and security attitudes. They report few reliable significant associations between demographics and privacy concerns or information security awareness or behaviours. Bulgurcu et al. (2010) included education, IT knowledge, organisational size, industry and information intensivity of the company in their study of compliance intention, and found no significant links. Canfield, Fischhoff and Davis (2016) included age, gender and college education as demographics in their test of phishing detection, and found no effect on the ability to detect phishing emails. Hu et al. (2011) included age and computer use as control variables to predict intention to violate security policies, and found no effects.

Anwar et al. (2017) explicitly tested for gender differences in cyber-security behaviours, and found that women scored lower on reported protective cyber-security behaviours, mostly explained by differences in reported computer self-efficacy and prior computer skills. They do report finding the same pattern of relationships between PMT variables and security behaviour, regardless of gender, suggesting that existing models are not influenced by gender, only scores on specific variables (in particular self-efficacy).

In their study of phishing susceptibility, Halevi, Lewis and Memon (2013) included gender as a control variable, and found that for women there was a medium correlation ($r = .5$) between susceptibility to a simulated phishing attack and neuroticism, but not for males. They did not find any relation between security expertise and susceptibility, and did not report any differences in susceptibility between males and females. In a later study. Halevi et al. (2016) considered age and gender in more detail in four countries (USA, India, UAE and Ghana) where they also collected data on personality and security behaviour and security self-efficacy. Culture (i.e. country) did not significantly predict security behaviour or self-efficacy. Males scored higher in computer security self-efficacy, but this did not translate to more secure behaviours (where males and females scored similarly).

Whitty and colleagues (2015) predicted that older adults would be more likely to share passwords, but they actually found the opposite - younger people were significantly more likely to report sharing passwords. In a study of phishing susceptibility and age, Oliveira et al. (2017) found that age determined how vulnerable users were to particular forms of attack, with younger users more likely to click on attacks based on scarcity and authority,

and older users more likely to click when the phishing email relied on reciprocation or liking as its influence technique. Across all types of attack, the authors report that older women were the most vulnerable group (although 40% of participants clicked on at least one phishing email in the study).

Finally, Ovelgönne et al. (2017) studied the number of executable binaries on 1.7m computers and malware attacks detected using an anti-virus system. They found a relationship between the number of binaries on a machine and the number of malware attacks, a relationship that was particularly strong for software developers.

In summary then, there is little to suggest that demographics contribute substantially or reliably to either cyber-security attitudes, intentions, or behaviour. There is some evidence that males have higher computer-related self-efficacy, suggesting that they report feeling more sure that they could take appropriate protective actions. However, this doesn't translate into men acting in a more secure manner in most of the studies in the review. In terms of age, there is evidence that both older adults and young people may be particularly vulnerable to security threats, but for different reasons, suggesting that interventions may need to be designed differently for different segments of the population.

3.4.10 Personality and other traits

Relatively few studies included personality as a variables in determining cyber-security behaviour or attitudes. As noted above, Halevi, Lewis and Memon (2013) found an association between neuroticism and susceptibility to phishing, but only for women. They found no other links between personality (openness, conscientiousness, extraversion and agreeableness) and susceptibility. Halevi et al. (2016) also used the 'big 5' traits to predict security behaviour and self-efficacy, but found only relationships between increased conscientiousness and reported security behaviour, and openness and reported security self-efficacy. Egelman and Peer (2015) tested the relationship between the 'big 5' personality traits (study 1) and decision making (study 2) on a range of privacy and cyber-security measures. Overall, the ability to predict any of their measures was lower for personality (under 10% of variance explained) than when using decision making approach (between 10-27% of variance). In their third study they found that people who were more inquisitive, rational decision makers who tended to focus on long-term outcomes reported higher security protective behaviours, while those with an avoidant decision making style (i.e. procrastinators) tended to have fewer good security practises.

Hu et al. (2011) included low self-control in their model of intention to violate company security policies. They found that people with low self-control tend to focus more on the perceived benefits of violation (which in turn predicts intention to violate policy). Hadlington & Parsons (2017) found that self-report scores on measures of excessive internet use and major cyberloafing (including items such as visiting adult or gambling websites at work) predicted scores on a measure of cyber-security attitudes and behaviour.

Finally, two studies considered the potential for ongoing tasks or multi-tasking to influence security behaviour. Hadlington & Murphy (2018) found that people who reported being high media multitaskers (e.g. watching TV while browsing web or responding to email) also reported more risky security behaviours (such as sharing passwords or using cloud storage for personal, sensitive files). Williams, Morgan and Joinson (2017) found that people were more likely to accept fake update messages when engaged in a cognitively demanding task.

3.5 Conclusion

The majority of the studies reviewed reported relatively weak links between 'human aspects' variables (e.g. threat appraisal, coping, attitudes) and either cyber-security behaviour or intention to behave. In many cases, the studies were overly reliant on self-reported behaviour or intention. It should also be noted that there are significant differences between the behaviours studied - complying with the security policy of an organisation is *not* the same as (not) clicking on a phishing email, which in turn is different from sharing a password with others, which in turn is different from installing updates and patches. This strongly suggests that as a field cyber-security needs to more carefully differentiate between *compliance* behaviours, *risky actions*, and *protective behaviours*. In many cases the causes of good (or poor) security behaviour are likely to differ - we may not comply with a policy because it interferes with our ongoing work, while we click on a phishing email because it is difficult to detect, and we do not install an update because it both interferes with our work (i.e. it is *costly*) and because we lack awareness of the risk of not doing so.

Some more reliable indicators did emerge from the review. From protection motivation theory, there is more evidence that improving **coping appraisal**, specifically both self- and response efficacy, improves security behaviour and intentions. This is partially supported by work on the theory of planned behaviour studies that also found that increased **self-efficacy** was associated with better security outcomes. This suggests that there is a role for awareness training within organisations that stresses not only the risks, but also provides employees with the appropriate skills and knowledge to mitigate such risks. Importantly, campaigns should also stress how taking the appropriate action is effective. This also suggests that work on usable security (that seeks to make good cybersecurity behaviour easier for users) can help by improving users' coping ability in the face of myriad threats.

While attitudes towards security did reliably predict a range of security intentions, there was little evidence to suggest that **increasing the perceived threat** posed by cyber-security threats would be a productive endeavour. Indeed, in some cases increasing the threat posed by either punishment for non-compliance or vulnerability to external threats served to **boomerang** and reduce security behaviour. It may be that combining increased threat appraisal with increased skills to respond may lead to better outcomes, but the evidence base is not yet sufficient to reliably make that claim at present.

For practitioners, we make the following recommendations:

1. Conduct a detailed analysis of the security behaviour you wish to address. Use workgroups / focus groups to identify whether a behaviour is not being conducted due to ability, motivation or another factor before designing any intervention.
2. Design interventions only following this detailed analysis of the reasons behind a specific (non) action.
3. Assume that increased severity of punishment or merely re-iterating a threat may have a counter-productive impact and increase vulnerabilities
4. Focus on improving users' coping skills - either by training, understanding of the effectiveness of simple actions, or by making security easier to do.
5. Know what success looks like - be able to identify what metrics or measures will change in response to any intervention ahead of time (and ideally measure before and after the intervention).

4. Evidence review: Beyond surveys - qualitative and mixed-method studies

There are examples of **survey questions put to employees in organisations were generated from in-depth qualitative work** with a subset of that population. Ashenden (2015) examined employee attitudes to information security, and the link to behaviour. She extracted relevant constructs from existing social psychology literature, and probed this interviews with 12 employees in a major UK insurance company using a structured qualitative method (repertory grid). From these interviews, she concluded that the most influential factor of 'good' security behaviour was self-efficacy - employees who thought that their behaviour had an impact on the organisation's security reported better behaviour and seeking more information, whereas those who thought their behaviour had little or no bearing so no need to. She constructed a survey that mapped previously validated survey questions on to information security behaviours, and conducted that survey with 474 employees of that organisation. The survey results showed a significant correlation with self-reported security behaviours. She concluded there two segments of employees – the *'I Can Handle It Group*' and the '*It's Out Of My Control Group*' - and that this attribution links attitudes and security behaviour - and different attributional framing is required to persuade members of both groups towards a specific security behaviour. This effectiveness of targeted persuasive messaging was validated through an intervention study in a second organisation, where employees were profiled in a survey, followed by a targeted intervention (attributional framing was used to tailor persuasive messages to both groups). Subsequently, she performed an online experiment with 201 employees in a different organisation, tailoring the security awareness method to the attitude of half the respondents, and found that those who received messages tailored to their control attitude were more likely to carry out the mandated security behaviour than those who had not. This research follows a careful multi-stage, mixed-methods approach: established psychological constructs are applied to information security, but selects those constructs as a result of a qualitative enquiry, and validates the concepts in a study within the same organisation, and then another one.

A multi-stage development process of understanding drivers of security behaviours was also pursued by Beautement et al. (2008, 2016). Beautement et al. (2008) initially interviewed 14 employees of one organisation. They analysed the transcripts using a qualitative method called Grounded Theory coding (Strauss & Corbin, 1990), and concluded that although interviewees reported not complying with a range of different security policies and methods, the common driver was when security reduced productivity too much - when their Compliance Budget was exhausted. Beautement et al. (2016) report a series of organisation-based studies that built on this approach - eliciting descriptions of specific security behaviours via **interviews with a sub-group of employees** (max. just over 100) 3 different organisations. The interviews probed the circumstances surrounding

non-compliance such as tail-gating, password sharing, bypassing of access control, and use of removable media). A range of scenarios emerged in which employees' security behaviour diverged from policy, and a number of factors that drove that behaviour. These scenarios were used as questions in a survey that was then completed by hundreds of employees. The constructs across the specific behaviour scenarios were *level of risk-awareness* and the *level of risk taking*. Beautment et al. concluded from the interviews that most employees were well aware of security risks, but did not comply because badly designed security meant following the rules would have lowered individual and organisational productivity. This hypothesis was tested in the surveys, while geographical location and business department were control variables. The results revealed 'hotspots' where employees indicated they knew the risks - yet reported they did not follow the policy. Based on statement in the interviews, the authors say such high-risk awareness/low compliance 'hotspots' are caused by security solutions that interfere with employees goals and tasks. Employees know they correct security behaviour, but sense that following them would lead to a reduction in their personal productivity that they think would reflect badly on them, and thus put productivity above security. In a further analysis of the interview and survey data, Kirlappos et al. (2015) point out the organisations in these cases studies were 'tacitly complicit' in their employees' decisions since management by tolerating non-compliance justified by 'productivity first' arguments. The authors also found that employees did not ignore security altogether, but more often than not devised make-shift security solutions (*'shadow security'*) to manage the risks they recognised as best they knew. The solutions they devised were often not secure, but Kirlappos et al. argue that - rather than 'stamping out' this type of non-compliance - security should see this as an opportunity to engage and create secure versions of workable solutions. These grounded, qualitative studies have produced one two main insights:

1. That non-compliant behaviour is largely triggered by the situation (conflicts with other organisational demands, such as productivity) rather than individual traits or dispositions, and

2. That engaging with employees to co-design and/or negotiate workable security solutions may be a more successful route to effective security than the hitherto prevalent 'diagnose and target' approaches.

Consequently, a range of techniques for creative security engagements have been developed. Dunphy et al. (2014) go describe creative security engagements encouraging participants (employees in the company context or consumers or citizens in wider engagement) to reflect on their *environment*, the *emotions* they feel, the *constraints* they experience, the *pressures* that they undergo as well as the *actions* and the *tasks* that they

perform when generating and sharing information. The EU Trespass Project[1] developed a One particular technique for creative engagements using Serious Lego for physical modelling of information security threats was developed by This type of physical modelling bridges the space between the typical diagrams

(flow-charts and UML diagrams for example) that security practitioners commonly work with, and the everyday practices of the consumers who are affected by security design. Heath, Hall & Coles-Kemp (2018) report a successful case study where this method to model security for a home banking application, which identified areas where human intervention and support needed to be provided to make security work overall. These **projective techniques** take employee or citizen involvement (Coles-Kemp, Ashenden & O'Hara) a step further by grounding the discussion of security behaviour in the representations of daily activity. This not only identifies security conflicts with goals and tasks - similar

These studies provide examples of different ways of engaging with employees, consumers and citizens on security. They are part of a growing trend in research, moving away from the mechanistic approach of looking for traits within individuals that are conducive to desired security behaviour, or trying to change behaviour by addressing or tweaking those traits.

4.1.1 Studies with other stakeholders - developers

Since many security vulnerabilities are in the software we use today, developers clearly have an important role to play. Security experts often bemoan that developers repeat well-known mistakes and continue to re-use code that known vulnerabilties - and assume this is because they "don't care". But Zurko & Simon (1996) pointed out in one of the earliest papers on human behaviour in security, not only end-users (such as employees and consumers) have difficulty doing what security experts want them to do - with rapid development of technology and specialisation, other technical stakeholders such as developers and system administrators struggle.

Over the past 5+ years a strand of research that takes a more collaborative and constructive approach has explored why known vulnerabilities are reproduced in new code. Most notably are the studies by Fahl et al. 2013 and Acar et al. 2016. Fahl et al. tracked SSL vulnerabilities in code to the developers that had produced them, and invited them to take part in interviews and surveys to establish how the vulnerability had occured. It was notable that of the X developers who were contacted, a large number (N) were willing to provide information, but only 13 were interviewed because companies refused permission for them to do so. From the interviews, Fahl et al. found that developers had little to no

[1] https://www.trespass-project.eu/

security training and were under extreme pressure to complete the app quickly - and that was the reason for the mistakes that led to vulnerabilities. The researchers developed a solution (certificate pinning) and provided code that developers could use to avoid those mistakes being repeated. The results has been a notable reduction in those vulnerabilities since, and this study received was runner-up in the 2014 NSA's Best Science of Cyber Security Paper Award.

In a study that received the NSA's Best Science of Cyber Security Paper Award in 2017, Acar et al. recruited 54 developers to develop an Android app to investigate whether a common complaint of security experts - that vulnerabilities come from inexperienced developers who copy and paste code from repositories such as Stack Overflow. Developers were either given official documentation, access to a code repository (Stack Overflow), a textbook, or a free choice of which resources they wanted to consult. They judged the quality of the code produced based on functional correctness, and the security based on whether 4 common vulnerabilities had been avoided. ⅔ of the developers who used Stack Overflow or the book managed to produce a functionally correct solution in the allocated time, whereas only 40% of those using official documentation did. In terms of the security tasks, the results were reversed - those using official documentation produced the most secure code, those using the Stack Overflow the least. A traditional security response to this results would be "use of Stack Overflow should be forbidden." But clearly, the productivity price developers and their organisations would pay would be a hefty one. The researchers conclude that given time pressures under which developers have to work, investing in ensuring code in repositories is secure would be the most efficient way to support developers in producing functioning and secure code.

Producing code without known vulnerabilities improves security; however, since research reviewed in the previous section (Beautment et al. 2008, 2016) shows that non-compliance by employees is often induced by unusable security, developers and software development organisations should also ensure the security within their products is usable. The US National Institute for Standards and Technology (NIST) has been aiming to foster usble security through a series of studies and workshop activity since 2011; one repeated request it received was for examples of how to deliver usable security. To produce those examples, Caputo et al. (2016) conducted 3 case studies in with project teams in major software development organisations that claimed to have 'secure and usable' products, and studied one specific product in each. The three organizations studied were selected to explore 3 aspects:

1. why each organization added usability and security elements to its software

2. development process how and where they added them,

3. how the organization determined that the resulting software was usable and secure.

The case studies looked for evidence of 3 hypotheses: that usable security could be delivered because of:

1. a “key individual” who promoted usability and/or provided specific knowledge on making security usable,

2. an “experienced team” that had built knowledge on how to make security usable, or

3. “incentives” provided by the organisation.

The authors total of 23 interviews with individual members of the project teams - developers, product managers, and senior managers from 3 large organisations who developed large number of software products, with business units in a number of locations.

The authors conducted a series of pilot interviews in a 4th organisations, and revised their questions because they were too long, and somewhat repetetive. The final set of questions put to each of the 3 groups are reproduced in the paper.

For each interview, detailed notes were taken by 3 team members, and subsequently reviewed, reconciled, and sent to the interviews for review and comments. The final transcripts were analysed using Grounded Theory coding by 3 team members individually, followed by 2 rounds of review and reconciliation.

The results found the following in common:

1. They had small development teams (even though all three were large companies)

2. They followed an agile-inspired, informal development process (developers followed the spirit of agile, rather than a specific process)

3. They did not have defined criteria or measurements for usability or security.

4. They did not perform formal usability testing (so could not demonstrate that usability improved)

5. They did little or no formal security evaluation (so the teams also could not assess if better usability improved security), but interviews talked as if formal security evaluations were performed.

6. They did not use business modeling to determine impact of usability and security.

7. The concept of ‘tradeoffs’ was mentioned by most interviewees - specifically between usabililty and security - but in the absence of any metrics or business modelling, there

was no basis for performing a trade-off. Smith & Sasse (1996), in the introduction to the Special Issue in which the study was published, see this as clear evidence that the tradeoffs are fake - they are a meme to justify security not bothering with usability. Economics and incentives were the key factors that initiated a push for usable security. Because those who deliver secure applications with poor usability generally don't bear the resulting cost, complaints about unusable security are relayed to developers and then often ignored. Developers didn't understand the impact of lack of usability on individual performance and wellbeing.

4.1.2 Studies with other stakeholders - security experts

Most developers confused their own 'knowledge about the product' with 'knowledge about how it was used by customers', thought usability knowledge and methods as "common sense" (which they naturally felt they possessed in abundance), not as a specialist discipline with relevant knowledge and methods

Despite not having usability training in education or on-the-job, and not knowing how a product might be used "in the wild," developers believe they know a product well and know how to improve it.

Only incentives hypothesis supported

- Massive cultural divide
- Absence of criteria and metrics

What if the number of user complaints was a metric affecting the performance reviews of the soft ware's designers or developers? Or, what if usable security was defined to include not only features but also lack of failures? Developers might then take increased ownership of failures in usable security and eventually take steps to design usable security into products during early development stages.

"Developers need a basic understanding of the complexity of human capabilities and limitations, as well as human activity and productivity, to appreciate the complementary expertise offered by a usability expert."

"Mutual respect between developers and usability experts might encourage the developers to see through diff erent eyes and observe that unusable security isn't secure because users will find workarounds to get their primary tasks done that will reduce intended security."

"What we saw is that textbook or mandated soft ware development processes weren't followed routinely or even valued; their adoption occurred only when developers valued them and were personally motivated to use them."

"What if the value of usable security were monetized so that the cost of putt ing usability analysis into the process is weighed against the expense of the help desk support needed when products aren't usable?"

"An organization can motivate developers by providing incentives or disincentives, or it can implement a process that ensures that usability is considered, regardless of whether people are motivated. If developers knew there was a formal usability evaluation, and products not meeting the usability threshold wouldn't get released, they would very likely pay a lot more attention to the test criteria." Applies to security as well?

Much talk about security shifting from compliance-based to risk-based – but in absence of metrics, that is hollow talk.

False tradeoffs that are prevalent

- Usability is common sense; no experts needed.
- If we make a product harder to use, then it's more secure.

There's always a tradeo ff between usability and security.

4.1.3 Studies with other stakeholders - security experts

Clearly, security experts themselves are a key stakeholder, and their behaviour has a significant impact on security outcomes. The two groups of experts that have been studied empirically using social scientist are Chief Information Security Officers (CISOs) and Cyber Security Incident Response Teams (CSIRTs) and Computer Emergency Response Teams (CERTs).

Ashenden & Sasse (2013) interviewed 5 CISOs, and analysed them organisational behaviour theory. The results show that CISOs struggle to gain credibility within their organisation due to a perceived lack of power, confusion about their role identity, and their inability to engage effectively with employees. They concluded that CISOs need to acquire skills to communicate effectively with employees and engage them in security initiatives. Ashenden & Lawrence (2016) conducted an Action Research[2] study with CISOs and other security professionals working in organisations. 6 scoping interviews revealed there was a

2 Ashenden and Lawrence provide an excellent summary of the Action Research process, paraphrased here: it is an applied research method that aims to analyze and achieve practical change in a particular environment, such as an organization. Researchers bring state-of-the-art knowledge (previous research) and their academic experience to structure and guide the process. .The process itself is an iterative one, to address a (set of) specific organizational issue(s). As with other participative methods (such as participatory design) participating employees become part of the research process and their views and organizational knowledge contribute to the final specific solution. The research output has implications beyond the immediate project; participants have a chance to review and comment on the research output created by researchers (and these are supposed to be noted if they diverged).

lack of engagement and trust between security practitioners and staff, who saw security as a hurdle to overcome. Developers particularly felt judged by security experts, and tended to avoid them for fear of them 'shooting their baby'. The researchers aim was to help security staff to develop relationships with other staff, and increase engagement and trust. The first step was to help them understand why staff acted the way they do. They ran 5 focus groups with 4-8 staff each, allowing them to express problems with security as they had experienced them. A key result was that staff still saw security practitioners as 'not on their side' and the source of problems as far as everyday work was concerned. As a result, they did the minimum required, and never volunteered information or sought help. The researchers then arranged 3 3-day workshops with 18 security practitioners in total. In the workshops, the secrurity practitioners learnt first of all learnt why staff acted the way they did. The researchers then taught (and made them practice) skills to engage and support staff instead (drawn from research on how to achieve adherence to medical treatment plans, question and conflict resolution skills from counselling, social market theories of exchange and influence, and how to design behaviour change interventions).

The workshops were captured through field notes, the output produced by practitioners, and daily feedback sheets completed by them. These were then analysed using qualitative analysis methods. One tangible output was the metrics the practitioners themselves developed for assessing the relationship between themselves and staff in future (the number of emails that were exchanged with security, the tone of emails, how early on security practitioners became engaged in the project, and the number of security features that had to be retrofitted into a process or system).

4.1.4 CSIRT/CERT

The name Computer Emergency Response Team is the historic designation for the first

Team (CERT/CC) 1 at Carnegie Mellon University (CMU). CERT is now a registered service mark of Carnegie Mellon University that is licensed to other teams around the world. Some teams took on the more generic name of CSIRT (Computer Security Incident Response Team).

4.2 Conclusion Evidence Review 3

- Scientifically
- General conclusions for practitioners

Stop looking for a psych theory or concept as a silver bullet - fear appeals, PMT, etc.

From ‘fixing the human’ to engagement and organisational learning

“Trust and collaboration are the foundations of functioning cyber security”

Coles-Kemp, Ashenden & O’Hara 2018.

Concepts that are worth investigating - skills and efficacy - awareness and training should focus on those, be integrated, build on each other - anchor in organisational behaviour - link to Thomas’ model + behaviour change model (links also to COM-B model - opportunity is linked to culture, policy and workplace practise).

If you do this (surveys, target your employees) this is how (what does a principled approach look like?) - long-term view, building repeatable, validated measures, aim for constant improvement over time

Stop over- fixating on fixing the individual

- Responsibilization of individuals - victim-blaming (ISO risk management link - fails cardinal principle of assigning responsibilities only to those capable/with resources to so) the learning process in organizations must bebased on the user-centered approach, paying attention to target groups, gender, and culture, which is based on individual knowledge and skills as well as on concrete work connections. The user-centered approach should also enable exchange in informal learning processes in certain social conditions within the organizational setting. The integration of formal and informal mechanisms can enhance the interaction between employees. Frequent interaction is the basis for the formation of interpersonal relationships and psychological attachment to the organization. Since threat analysis, self-efficacy, and response effectiveness have a significant impact on the intention to comply with the IS guidelines, such aspects of emotionalization and motivation should be incorporated into the sensitization to and training of ISA. Scholl, Fuhrmann & Scholl 2018
- fix your security first so it works for your employees, and your business
- fix the organisational culture: leadership, etc. has been under-studied and under-- as outlined in Thomas’ model - link to current efforts to educate/support boards

5. Evidence review: Current Practice

By awareness, training and education, organisations hope to change the security behaviour of their staff. Unfortunately, this expectation is very often not satisfied. Changing behaviour is more difficult than most IT professionals think, because they lack required know-how and often seek simple solutions to complex issues. In order to influence the behaviour positively and to choose effective improvement actions, it is not only necessary to understand theories and models of behaviour change, it is also required to have a set of methods for its study.

5.1 Common metrics used in practice and their issues

According to NIST (National Institute of Standards and Technology, 2008), three general types of security metrics can be identified:

- *Implementation measures* to measure execution of security policy (e.g. compliance with ISO/IEC 27001 or regulations);
- *Effectiveness/efficiency measures* to measure results of security services delivery (e.g. costs of single activities or whole programme, user satisfaction, change in risk exposure); and
- *Impact measures* to measure business or mission consequences of security events (e.g. costs of security incidents, cyber security budget vs. IT budget).

The maturity of an organization's cyber security awareness programme determines the type of measures that can be gathered successfully.

Organizations derive those security metrics usually from statistical numbers, performance metrics, tests/inspections or audit results. While those measures increase accountability and effectiveness, and demonstrate compliance of security controls, they do not provide good enough insights into organisational behaviour and the strategies to influence it (see Table 1).

Source	Example	Issue
Statistical numbers	No. of IT Service Desk tickets related with security	Statistical numbers are often hard to interpret. In the given example, an increase in tickets related with security could mean either that security awareness has dropped and users behave more insecure, or that security awareness has increased and users detect and report more incidents
Performance metrics	No. of staff trained No. of visits of Intranet security page	Performance numbers often look good at first sight, but do not help to understand the organisational performance in a way that informs future strategies. Such metrics are called *vanity metrics*.
Tests / Simulations	Phishing tests Cyber defense simulations Red team vs. blue team	Tests and simulation can give valuable insights into human behavior patterns. But they are very limited to specific situations and do not provide information about strategies to influence the secure behavior.
Audit results	ISO/IEC 27001 PCI/DSS	While audit results are the most complete metrics, todays standards and best practice catalogues do not cover the full spectrum of social and psychological items that influence human behavior.

Table 1 Source of metrics to measure cyber security awareness and their issues

- Jaeger (2018) has performed a literature review of 40 studies and has categorized the variables used by researchers to study the security culture in a framework. This integrative framework categorizes the variables used in the studies in input (antecedents) and output (outcomes) factors:

- Input: individual factors (knowledge, experience, computer anxiety), organisational factors (procedures, communication, value of information, management support), social-environmental factors (public expectations, regulations, peers), technological factors (technical awareness and security tools)
- Outcome: beliefs (instrumental, behavioral, normative and control beliefs) , attitude, behavioral intention, actual behavior

While this integrative framework helps to categorize current research projects and to visualize the various factors influencing an individual's cognitive state of mind, an ingenuous reader might be misled by the deterministic input-output model promoted. For being able to successfully implement effective behavior change it is important to understand and model the dynamics behind organizational behaviour.

5.2 Approaches to assess the cyber security culture

To study and measure human behaviour, it is therefore recommended to rely on behavioral sciences and organizational theory. This is usually summarized as the assessment of organizational security culture, see e.g. (von Solms, 2000; Martins & Eloff, 2002; Schlienger & Teufel, 2002).

The benefits of assessing the cyber security culture are:

- Baseline for subsequent assessments to demonstrate effectiveness and track changes over time
- Setting priorities in investments based on sound facts
- Optimization of current activities and planning of new improvement actions
- Understand the organizational (sub-)cultures and drivers that influence the behavior

Organisational theory suggests various instruments to gather data about the organisational cultures (see Table 2).

Approach	Instrument	Pros	Cons
Quantitative (Outsider perspective, deductive)	Standardized surveys	Quantitative statistical analysis Cost effective large samples possible (representativeness) Researcher is emotionally not involved	May not respect the specific cultural setting A priori assumptions might be wrong Answers might reflect only the desired state and not the current state
	Structured interviews		
	Scenario-based surveys		
Qualitative (Insider perspective, inductive)	Workshops	Ability to assess all facets Flexible	Analysis is complex and time consuming Expensive (only small samples possible) Might be biased due to Interviewee and/or Interviewer Difficulties with interpretation
	Unstructured interviews		
	Projective techniques		

Table 2 Quantitative vs. qualitative data gathering instruments to assess the organisational culture (see Sackmann, 2017)

5.3 Assembling the puzzle: Triangulation

For a general discussions of these methods in the scope of organisational behavior, see e.g. Sackmann (2017), Schein (1985; 2016), Vecchio (2004), Despres (1995).

Using triangulation, the combination of different instruments, as suggested by most of these researchers, equalizes the pros and cons of the selected instruments. This allows to verify the results with other instruments and to use different viewpoints in interpreting them. Ideally, one should combine quantitative and qualitative measures.

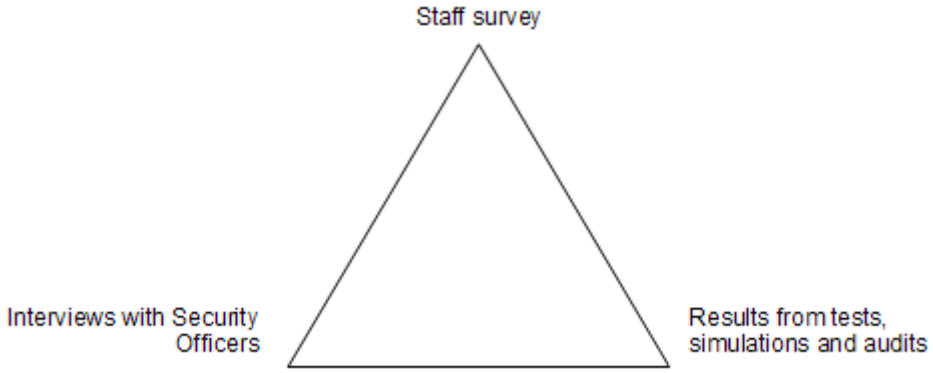


Figure 1 Example for triangulation

5.3.1 Quantitative survey

Schlienger (2006) has developed a model of organizational behaviour to study the security culture within organisations. The model has been applied and validated in several studies. It is used to measure the cyber security culture with a standardized questionnaire on three levels and twelve domains. Since the questionnaire is based on a model of organizational behavior, the measurement results indicate concrete starting points for improving and changing the cyber security culture.

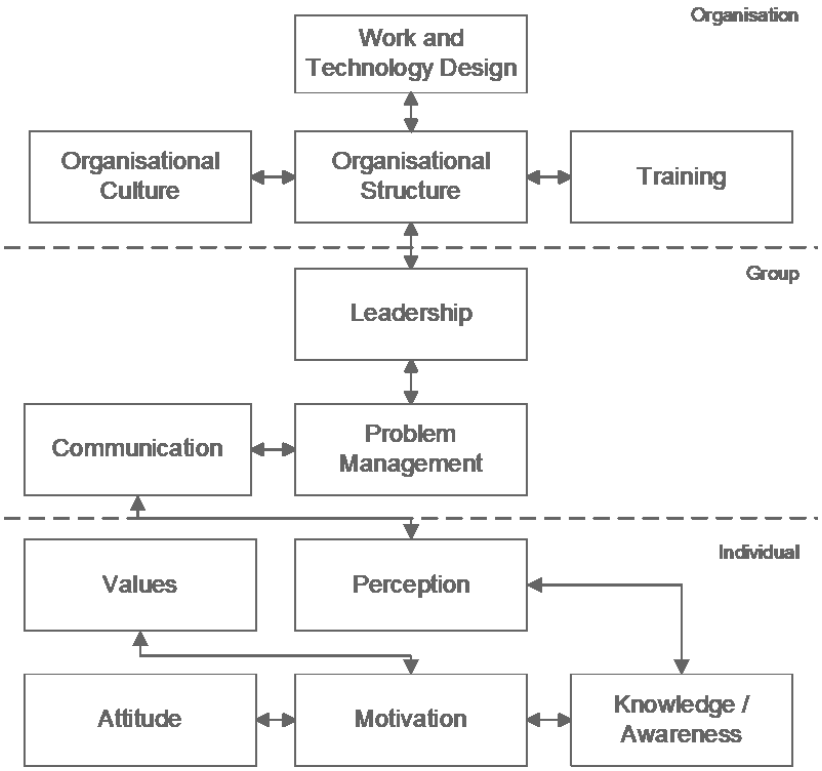


Figure 2 Organisational behaviour model to assess cyber security culture (arrows are only exemplary to demonstrate interconnections between domains)

Because of the measurement and interpretation problems of surveys described in the above section, Schlienger recommends to verify the results of the measurement with other data from interviews or audits. He suggests also to use a free text question, where the survey participants can give further comments.

The model is successfully used for 15 years now and has proven, that a sound model for assessing the cyber security culture gives valuable insights into the drivers that influence the behavior and successfully supports organisations in changing the behaviour and improving the cyber security.

The experience with this model has shown, that the main pain points most often are not in awareness and training, but in supporting domains that strongly influence the work environment of the users: work and technology design (usable security tools, understandable and meaningful policies and guidelines), organisational structure (clear and realistic work processes, well known points of contacts), leadership (security support by the middle and upper management as well as the executive board) and problem management (will reported incidents or suspected incidents be well received and the situation improved?). These domains strongly influence the perception and the motivation, which in fact finally have a strong impact on the behaviour.

6. References

6.1 References (Evidence Review 1 - Constructs)

Adams, A. & Sasse, M. A. (1999): Users are not the enemy: Why users compromise computer security mechanisms and how to take remedial measures. *Communications of the ACM*, 42(12):40-46, December 1999.

Ajzen, I. (1991): The theory of planned behaviour. Organizational Behavior and Human Decision Processes. Volume 50, Issue 2, December 1991, Pages 179-211.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215.

Becker, I. & Sasse, M. A. (2018): Separating Security Science from Pseudo-Science:

A Systematisation of Knowledge (SoK) review of survey constructs measuring security behaviour. Report and analysis available at https://verdi.cs.ucl.ac.uk/constructDB/

J.Bonneau, C.Herley, P.C.vanOorschot, and F.Stajano. Passwords and the evolution of imperfect authentication. Communications of the ACM 58(7):78–87, June2015.

Herley, C. (2009): So long, and no thanks for the externalities: the rational rejection of security advice by users. Proceedings of the New Security Paradigms Workshop (NSPW) 2009. ACM.

Safa, N. S., & Von Solms, R. (2016). An information security knowledge sharing model in organizations. *Computers in Human Behavior, 57*, 442-451.

6.2 References (review of interventions)

Abraham, C., & Michie, S. (2008). A taxonomy of behavior change techniques used in interventions. *Health psychology*, *27*(3), 379.

Anwar, M., He, W., Ash, I., Yuan, X., Li, L., & Xu, L. (2017). Gender difference and employees' cybersecurity behaviors. *Computers in Human Behavior, 69*, 437-443. doi:10.1016/j.chb.2016.12.040

Bauer, S., & Bernroider, E. W. N. (2017). From Information Security Awareness to Reasoned Compliant Action: Analyzing Information Security Policy Compliance in a Large Banking Organization. *SIGMIS Database, 48*(3), 44-68. doi:10.1145/3130515.3130519

Ben-Asher, N., & Meyer, J. (2018). The Triad of Risk-Related Behaviors (TriRB): A Three-Dimensional Model of Cyber Risk Taking. *Hum Factors*, 18720818783953. doi:10.1177/0018720818783953

Boss, S. R., Galletta, D. F., Lowry, P. B., Moody, G. D., & Polak, P. (2015). What Do Systems Users Have to Fear? Using Fear Appeals to Engender Threats and Fear That Motivate Protective Security Behaviors. *Mis Quarterly, 39*(4), 837-U461. doi:Doi 10.25300/Misq/2015/39.4.5

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.

Bulgurcu, B., Cavusoglu, H., Benbasat, I., & Ieee. (2010). Quality and Fairness of an Information Security Policy as Antecedents of Employees' Security Engagement in the Workplace: An Empirical Investigation. In *43rd Hawaii International Conference on Systems Sciences Vols 1-5*(pp. 4098-4104).

Canfield, C. I., Fischhoff, B., & Davis, A. (2016). Quantifying Phishing Susceptibility for Detection and Behavior Decisions. *Hum Factors, 58*(8), 1158-1172. doi:10.1177/0018720816665025

Caputo, D. D., Pfleeger, S. L., Freeman, J. D., & Johnson, M. E. (2014). Going Spear Phishing: Exploring Embedded Training and Awareness. *IEEE Security & Privacy, 12*(1), 28-38. doi:10.1109/MSP.2013.106

Chou, H.-L., & Chou, C. (2016). An analysis of multiple factors relating to teachers' problematic information security behavior. *Computers in Human Behavior, 65*, 334-345. doi:10.1016/j.chb.2016.08.034

Dang-Pham, D., Pittayachawan, S., & Bruno, V. (2017a). Applying network analysis to investigate interpersonal influence of information security behaviours in the workplace. *Information & Management, 54*(5), 625-637. doi:10.1016/j.im.2016.12.003

Dang-Pham, D., Pittayachawan, S., & Bruno, V. (2017b). Why employees share information security advice? Exploring the contributing factors and structural patterns of security advice sharing in the workplace. *Computers in Human Behavior, 67*, 196-206. doi:10.1016/j.chb.2016.10.025

Djajadikerta, H. G., Roni, S. M., & Trireksani, T. (2015). Dysfunctional information system behaviors are not all created the same: Challenges to the generalizability of security-based research. *Information & Management, 52*(8), 1012-1024. doi:https://doi.org/10.1016/j.im.2015.07.008

Egelman, S., Harbach, M., & Peer, E. (2016). *Behavior Ever Follows Intention?: A Validation of the Security Behavior Intentions Scale (SeBIS)*. Paper presented at the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, California, USA.

Egelman, S., & Peer, E. (2015). Predicting privacy and security attitudes. *SIGCAS Comput. Soc., 45*(1), 22-28. doi:10.1145/2738210.2738215

Hadlington, L., & Murphy, K. (2018). Is Media Multitasking Good for Cybersecurity? Exploring the Relationship Between Media Multitasking and Everyday Cognitive Failures on Self-Reported Risky Cybersecurity Behaviors. *Cyberpsychol Behav Soc Netw, 21*(3), 168-172. doi:10.1089/cyber.2017.0524

Hadlington, L., & Parsons, K. (2017). Can Cyberloafing and Internet Addiction Affect Organizational Information Security? *Cyberpsychology, behavior and social networking, 20*(9), 567-571. doi:10.1089/cyber.2017.0239

Halevi, T., Lewis, J., & Memon, N. (2013). *A pilot study of cyber security and privacy related behavior and personality traits*. Paper presented at the Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil.

Halevi, T., Memon, N., Lewis, J., Kumaraguru, P., Arora, S., Dagar, N., . . . Chen, J. (2016). *Cultural and psychological factors in cyber-security*. Paper presented at the Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services, Singapore, Singapore.

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. doi:doi:10.1111/j.1365-2575.2012.00420.x

Herath, T., & Rao, H. R. (2009a). Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. *Decision Support Systems, 47*(2), 154-165. doi: https://doi.org/10.1016/j.dss.2009.02.005

Herath, T., & Rao, H. R. (2009b). Protection motivation and deterrence: a framework for security policy compliance in organisations. *European Journal of Information Systems, 18*(2), 106-125. doi: 10.1057/ejis.2009.6

Hu, Q., Xu, Z., Dinev, T., & Ling, H. (2011). Does deterrence work in reducing information security policy abuse by employees? *Commun. ACM, 54*(6), 54-60. doi:10.1145/1953122.1953142

Johnston, A. C., & Warkentin, M. (2010). Fear appeals and information security behaviors: an empirical study. *MIS Q., 34*(3), 549-566.

Kajtazi, M., Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2014, Jan 06-09). *Assessing Sunk Cost Effect on Employees' Intentions to Violate Information Security Policies in*

*Organizations.*Paper presented at the 47th Annual Hawaii International Conference on System Sciences, Waikoloa, HI.

Kelley, T., Amon, M. J., & Bertenthal, B. I. (2018). Statistical Models for Predicting Threat Detection From Human Behavior. *Front Psychol, 9*, 466. doi:10.3389/fpsyg.2018.00466

Kim, S. H., Yang, K. H., & Park, S. (2014). An integrative behavioral model of information security policy compliance. *Scientific World Journal, 2014*. doi:10.1155/2014/463870

Mamonov, S., & Benbunan-Fich, R. (2018). The impact of information security threat awareness on privacy-protective behaviors. *Computers in Human Behavior, 83*, 32-44. doi:10.1016/j.chb.2018.01.028

Martin, J., Dube, C., & Coovert, M. D. (2018). Signal Detection Theory (SDT) Is Effective for Modeling User Behavior Toward Phishing and Spear-Phishing Attacks. *Hum Factors*, 18720818789818. doi:10.1177/0018720818789818

Mayer, P., Kunz, A., & Volkamer, M. (2017). *Reliable Behavioural Factors in the Information Security Context*. Paper presented at the Proceedings of the 12th International Conference on Availability, Reliability and Security, Reggio Calabria, Italy. http://delivery.acm.org/10.1145/3100000/3098986/a9-

Michie, S., Van Stralen, M. M., & West, R. (2011). The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation science*, *6*(1), 42.

Oliveira, D., Rocha, H., Yang, H., Ellis, D., Dommaraju, S., Muradoglu, M., . . . 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 Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.

Ovelgönne, M., Dumitra, T., Prakash, B. A., . . . Wang, B. (2017). Understanding the Relationship between Human Behavior and Susceptibility to Cyber Attacks: A Data-Driven Approach. *ACM Trans. Intell. Syst. Technol., 8*(4), 1-25. doi:10.1145/2890509

Pahnila, S., Siponen, M., & Mahmood, A. (2007, 3-6 Jan. 2007). *Employees' Behavior towards IS Security Policy Compliance.*Paper presented at the 2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07).

Safa, N. S., & Von Solms, R. (2016). An information security knowledge sharing model in organizations. *Computers in Human Behavior, 57*, 442-451. doi:10.1016/j.chb.2015.12.037

Shahri, A. B., Ismail, Z., & Mohanna, S. (2016). The Impact of the Security Competency on "Self-Efficacy in Information Security" for Effective Health Information Security in Iran. *Journal of Medical Systems, 40*(11). doi:10.1007/s10916-016-0591-5

Shoshitaishvili, Y., Invernizzi, L., Doupe, A., & Vigna, G. (2014). *Do you feel lucky?: a large-scale analysis of risk-rewards trade-offs in cyber security*. Paper presented at the Proceedings of the 29th Annual ACM Symposium on Applied Computing, Gyeongju, Republic of Korea.

Siponen, M., Adam Mahmood, M., & Pahnila, S. (2014). Employees' adherence to information security policies: An exploratory field study. *Information & Management, 51*(2), 217-224. doi:https://doi.org/10.1016/j.im.2013.08.006

Siponen, M., Pahnila, S., & Mahmood, A. (2007). *Employees' Adherence to Information Security Policies: An Empirical Study*, Boston, MA.

Siponen, M., Pahnila, S., & Mahmood, M. A. (2010). Compliance with Information Security Policies: An Empirical Investigation. *Computer (Long Beach Calif), 43*(2), 64-71. doi:10.1109/mc.2010.35

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. doi:10.1016/j.chb.2017.05.038

Wang, J., Xiao, N., & Rao, H. R. (2010). Drivers of information security search behavior: An investigation of network attacks and vulnerability disclosures. *ACM Trans. Manage. Inf. Syst., 1*(1), 1-23. doi:10.1145/1877725.1877728

Whitty, M., Doodson, J., Creese, S., & Hodges, D. (2015). Individual differences in cyber security behaviors: an examination of who is sharing passwords. *Cyberpsychology, behavior and social networking, 18*(1), 3-7. doi:10.1089/cyber.2014.0179

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.

Williams, E. J., Morgan, P. L., & Joinson, A. N. (2017). Press accept to update now: Individual differences in susceptibility to malevolent interruptions. *Decision Support Systems, 96*, 119-129.

Yazdanmehr, A., & Wang, J. (2016). Employees' information security policy compliance: A norm activation perspective. *Decision Support Systems, 92*, 36-46. doi:10.1016/j.dss.2016.09.009

Zhang-Kennedy, L., Chiasson, S., & Biddle, R. (2016). The role of instructional design in persuasion: A comics approach for improving cybersecurity. *International Journal of Human-Computer Interaction, 32*(3), 215-257. doi:10.1080/10447318.2016.1136177

6.3 References (Review of Qualitative Studies)

Acar, Y., Backes, M. Fahl, S., Kim, D., Mazurek, M., Stransky, S. (2016): You Get Where You're Looking For - The Impact of Information Sources on Code Security. Proceedings of the 2016 IEEE Symposium on Security and Privacy - Oakland'16.

Ashenden, D. : Information Security Awareness: Improving Current Research & Practice. PhD thesis, UCL Department of Computer Science. 2015

Ashenden, D., Lawrence, D. (2016): Security Dialogues: Building Better Relationships between Security and Business. IEEE Security & Privacy Magazine, May/June 2016.

Ashenden, D. & Sasse, M. A. (2013): CISOs and organisational culture: Their own worst enemy? Computers & Security, Volume 39, Part B, November 2013, Pages 396-405.

Bada, M., Sasse, M. 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, 118–131.

Beautement, A. Sasse, M. A., Wonham, M. (2009): The compliance budget: managing security behaviour in organisations. In New Security Paradigms Workshop (NSPW), 2008, pages 47-58.

Beautement, A., Becker, I., Parkin, S., Krol, K., & Sasse, M. A. (2016). Productive Security: A Scalable Methodology for Analysing Employee Security Behaviours. Procs. SPOUPS 2106 USENIX Association.

Becker, I., Sasse M. A. (2018): Separating Security Science from Pseudo-Science:

A Systematisation of Knowledge (SoK) review of survey constructs measuring security behaviour. Report and analysis available at https://verdi.cs.ucl.ac.uk/constructDB/

Beris, O., Beautement, A., & Sasse, M. A. (2015). Employee Rule Breakers, Excuse Makers and Security Champions:: Mapping the Risk Perceptions and Emotions That Drive Security Behaviors. *Proceedings of the 2015 New Security Paradigms Workshop*, 73-84.

Caputo, D.D., Pfleeger, S.L., Sasse, M.A., Ammann, P., Offutt, J., Deng, L. (2016). Barriers to Usable Security? Three Organizational Case Studies. IEEE Security and Privacy, 14 (5), 22-32. doi:10.1109/MSP.2016.95

Chen, T. R., Shore, D. B. Zaccaro, S. J., Dalal, R. S. Tetrick, L. E. & Gorab, A. K. (2016): An organizational psychology perspective to examining computer security incident response.teams. IEEE Security & Privacy, (5):61–67, 2014.

Coles-Kemp, L., Ashenden, D., O'Hara, K (2018): Why Should I? Cybersecurity, the Security of the State and the Insecurity of the Citizen. Politics and Governance (ISSN: 2183–2463) 2018, Volume 6, Issue 2, Pages 41–48.

Dunphy, P., Vines, J., Coles-Kemp, L. Clarke, R., Vlachokyriakos,V. Wright, P., McCarthy, J. & Olivier, P. “Understanding the Experience- Centeredness of Privacy and Security Technologies,” in Proceedings of the 2014 Workshop on New Security Paradigms Workshop NSPW 2014, pp. 83–94.

Fahl, A., Harbach, M. Perl, H. Kötter, M., Smith, M. (2013): Rethinking SSL Development in an Appified World. Proceedings of the 2013 ACM conference on Computer and Communications Security - CCS'13.

Heath, C., Hall, P. & Coles-Kemp, L. (2018): Holding on to dissensus: Participatory interactions in security design. Strategic Design Research Journal, 11(2): 65-78 May-August 2018.

Kirlappos, I., Parkin, S., Sasse, M.A. (2015). "Shadow security" as a tool for the learning organization. ACM SIGCAS Computers and Society, 45 (1), 29-37. doi:10.1145/2738210.2738216

Pallas, F. (2009): Information Security Inside Organizations - A Positive Model and Some Normative Arguments Based on New Institutional Economics (August 11, 2009). Available at SSRN: https://ssrn.com/abstract=1471801 or http://dx.doi.org/10.2139/ssrn.1471801

Pfleeger, 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. doi:10.1515/jhsem-2014-0035.

Poller, Andreas, Laura Kocksch, Sven Türpe, Felix Anand Epp, and Katharina Kinder-Kurlanda. "Can Security Become a Routine? A Study of Organizational Change in an Agile Software Development Group." In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, pp. 2489-2503. ACM, 2017.

Sasse, M.A., Smith, M., Herley, C., Lipford, H., Vaniea, K. (2016). Debunking Security-Usability Tradeoff Myths. IEEE SECURITY & PRIVACY, 14 (5), 33-39.

Safa N. S., Von Solms, R. & Furnell, S. Information security policy compliance model in organizations. Computers & Security, Volume 56, February 2016, Pages 70-82.

SC Sundaramurthy, J Case, T Truong, L Zomlot, M Hoffmann (2014):

A tale of three security operation centers. Proceedings of the 2014 ACM workshop on security information workers, 43-50.

Sathya Chandran Sundaramurthy, Alexandru G Bardas, Jacob Case, Xinming Ou, Michael Wesch, John McHugh, S Raj Rajagopalan (2015) A Human Capital Model for Mitigating Security Analyst Burnout. Procs SOUPS 2015 pp. 347-359.

Sathya Chandran Sundaramurthy, Michael Wesch, Xinming Ou, John McHugh, S Raj Rajagopalan, Alexandru Bardas. (2017): Humans are dynamic. Our tools should be too. Innovations from the Anthropological Study of Security Operations Centers. IEEE Internet Computing.

Weirich, D. (2006). Persuasive Password Security. PhD Thesis, University College London 2006. http://discovery.ucl.ac.uk/1446157/1/Weirich_thesis.pdf

Whitten, A. & Tygar, D. (1999): Why johnny can't encrypt: A usability evaluation of pgp 5.0. In *Proceedings of the 8th USENIX Security Symposium*, August 1999.

6.4 References (current practise)

Despres, C. J.-N., 1995. Culture, Surveys, Culture Surveys and Other Obfuscations: A Reply to Migliore & Martin. The Journal of Strategic Change, Band 4, pp. 65-75.

Jaeger, L., 2018. Information Security Awareness: Literature Review and Integrative Framework. Hawaii, USA, s.n.

Martins, A. & Eloff, J. H. P., 2002. Information Security Culture. In: Security in the information society: visions and perspectives. Cairo, Egypt: IFIP TC11 International Conference on Information Security.

National Institute of Standards and Technology, 2008. Performance Measurement Guide for Information Security, Gaithersburg, USA: NIST Special Publication 800-55 Revision 1.

Sackmann, S., 2017. Unternehmenskultur: Erkennen - Entwickeln - Verändern. München, Germany: Springer Gabler.

Schein, E. H., 1985. Organizational Culture and Leadership: A Dynamic View. San Francisco, USA: Jossey-Bass.

Schein, E. H., 2016. The corporate culture survival guide: sense and nonsense about culture change. Hoboken, USA: Wiley.

Schlienger, T., 2006. Informationssicherheitskultur in Theorie und Praxis. Doctoral Thesis. Fribourg, Switzerland: iimt University Press.

Schlienger, T. & Teufel, S., 2002. Information Security Culture - The Socio-Cultural Dimension in Information Security Management. In: Security in the information society: visions and perspectives. Cairo, Egypt: IFIP TC11 Internation-al Conference on Information Security.

Vecchio, R. P., 2004. Organizational behavior: core concepts. 6 ed. Fort Worth, USA: South-Western College Pub.

von Solms, B., 2000. Information Security - The Third Wave. Computers & Security, 19(7), pp. 615-620.

7. Appendix

7.1 APPENDIX A:

Commercial options for measurement of security awareness or security culture

Tests

- Countless Phishing Test providers [PhishMe or Wombat - why this does not improve behaviour/culture - drawbacks: cowed employees who don't click on anything, lost business]
- Social Engineering offered by many IT Security companies [more research needed - we should provide at least one example - Angela]

Culture Assessments

- AskIt by KnownSense
- CTRLRe by Roer Group
- SABR by The Security Company
- SAM by Steinbeis (Prof. Zerr)
- TWISK by TreeSolution

ENISA

European Union Agency for Network
and Information Security
1 Vasilissis Sofias
Marousi 151 24, Attiki, Greece

Heraklion Office

Science and Technology Park of Crete (ITE)
Vassilika Vouton, 700 13, Heraklion, Greece

ISBN 978-92-9204-326-1 DOI 10.2824/717888

TP-02-20-012-EN-N

1 Vasilissis Sofias Str, Maroussi 151 24, Attiki, Greece Tel:
+30 28 14 40 9710
info@enisa.europa.eu
www.enisa.europa.eu

ISBN: 978-92-9204-326-1
DOI: 10.2824/717888