# An Insider Threat Neutralisation Mitigation model predicated on Cognitive Dissonance ($ITNM^{CD}$)

Keshnee Padayachee

Institute of Science and Technology Education, University of South Africa

## ABSTRACT

The insider threat concern is a complex issue, as the problem domain intersects the social, technical and socio-technical dimensions. Consequently, counteracting the insider threat involves influencing the insider's perceptions and behaviour in order to ensure compliance. When an individual's actions and beliefs are incongruent, this induces a phenomenon known as cognitive dissonance. In order to reduce this dissonance, individuals are self-motivated either to change their behaviours or beliefs, or to rationalize their behaviour. Neutralisation is a technique used by criminals to rationalize maleficence. In terms of the insider threat, it has been proposed that if the rationalisations for committing an offence are eliminated, then the insider is less likely to commit the offence. This process is known as neutralisation mitigation. The research reported on here proposes inducing cognitive dissonance in order to counter the resultant neutralisations that may ensue with neutralisation mitigation. To test these concepts, a pragmatic implementable solution, the Insider Threat Neutralisation Mitigation model predicated on Cognitive Dissonance ($ITNM^{CD}$), is proposed. A proof-of-concept was developed and the model concept was evaluated using the design science method.

KEYWORDS: Cognitive dissonance, neutralisation theory, situational crime prevention theory, neutralisation techniques

CATEGORIES: K.6.5 [**Security and protection**]: Unauthorised access (e.g. hacking, phreaking)

ARTICLE HISTORY

Received 8 July 2014
Accepted 9 May 2015

## 1 INTRODUCTION

An insider threat is more hazardous than an external threat, as an insider may use skills and knowledge gained through legitimate work duties for illegitimate gain [1]. According to the 'US State of Cybercrime Survey', 32% of the respondents acknowledged that 'insider crimes are more costly or damaging than incidents perpetrated by outsiders' [2]. Farahmand and Spafford [3] found that most law enforcement agents use the fraud triangle to investigate the insider threat. The fraud triangle, which is used as a framework to explain crime, is composed of three elements—pressure (i.e. motivation), opportunity and rationalisation [4]. A comprehensive insider threat strategy will involve addressing all three elements of the fraud triangle.

However, this paper focuses primarily on the rationalisation element of the fraud triangle; it is clear that if an insider is unable to excuse and justify an offence, then the offence is not considered to be a suitable opportunity [5]. Related to the notion of rationalisations is the concept of neutralisation techniques. These techniques are used to rationalize behaviour, 'whether in response to cognitive dissonance, as precondition to acting, or other factors' [6]. When an individual's actions and beliefs are incongruent, this induces a phenomenon known as cognitive dissonance [7]. In order to reduce this dissonance, individuals are self-motivated either to change their behaviours or beliefs, or to rationalize their behaviour.

The research reported on here proposes that inducing cognitive dissonance to counter the resultant neutralisations (i.e. rationalisations) that may ensue with neutralisation mitigation, which is a process employed to remove subsequent rationalisations [8], may reduce the insider threat in a complementary manner. This would involve simulating an environment where an insider, in the process of committing maleficence, is simultaneously challenged by neutralisation mitigation. In order to simulate such an environment, honeypots are deployed to bait (i.e. entice) the insider into maleficence. It is envisaged that the insider will no longer be subject to those preconceived rationalisations when a real opportunity to commit maleficence arises. The

**Email:** Keshnee Padayachee padayk@unisa.ac.za

aim of the research was to propose and evaluate a model that incorporates these concepts into an implementable solution to test this proposition—the Insider Threat Neutralisation Mitigation model predicated on Cognitive Dissonance ($\text{ITNM}^{\text{CD}}$).

The theoretical underpinning for this model is the convergence of three theories concerned with rationalisations, namely situational crime prevention theory [9], neutralisation theory [10] and cognitive dissonance theory [7]. The first two theories have their basis in criminology, while the third is a seminal theory in social psychology. These three theories provide responses to three questions:

1. How does an insider rationalize cybercrime?
2. What intrinsic processes influence an insider to rationalize cybercrime?
3. How can organisations deal with these rationalisations of cybercrime, in other words, reduce the excuses made by a maleficent insider?

The response to the first question is to be obtained from neutralisation theory, which provides a set of techniques that individuals use to justify crime, such as denial of injury; these are also applicable to the insider threat problem [8, 11]. The answer to the third question is supplied by situational crime prevention theory, which provides a set of techniques to remove excuses. The second question is the most difficult one to answer, as these processes are intrinsic to the criminal. However, cognitive dissonance theory may offer an answer to this question in terms of behaviour and attitude.

A cognition is any piece of knowledge—that is, knowledge about one's behaviour or attitude, or about the state of the world: 'If a person holds cognitions A and B such that A follows from the opposite of B, then A and B are dissonant' [12]. Cognitive dissonance directs the tension experienced when a person holds two or more conflicting cognitions simultaneously: these could be ideas, beliefs, values or emotional reactions [7]. Redondo and Charron [13] used cognitive dissonance to understand the payment patterns of downloaders of music and movies to explain how users diminish the importance of compensating authors to reduce the dissonance arising from downloading without paying. The study reported on in this article reasoned similarly about the advantages of using cognitive dissonance to underpin the proposed model. First, cognitive dissonance is a well-researched phenomenon in social psychology [12]. Second, the theory offers a flexible framework to explain behaviour. Cognitive dissonance usually compels an individual to rationalize their actions or attitudes rather than propelling a fundamental change [12].

Sykes and Matza [10] hypothesize that criminals use neutralisation techniques (or rationalisations) before committing an illegal act. For instance, insiders may rationalize their actions by labelling cybercrime as a victimless crime; however, it has been posited that if the rationalisations are removed, then the insider threat can be circumvented through a process termed 'neutralisation mitigation' [8]. Situational crime prevention theory offers a set of specific techniques to remove the rationalisations that criminals may use to commit crime. The model described here leverages these techniques. In the past, situational crime prevention theory has been applied in order to solve the problem of the insider threat [14]. If offenders can be prevented from rationalizing and excusing their criminal actions, they will be open to feelings of guilt and shame [15], and will consequently refrain from committing the crime.

The aim of the research was to propose a model for neutralisation mitigation that is activated by inducing cognitive dissonance. The model relies on dissonance-generated self-persuasion rather than the direct persuasion offered by training. The model was empirically evaluated using the design science methodology. This methodology was selected because it provides more than a usability study of the prototype, and allows participants to reason about the implementation and operational requirements of the system as well as possible deficiencies of the model.

This article is structured as follows: Section 2 presents related work dealing with the problem of the insider threat, while Section 3 presents the theoretical framework for the model. Section 4 presents the model itself, while the proof-of-concept is elaborated on in Section 5. Section 6 contains a discussion of the concepts underlying the research methodology, while the results of the study are presented in Section 7. Section 8 provides a discussion of the findings, and Section 9 outlines the implications for practice. The article concludes with Section 10 and possible future research opportunities.

## 2 RELATED WORK

An essential difference between outsiders and insiders is that outsiders have limited opportunities to carry out their attacks. The latter have to exploit vulnerabilities in the system, while insiders have privileged access and hence greater opportunity [16]. Insiders have a significant advantage, as not only do they have knowledge about vulnerabilities in policies, networks or systems [17], but also the requisite capability. Walton [16] observes, on a positive note, that the insider (unlike an outsider) is subjected to policies, procedures and agreements. For instance, if an insider agrees to be subject to monitoring by a honeypot, then unlike an outsider he/she may have no legal recourse. An insider has 'legitimate access to an organisation's computers and networks', while an insider threat is an entity that places organisations' assets at risk [18]. Bishop and Gates [19] further qualify the modus operandi of a malicious insider as constituting a violation of information security policy using either authorised access or unauthorised access. There are two classes of insider threat: 'traitors' and 'masqueraders'. Traitors are legitimate users that abuse their privileges, and masqueraders are insiders that impersonate other users for malicious purposes [20].

Misuse can be classified as intentional or accidental.

Examples of attacks include unauthorised extraction, duplication or exfiltration of data, tampering with data, deletion of critical assets, etc. [21]. The motivations for intentional misuse range from revenge, disgruntlement, avarice and divided loyalty to delusion; accidental misuse could be due to inadequate system knowledge, stress, or a lack of rules [22]. Interestingly, it was found empirically that 'private or sensitive information unintentionally exposed' accounts for 82% of security incidents committed by insiders, while 'confidential records compromised or stolen' accounts for 76% of security incidents committed by insiders [2]. Hence unintentional incidents by benign insiders are just as significant as intentional incidents by malicious insiders. It is important to note that there is grey area between intentional and accidental misuse [23]. This can imply that an insider who committed a crime intentionally may claim to have committed a cybercrime by mistake, and it is difficult to tell the difference.

Wood [24] describes the attributes of an insider threat as:

- access,
- knowledge,
- privileges,
- skill,
- risk,
- tactics,
- motivation, and
- process.

The process ranges from the intrinsic motivation of the insider, to identifying a target, to planning and finally launching the attack [24]. The knowledge and skill factors are related to capability. The capability of an insider to commit a crime is a significant factor, as an insider who may be motivated to commit a crime must also have the capability to commit the crime.

In designing an insider threat detection or prevention program, it may be practical to map the capability of insiders. For example, Bowen et al. [25] mapped the following capability characteristics to design honeypots to account for each competency level:

- Low: This type of insider relies on what may be discerned from a cursory scan such as shoulder surfing.
- Medium: This type of insider performs verifications from other sources to check the authenticity of the information, for example by using a search engine.
- High: This type of adversary has access to the most sophisticated tools, for example key stroke loggers.
- Highly privileged: This type of adversary may know that there are decoys or detection systems and will attempt to disable or avoid them.

Sophisticated users are more dangerous as they are more likely to cause detrimental attacks and able to cover their tracks more effectively [26]. It may be prudent to monitor those insiders with higher privileges more closely. A methodical review of technological tools available to minimize the insider threat may be found in Zeadally et al. [27]. Mechanisms to undermine the insider threat include monitoring, detection, mitigation [25] and deterrence and profiling.

Data Loss Prevention (DLP) tools may be used to monitor data usage so as to detect and mitigate insider threats [28]. DLP is about preventing leakage of sensitive information and it involves managing, classifying, monitoring and protecting data according to its state. These states include data at rest (e.g. data residing in a database), data at the end point (e.g. data residing in mobile devices) or data in motion (i.e. data moving through to the outside world via communication mechanisms, such as e-mail) [29]. According to Guido and Brooks [30], applying a DLP policy to control network services such as e-mail may deter potential insiders. DLP can, for example, help to identify an insider who downloads classified documents and attaches them to an e-mail [31]. Aside from system performance issues, a DLP may not be able to intercept a message that is encrypted [31] or detect an insider who uses steganography to obfuscate an e-mail message [32].

Although intrusion detection systems are deployed to manage the insider threat, these systems have typically been designed for the external rather than the insider threat. Bowen et al. [25] argue that intrusion detection mechanisms present a number of challenges, ranging from false positives to difficulty in correctly identifying anomalous behaviour. Zeadally et al. [27] remark that intrusion detection systems may be ineffective if an insider leaves no traces behind because they have knowledge of how to disable the intrusion detection system [33]. Intrusion detection systems are susceptible because they cannot discern a pattern of crime committed sporadically over a long period of time, and this is further complicated by the fact that insiders will perform malicious acts in the course of normal activities [23].

Unlike intrusion detection systems, honeypots are unlikely to be beset by false positives, as any interaction with a honeypot is likely to be illicit [34]. A number of studies have been conducted on using honeypots to detect the insider threat [25, 35]. For example, McGrew et al. [35] found that honeypots succeed in 'sandboxing' (i.e. containing) activities related to an insider. However, according to Spitzner [34], honeypots have several disadvantages. There is a risk that an attacker may use a honeypot to harm other systems. Honeypots are only of value when an attacker interacts with them and they manage to capture actions related to this activity. Several legal [36] and ethical concerns [37] are also associated with deploying honeypots. Honeypots provide an opportunity for an insider to commit maleficence in a controlled manner:

> Moreover, honeypots may also act as a warning device for more serious maleficence as malicious activity in a honeypot may also point to malicious activity elsewhere in the system. [38]

Several researchers have advocated profiling to predict future threats. An accurate profile of the insider may help to identify threats both prospectively and retro-

spectively [39]. For instance, Schultz [33] recommends the following indicators to detect the insider threat:

- personality traits
- deliberate markers
- preparatory behaviour
- correlated usage patterns
- verbal behaviour
- meaningful errors

Some of these indicators may not be the most tangible sources of information; however, the elements relating to preparation may be. An insider has to prepare for an attack by searching for information and in doing so may make mistakes. A pattern to this preparation may be detected, such as issuing the same commands across all systems. However, as Hunker and Probst [40] claim, profiling has its drawbacks as it assumes that human behaviour is predictable.

There will always be a prerequisite for 'technical controls such as encryption, access control, minimum privilege, monitoring, auditing and reporting' [41]. However, in a survey commissioned by Raytheon, it was found that following challenges with respect to security tools, detecting whether an event is an insider threat was ranked highest, followed by insufficient contextual information (69%), 'too many false positives' (56%) and copious data (45%) [42]. The fourth-ranked challenge indicates that the 'behaviour involved in the incident is consistent with the individual's role and responsibilities (28%)'; this is also significant as it is another reason why it is so difficult for security tools to discern the difference between normal activities and misuse.

Deterrence countermeasures are based on four factors: awareness of security policies, monitoring, preventive software, and training [43]. Monitoring alone is not sufficient to manage the insider threat, as it captures the intent but not the motivation. It is also difficult to identify patterns of misuse [40]. Deterrent mechanisms cannot provide insight into the actual act of a malicious insider [15]. In addition, deterrent mechanisms involving penalties do not work, as an insider will more likely focus on 'Will I be caught?' than on 'What is the punishment if I am caught?' [3].

An information security awareness and training program is an absolute necessity in any information security management plan. However it does have drawbacks: Schultz [44] indicates that it is often difficult to measure the benefits and sometimes it leads to a

> 'one size fits all' approach that leaves many attendees puzzled and many others bored, disappointed, and even hostile because they have learned nothing new.

This perspective is also shared by Kazjer et al. [45], who found that applying a 'one size fits all' approach to information security awareness training is imprudent. Kajzer et al. [45] demonstrate empirically that security awareness campaigns may backfire, as the personality traits of an individual may affect their receptiveness to the typical messages of deterrence, morality, regret, feedback and incentive that are often contained in an information security awareness program. For example, people with a high desire for social approval are not strongly influenced by morality-type messages, whereas neuroticism is positively associated with deterrence-type messages.

Given the shortcomings associated with each technique, some researchers have considered an integrated approach. For instance, Brdiczka et al. [46] used profiling and anomaly detection to detect the insider threat, while Salem and Stolfo [20] combined profiling and honeypots to reduce false positives. It is clear that managing the insider threat requires a wide-ranging approach and that no technique used in isolation is fully satisfactory. Hence the technique presented in this paper is intended to be used as part of an integrated solution.

## 3 THEORETICAL FRAMEWORK

This article draws on three prevailing theories: cognitive dissonance theory, neutralisation theory and situational crime prevention theory. The neutralisation theory explains the rationalisations used by criminals, while cognitive dissonance focuses more on behaviour and attitude. Situational crime prevention theory forms the basis of the mitigation techniques.

### 3.1 Cognitive dissonance

Although cognitive dissonance is not a recent concept it remains relevant, as even though technology changes, fundamentally human beings do not. Although cognitive dissonance is a social science concept, it has been used in other areas such as marketing [47], education [48] and management [49]. Aside from individual dissonance, the concept of organisational dissonance also exists, where tension 'arises from the misalignment of key organisational elements', and cognitive dissonance may be used to explain and understand the actions of organisations [49]. Cognitive dissonance can be applied in information security, especially when new policies are enforced and the changes will be met by resistance; if management is consistent, they change behaviour by cognitive dissonance [50]. Cognitive dissonance has been applied in information security. Workman [51] considered the role of cognitive dissonance in social engineering attacks, while Lawrence and Caputo suggest that a system could

> emphasize a user's sense of foolishness concerning the cyber risks he is taking, enabling dissonant tension to be injected suddenly or allowed to build up over time. Then, the system can offer the user ways to relieve the tension by changing his behaviour. [52]

The model presented in this article is based on challenging the rationalisations made by an insider with neutralisation mitigation.

Reodondo and Charron's [13] empirical study on digital piracy also used cognitive dissonance as a theoretical framework to explain the phenomenon. They propose that piracy is a not a crime of ignorance, but rather that users 'underestimate' the importance

of compensating creators of the work to 'reduce the tension arising from the behavioural inconsistency of downloading without paying'. They advise that negative campaigns advising users not to pirate are ineffective as the users rationalize their actions by devaluing the artistry of the creator. They liken the negative campaigns against piracy to the negative campaigns against smoking, which are also, coincidentally, ineffective as smokers undervalue the health warnings [53]. Hence they suggest that positive advertisement highlighting the ingenuity of the artist (i.e. the skill, creativity, originality etc.) will be more successful, and this includes affirmations that praise users who pay. They also suggest that eclectic downloaders (i.e. users who pay occasionally) should be coerced into seeing the hypocrisy of paying in one instance and not in another.

This shows that cognitive dissonance is more than a theory for explaining behaviour; it may be used a practical tool to change attitudes or behaviour. In other words, one can induce dissonance for positive change: for instance, the notion of inducing hypocrisy is based on the idea of cognitive dissonance. Aronson et al. [54] show that hypocrisy (what participants 'preach' versus their actual behaviour) can be an effective technique in changing behaviour. This approach combines two techniques: commitment and mindfulness [55]. This approach was demonstrated by Aronson et al. [54] and Morrongiello and Landa [56]. The idea is to indirectly prompt an individual to reflect on the hypocrisy of their behaviour. The reason that this approach works is that individuals want to maintain self-consistency [57]. Commitment involves advocating a positive message such as signing a commitment statement [58]. Mindfulness involves 'making the individual aware of instances when he/she did behave in line with the advocated standards' [55] and reneged on their earlier commitment.

An application of this technique was demonstrated by Dickerson et al. [58], which involved instructing individuals to sign a commitment statement about water conservation. Later the same individuals were shown the hypocrisy of their behaviour (i.e. that they did not in reality conserve water). In this manner, they were made aware that their behaviour showed inconsistency with their earlier commitment. This notion is termed dissonance-generated self-persuasion. The dissonance-based persuasion is a 'powerful behaviour-altering force likely to be more effective than straightforward persuasive or coercive appeals' and more 'enduring' [58]. Persuasion by hypocrisy involves inducing an individual to commit to something they believe in and then reminding them their behaviour is not consistent with that belief [59]. It was claimed that this type of persuasion was more effective than typical direct persuasion methods like training. Dickerson et al. [58] demonstrate that dissonance-generated self-persuasion is more successful than typical information campaigns.

In Dickerson's [58] experiment it was shown that hypocrisy is more potent when combined with mindfulness and commitment manipulations. This concept of hypocrisy is used to some extent in this study, where the model concept proposes dissonance-generated self-persuasion by entrapment using honeypots; subsequently the rationalisations users use to justify 'attacking' the honeypot are countered by neutralisation mitigation. In a way users are 'committing' to a maladaptive position, and they are subsequently exposed to neutralisation mitigation that is posed in a way that allows them to become mindful of their actions. In this induced-dissonance state, they are shown the hypocrisy of their rationalisations. Typically 'the discrepancy between prosocial advocacy and the past transgressions arouse[s] dissonance' and consequently modifies the user's subsequent behaviour to be consistent with the prior advocacy [55]. In this research the discrepancy between users' awareness of their transgressions and their maladaptive commitment should induce behavioural change.

The dissonance-inducing effect does not work consistently for all individuals, as verified by Murray et al. [60]. They found that cognitive dissonance in an induced-compliance paradigm is ineffective 'among individuals with high levels of psychopathic traits' (i.e. lack of guilt and empathy). The induced-compliance paradigm is a form of induced-dissonance; however, it involves convincing an individual to perform counter-intuitive tasks such as deliberately lying to another person by saying that a boring task was actually enjoyable. In the classic experiment by Festinger and Carlsmith [61], individuals were expected to lie and claim that a boring task was actually 'fun'. These individuals were paid either \$1 or \$20 to lie. The group that was paid less felt more dissonance than the group that was paid more. The dissonance effect is weakened 'when individuals perceive that their dissonant behaviour occurred in response to an external motivator' [60]. Murray et al. [60] found that guilt-induced dissonance may be weak amongst individuals with higher psychopathic traits. However Freijy and Kothe [62], who conducted a meta-analysis of studies conducted in health using cognitive dissonance to change behaviour, found that the induced-compliance paradigm was not as effective as the hypocrisy paradigm. It was also found that self-concept is important for the dissonance effect. If one's self-concept is lowered then one feels more inclined to be immoral [63]. Hence the higher one esteems one's sense of morality, the greater the hypocrisy effect [59].

Aside from possible personality traits, there is a problem of determining when induced dissonance may be most effective. For example, McClurg [59], who considered the benefits of inducing dissonance to prevent corruption in the police force, suggested that it may be too late to change behaviours through induced dissonance once an individual is already corrupt. Hence the sooner one is exposed to induced dissonance the more effective it is. McClurg [59] proposes that the induced dissonance is a means to prevent good policemen from becoming corrupt and corrupting others. Once a cycle of justification is created it is very difficult to break. McClurg [59] proffers that the goal is to maintain police

officers in the pre-hypocrisy state, as it may be too late to recover an individual who is already corrupt. He termed it an 'honesty maintenance' program: that is, if a person believes that they are honest then any threat to this self-concept will result in maximizing the dissonance effect. McClurg [59] advances simulation-type exercises that give one the chance to face an integrity dilemma in advance, to think about the consequences and to learn from it.

The insider threat problem is considered to be a '"moral grey area" around the ownership of electronic data'; furthermore, the problem involves more than a disgruntled insider, it also includes 'misguided employees' [64]. Hence, unlike other moral issues such as homicide, which are unambiguous, the fuzzy nature of cybercrime allows insiders to undervalue their actions and to resort to rationalisations.

## 3.2 Neutralisation theory

Typically insiders 'do not consider themselves as criminals and have a tendency to justify their deeds' [3]. According to Farahmand and Spafford:

> rationalisation happens by insiders viewing themselves as: 1) essentially noncriminal, 2) justified, and 3) part of general irresponsibility for which they do not feel accountable. [3]

Neutralisation techniques are rationalisations that criminals use to justify a crime. These rationalisations include denial of responsibility, denial of injury, denial of the victim, condemnation of the condemners, and appeal to higher loyalties. Examples of arguments used in each type of neutralisation technique respectively are:

'It wasn't my fault.'
'It wasn't a big deal.'
'They could afford the loss.'
'They had it coming.'
'You were just as bad in your day.'
'My friends needed me. What was I going to do?' [10]

There are also other defences, such as 'defence of necessity' [65] and 'metaphor of ledger' [66]. According to Barlow et al. [8], insiders claim that because they are hardworking they can be allowed some indiscretion (metaphor of ledger), or that it was necessary to meet a deadline, for instance (defence of necessity).

As the study being reported on did not deal with specific techniques, it is beyond the scope of this article to demonstrate individual techniques as they relate to the insider theat. However, Siponen and Vance [11] and Willison and Warkentin [67] provide a thorough commentary on the role that neutralisation plays in behavioural intention with regard to the insider threat. Willison and Warkentin et al. [67] highlight the lack of empirical studies in this area and indicate that studying neutralisation techniques could enhance our understanding of the insider threat and help to determine a basis for developing intervention strategies. However, there have been two recent empirical studies on the subject [8,11]. The study conducted by Siponen and Vance [11] found neutralisation to be an excellent predictor of information security violations. Interestingly, they found that deterrence mechanisms were insignificant compared with the influence of neutralisation techniques. Barlow et al. [8], however, found that both deterrence and neutralisation mitigation are necessary to circumvent the insider threat. They also found that some types of neutralisation technique were more significant than others. However, it has been posited that if the rationalisations are removed then the insider threat can be circumvented; this is termed neutralisation mitigation.

## 3.3 Situational crime prevention

Curley and Zamoon [68] argue that for each neutralisation technique used to justify cybercrime, there must be an equal and opposing force termed 'counter-neutralisations'—the latter is particularly useful in understanding cybercrimes, as the nature of information technology is 'potentially ambiguous' with respect to 'ethical implications'. The notion of counter-neutralisations is most often used in other sociological problem areas such as alcoholism where it was found that anti-neutralisation-based interventions can be used as a dissonance-inducing strategy [69] that may lead to a person changing his/her action or perception. In the current study, the anti-neutralisation-based intervention strategy relies on techniques derived from the situational crime prevention theory.

Situational crime prevention theory considers five categories of opportunity-reducing measures, each of which is further divided into 25 specific techniques which are intended for the physical landscape:

- 'Increase the effort' includes 'target hardening', 'control of access to facilities', 'screen exits', 'deflecting offenders' and 'controlling tools'
- 'Increase the risks' includes 'extending guardianship', 'assisting natural surveillance', 'reducing anonymity', 'utilizing place managers' and 'strengthening formal surveillance'
- 'Reduce the rewards' includes 'concealing targets', 'removing targets', 'identifying property', 'disrupting markets' and 'denying benefits'
- 'Reduce provocations' includes 'reducing frustrations and stress', 'avoiding disputes', 'reducing emotional arousal', 'neutralizing peer pressure' and 'discouraging imitation'
- 'Remove excuses' includes 'setting rules', 'posting instructions', 'alerting conscience', 'assisting compliance' and 'controlling drugs and alcohol' [70]

These techniques were given 'digital analogies' by Beebe and Roa [71], Willison [15] and Coles-Kemp and Theoharidou [14]. We will consider only 'Remove excuses', as being the most appropriate to the study reported on.

In terms of 'setting rules', the typical information security policies, agreements and procedures have been proposed. In terms of 'posting instructions', e-mail disclaimers [71] are recommended as a comparable information security control, apart from the typical controls such as information security policy. Single

sign-on [15] and 'a single point of reference for security' [14] have been proposed as information security controls to realize the 'assisting compliance' technique. In terms of 'alerting conscience', the information security controls that are recommended include copyright protection [14], a code of ethics [14] and 'multi-level warning banners' [71]. The 'controlling drugs and alcohol' technique is incongruent with the domain of information security, and was not considered in this research. Note that these techniques may be modified according to the context.

According to Willison and Siponen [15], both the situation crime prevention theory and the techniques of neutralisation proposed by Wortley [72] are appropriate to the realm of information security. The techniques proposed by Wortley [72] include

rule setting, which subsumes 'setting rules' in situational crime prevention theory and involves reinforcing the illegitimacy of the targeted behaviour,

clarifying responsibility, i.e. nullifying the tactic of blaming others,

clarifying consequences, i.e. emphasising the cost to the victim,

increasing victim worth, i.e. personification of the victim.

Wortley et al. [73] elucidate that the 'remove excuses' category of situational crime prevention theory is a response to challenge neutralisations of offenders at a situational level, while the aforementioned techniques operate on the psychological and sociological dimensions. With regard to their study on cheating and shoplifting, Agnew and Peters [74] proved that 'acceptance of neutralisations will lead to deviance only when people believe that they are in situations in which the neutralisations apply'. Consequently counteracting neutralisations at the situational level is a pragmatic approach.

Ultimately, the situational crime prevention theory was selected as an intervention strategy for the following reasons. First, there is a body of work that shows that the techniques in this theory could be practically applied to the insider threat [1, 14, 15, 38] as well as to general information security concerns [5,71,75]. Second, situational crime prevention theory offers specific, practical and implementable measures (i.e. the 'remove excuses' category) to offset neutralisation. Third, the theory has been shown to be successful in other domains. Clarke [9] cites several case studies in which situational crime prevention theory has been used successfully—for example in subway systems and parking facilities. He goes on to state that even though situational crime prevention theory was initially intended for predatory crime, it has been extended to white-collar crimes (e.g. tax evasion) due to the inclusion of the 'remove excuses' techniques.

In the next section, a convergence of these three theories frames the model concept.

# 4 THE MODEL CONCEPT

As the problem domain intersects with the technical, socio-technical and sociological dimensions, these perspectives are organised into three distinct tiers within the model concept: that is, the technical indicators, sociological indicators and the socio-technical interventions tiers.

## 4.1 The technical indicators tier

Honeypots are more than just computer or physical resources—a honeypot may be anything from a Windows program to an entire network of computers. However, in its most rudimentary form it may be a credit card number, an Excel spreadsheet or a record in a database. In this form a honeypot is called a honeytoken [76]. Honeytokens are easily customizable and easily integrated [76], and for this reason the proof-of-concept of the model was based on honeytokens. A honeypot will not be effective if the insider decides not to select it because they recognize that it is in fact a trap. Interactions with a honeypot should therefore be detectable to the system administrator, but not to the insider. The model is based on a luring strategy, where the insider is baited with honeypots instead of attacking real data during their daily operations. The honeypot is used to induce cognitive dissonance.

## 4.2 The sociological indicators tier

Cognitive dissonance is a sociological indicator that directs the tension experienced when one simultaneously holds two or more conflicting cognitions (ideas, beliefs, values or emotional reactions) [7]. For example, introducing information security policy that shows the hypocrisy of their rationalisations via a neutralisation mitigation perspective will create cognitive dissonance between an insider's act of maleficence and the rules. Dissonance has magnitude: the more discrepant two cognitions are, the more dissonance is caused [12]. This forces the insider to change their actions or readjust their perception (a sociological indicator) to account for this new information. The changes caused by cognitive dissonance have been found to be more effective if the individual is intrinsically motivated to change [77]. However, cognitive dissonance may also result in the individual rationalizing their behaviour or perceptions. This issue intersects both the technical and sociological realms; hence, this model is predicated on socio-technical interventions to challenge the rationalisations of an insider threat.

## 4.3 Socio-technical interventions tier

Willison [15] proposes that the techniques advocated by situational crime prevention theory could reasonably be adopted by information security practitioners. In particular, the 'remove excuses' category is accomplished by interventions that decrease the rationalisations that criminals may use to justify their behaviour; this is highly appropriate to the study, and the interventions will be adopted here. The 'remove excuses'

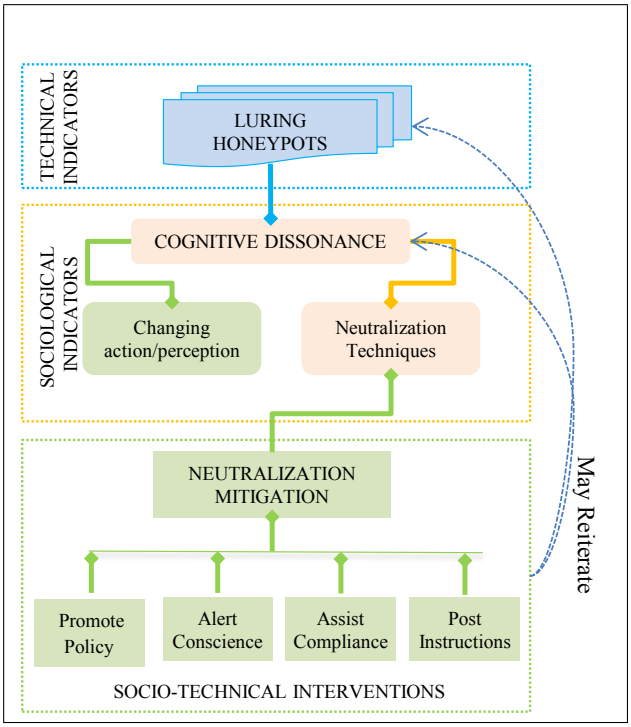


Figure 1: A first-order overview of the ITNM$^{CD}$ model

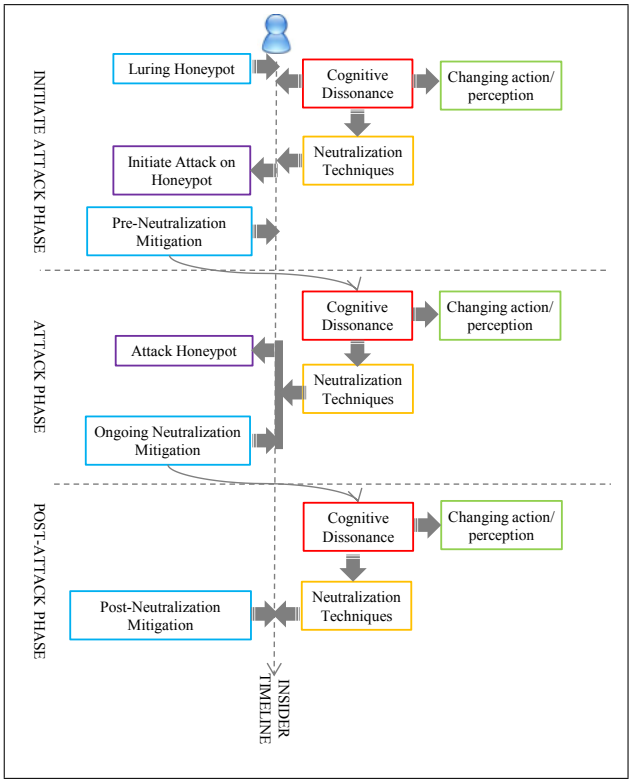


Figure 2: A timeline overview of the INTM$^{CD}$ model

category from the situational crime prevention theory provides socio-technical interventions that will be used by the model concept to decrease the rationalisations that criminals may use to justify their behaviour. The strategies proposed are 'setting rules', 'posting instructions', 'alerting conscience' and 'assisting compliance'. The 'setting rules' technique was renamed 'promoting policy' to make it more congruent with the cyber domain. However, there has to be a means to detect when to activate these inventions. Consequently, honeypots are a useful technical indicator in detecting the insider threat and activating the interventions in a controlled environment.

## 4.4 The model overview

Honeypots are technical indicators that are used to detect an insider threat and activate the socio-technical interventions. The sociological indicators of cognitive dissonance force an insider to either use neutralisation techniques (i.e. rationalisations) or to change their perception or behaviour. The neutralisation mitigation mechanisms encompass the socio-technical interventions of 'promoting policy', 'posting instructions', 'alerting conscience' and 'assisting compliance' to show the hypocrisy of their rationalisations.

This implies that neutralisation mitigation may be implemented from both a technical and a sociological perspective, depending on the context. Neutralisation mitigation techniques are a way of removing the rationalisations that criminals use to commit a crime by propagating the organisation's information security policy. The conflict between the insider's neutralisation techniques and the neutralisation mitigation causes cognitive dissonance. This forces the insider to change either their behaviour or their beliefs positively towards compliance. However, the insider may choose to further rationalize their behaviour or beliefs by choosing an alternative neutralisation technique, thus compromising compliance (see Fig. 1). The processes may iterate for reconnaissance purposes, or neutralisation mitigation may lead to state of cognitive dissonance.

Sykes and Matza hypothesised that criminals apply neutralisation techniques before committing an act [10]. However, research has shown that neutralisation techniques could be used prior to, during or after criminal involvement [78]. The model is reviewed from a timeline perspective involving the initiate-attack, attack and post-attack phases. This idea is loosely based on the usage control model of Sandhu and Park [79].

## 4.5 A timeline perspective of the model

Fig. 2 demonstrates a timeline perspective of the model concept.

In the initiate-attack phase, the honeypot places the insider in a state of cognitive dissonance—the insider may be motivated to attack the honeypot. While in this state of cognitive dissonance, the insider may experience a sense of unease or tension created by the desire to attack the honeypot and violate the information security policy. However, some individuals choose to rationalize their actions through neutralisation techniques. The pre-neutralisation mitigation will challenge their rationalisations once they initiate the attack on the honeypot. This leads to another state of cognitive dissonance—perhaps an uneasy tension while initiating the attack and the neutralisation techniques (i.e. the justifications for the attack). In the attack phase, the insider rationalizes the act and attacks the honeypot (for example by copying or editing the honeypot). This simultaneously triggers the

ongoing neutralisation mitigation that challenges the neutralisation techniques once again during the attack on the honeypot.

This leads to the post-attack cognitive dissonance that is probably a state of regret at having attacked the honeypot. In the post-attack phase, the insider may choose either to use neutralisation techniques to justify the act of maleficence or to change their behaviour or perceptions of compliance. Post-neutralisation mitigation then challenges the neutralisation techniques once again.

It is evident that if the insider continues after the post-attack phase to employ neutralisation techniques to justify their actions, the neutralisation mitigation process has failed. This is an indicator that a more stringent intervention is required. In this context, it is difficult to discern whether the neutralisation mitigation has succeeded or failed, as these socio-indicators are intrinsic to the insider unless the insider opts out. However, any interaction with a honeypot is detectable and guaranteed to be illicit. Hence, an insider who 'attacks' the honeypot should be subjected to alternative mitigations such as rigorous monitoring, auditing and training. To demonstrate the feasibility of the model concept, a simple prototype was implemented.

## 5 PROOF-OF-CONCEPT

Insider attacks take place at application level where insiders have access to client records—this scenario was used as a case study for the proof of concept (see Fig. 3). The insider is lured by honeytokens on querying the database. This technical design demonstrates one possible way of applying the model concept, which is an over-simplistic interpretation of the model concept.

In this interpretation, the technical design has the following core modules: the 'honeytoken generator' and the 'neutralization mitigator' (pre-, post- and ongoing). As the user queries the database, they are presented with real data and honeytokens (stored in the 'honeytoken database').

If the insider decides to act upon (i.e. 'initiate attack') the honeytoken, then it is evident that this act could be a threat. Thereafter the insider is subject to neutralisation mitigation, as this process interfaces with the socio-technical interventions—that is, the insider may be counteracted by means of both technical and sociological controls prior to, during and after the attack. In this interpretation of the model concept from a timeline perspective, the insider is confronted with a pre-neutralisation mitigation warning about accessing the data, in an attempt to 'alert conscience'. The insider is allowed to view the honeytoken data, but is subjected to ongoing neutralisation mitigation in the form of a banner designed to be displayed alongside the data to 'post instructions' on the usage of the data. This is a parallel process. After the insider has closed the access, they are confronted with post-neutralisation mitigation that attempts to 'alert conscience' by subtly determining the justifications for the access. Each justification is then mapped to a warning that will 'assist compliance' and 'promote policy'. Although this was not demonstrated in the prototype as this type of information is highly contextualised, the idea is to promote policy and assist compliance in a manner that directly challenges rationalisations. To gather more evidence and to provide another opportunity for neutralisation, the process reiterates and the 'honeytoken generator' generates more honeytokens based on the insider's queries to the database using the 'honeytoken database'. As the honeytoken generator was not the focus of this research, the proof-of-concept prototype did not demonstrate this aspect of the design. Screenshots of the prototype are shown in Appendix B.

## 6 RESEARCH METHODOLOGY

The design science research methodology was leveraged to conduct a small-scale experiment based on the following activities: build, evaluate, theorize and justify [81]. The experiment involved a problem identification stage, design and development of prototype stages and an evaluation stage [82]. The evaluation utilised information security practitioners from various organisations. The purpose of this process was to identify whether there were any vacuities, ambiguities or inconsistencies in the model concept. During the evaluation stage, the participants viewed a demonstration of the prototype and the model concept via online videos[1] and provided value judgements on them in terms of the efficacy of the security mechanism provided by the product concept. The research model is shown in Fig. 4. Both qualitative and quantitative data collections were employed, and involved open-ended questions and a structured questionnaire. The structured questionnaire was used to determine the participants' perceptions of the appeal of the model concept in terms of containing the insider threat. Participants had to consider the following in terms of the model concept: viability (i.e. feasibility of the model), utility (i.e. value), efficacy (i.e. effectiveness), usability and scalability. To facilitate the process, the issues relating to the evaluation were formulated in 13 statements. The participants provided a judgement on each statement (see Appendix A). To ensure rigor, the four principles of Österle et al. [83] were used as basis for the in-depth interview:

- Abstraction: Each artefact must be applicable to a class of problems (in other words, must not be specific to a unique problem).
- Originality: Each artefact must contribute substantially to the advancement of the body of knowledge.
- Justification: Each artefact must be justified in a comprehensible manner and must allow for its validation.
- Benefit: Each artefact must yield benefit either immediately or in the future for the respective stakeholder groups.

[1] https://sites.google.com/site/theinsiderthreatproject/

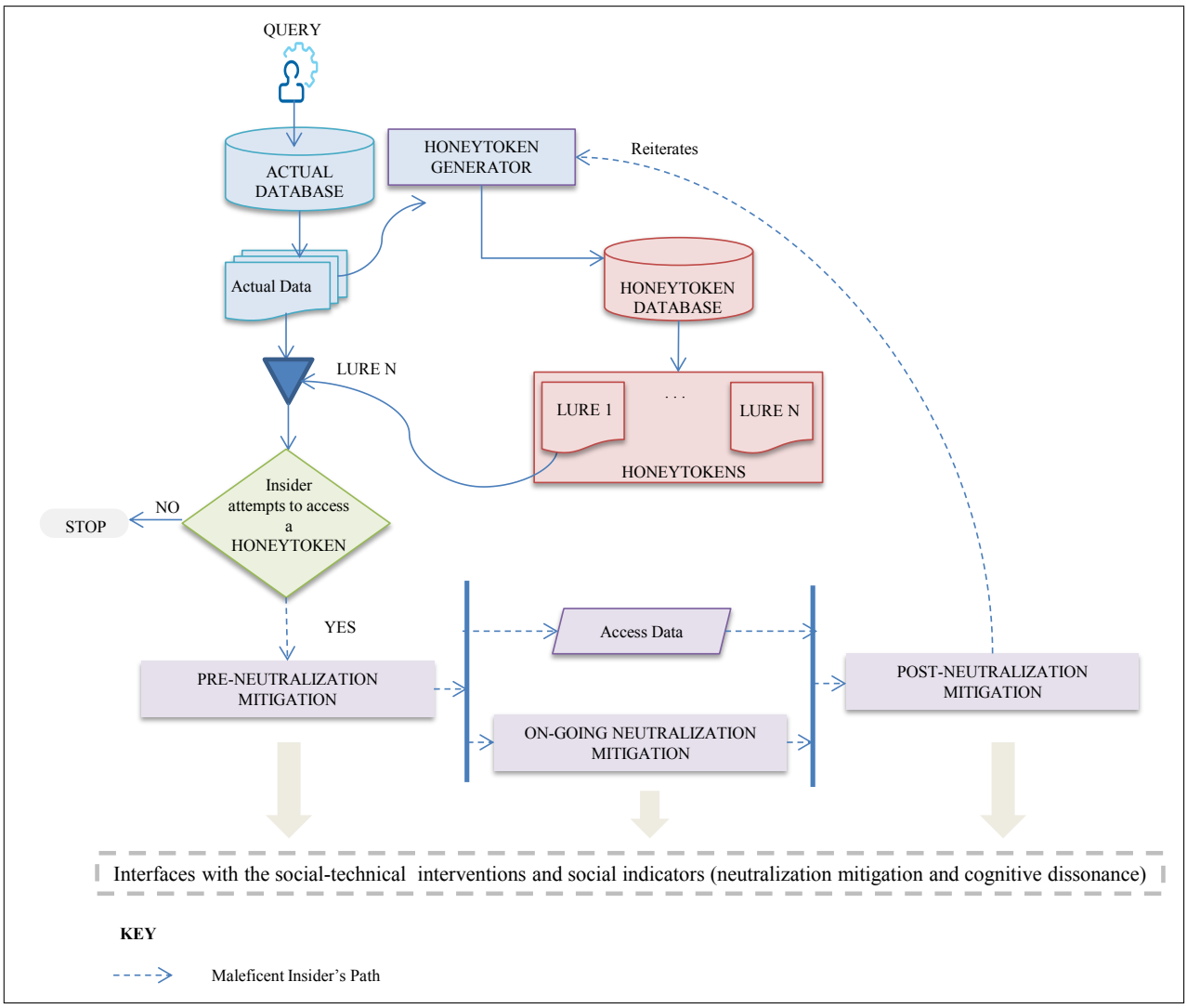


Figure 3: A proof-of-concept of the INTM$^{CD}$ model

According to Peffers et al. [80], the process of design science research involves defining the following: Problem statement, Objectives, Design and development, Demonstration and Evaluation and, finally, Communication. The problem statement is essentially concerned with the susceptibility of organisations to the insider threat. The hypothesis is that the model should increase an insider's compliance intention.

## 7 DATA ANALYSIS

The sample size was ($n = 25$), and purposive sampling was used. A professional social media network was used to invite 200 participants with information security expertise to join the researcher's professional network. The response rate was 12.5%. The majority of respondents (44%) occupied a supervisory role. The participants were grouped into categories as shown in Fig. 5: Administrators ($n = 1$), Information Security Analysts ($n = 2$), Information Security Operations Supervisors ($n = 3$), Information Security and Risk Supervisors ($n = 3$), Information Security Supervisors ($n = 5$), Information Security Technical Specialists ($n = 5$) and Information Security Specialists ($n = 6$).

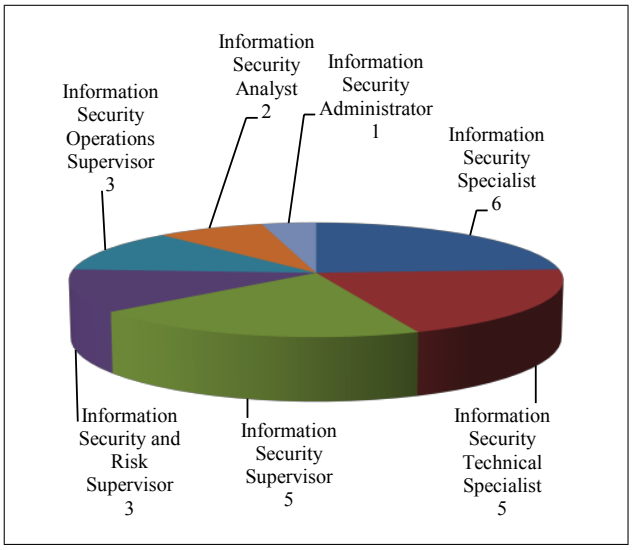


Figure 5: Profile of participants

### 7.1 Results of the value judgements

In terms of the value judgements, viability and scalability factors were ranked high; however, the usability factor was ranked poorly. The utility of the model con-

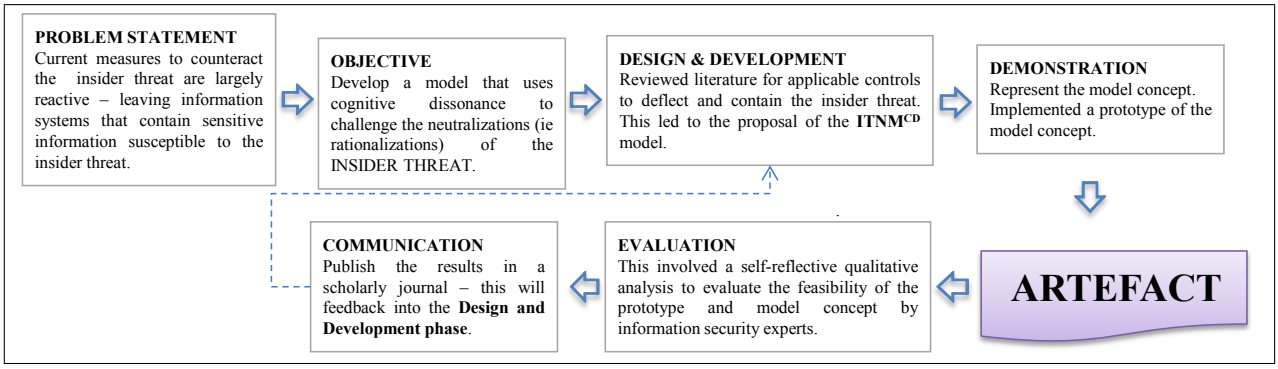


Figure 4: The phases in the research design (adapted from [80])

cept was ranked relatively high at 70%, while efficacy was ranked above average at 58% (Fig. 6).

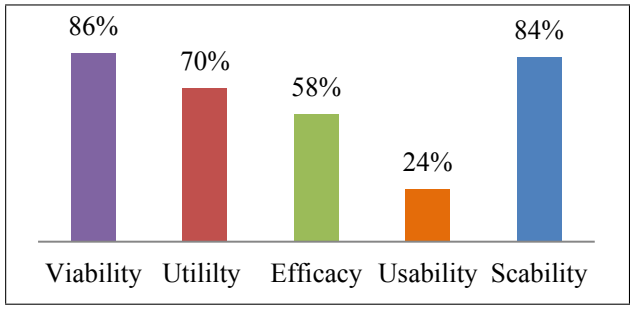


Figure 6: Value judgements

### 7.1.1 Viability

Overall, viability, which was based on the factors of implementability and integratability, was ranked at 86%.

#### 7.1.1.1 Implementability

In terms of implementability (question S1), 80% of the participants agreed that the model concept could be easily translated into an implementable product. However, a few criticisms were raised. The participants who disagreed with the statement cited the following concerns:

- limited applicability:

  > Although the idea behind the model is sound, I don't see how you will make this universal enough to make it implementable on a broad basis.
  >
  > (Participant #3)

- limitations of honeypots/honeytokens, such as false positives:

  > I don't understand why it is said that any interaction with the honey token is guaranteed to be malicious. If the honey token looks like real data, and access to the real data is part of day to day activities, why would access to the honey token be deemed a guaranteed attempt at breaching security? I understand the merit of the popups etc., my doubts lie with the honey token itself.
  >
  > (Participant #19)

- complexity of integration:

  > Although the concept could work from a compliance perspective and awareness perspective, IT consists of various artefacts or elements e.g. application, database, network and it could be difficult to translate the concept into all elements. For example an IT network specialist maintains a router. The model concept only focused on access to an application by a user.
  >
  > (Participant #22)

#### 7.1.1.2 Integrability

In terms of integrability (question S2), 92% of the respondents agreed that the model concept could be integrated into an existing system. Some respondents indicated that legacy systems with a closed architecture might pose a problem. This was exemplified by the following response by Participant #22:

> Proactive thinking will be required by information security practitioners to apply this model. For example where the user was accessing the honeypot practitioners will need to think what policies have been violated etc. Further systems are already designed and implemented where users have minimum rights or on a "need to know" basis. In addition to implement this on legacy systems will be a challenge.

Participant #20, who disagreed, qualified the response by indicating that it depends if there is a distinction between real and honeytoken data. This participant remarked:

> My answer is same as with S1. If there can be a better distinction between the honey token and real data, it might work.

This participant challenged the effectiveness of honeytokens in the previous question (S1), by asserting that if the honeytoken data is not distinct from real data, how can any interaction with a honeytoken be considered malicious? In other words, honeypots/honeytokens

may lead to false positives; however, a distinction discernible to the user will defeat the purpose.

#### 7.1.2 Utility

Utility was ranked overall at 70%; this characteristic was dependent on the deterrent and compliant effect of the model concept, the utility of honeypots and the utility of cognitive dissonance.

##### 7.1.2.1 Utility of deterrence

In terms of deterrence (question S3), 60% of the respondents agreed that the model concept would deter the insider threat from abusing their privileges. However, those respondents who disagreed also conceded that the model may be successful in certain situations. The concerns raised by those participants who disagreed with the statement related to whether the model concept would deter the insider threat, since so much is contingent on the user's disposition (i.e. motivation, status, morality and ethics).

The issue of motivation is complex and no model concept can account for the myriad of motivations that drive the insider to commit maleficence. Hence the success of the model is contingent on the motivation of the insider; this is evidenced by the following claims:

> The motivations driving insiders are complex. You will only be able to deter a small segment of insiders. At best I think this will only postpone malicious behavior [sic]. A malicious insider is likely to find the path of least resistance.
>
> (Participant #3)

> It will enhance social awareness and to a degree it may limit insider threat for example unnecessary access is limited or user will intentionally lock his pc, however, an insider is motivated by a number of reasons e.g. disgruntled employee or fraudster. Therefore there is a specific intend [sic] to commit a crime or violation and that neutralisation to this respect [sic] may not be possible.
>
> (Participant #22)

While Participant #3, who disagreed with the statement along with other participants, felt that the concept would deter the insider threat, they however indicated it may work for a subset of users. For example, Participant #4 indicated: 'This may work for some employees but there are those who would be unfazed'. Similarly Participant #21 stated: 'They will always find a justification, it may prevent those that are sitting on the fence though.'

The contingent nature of the model was further validated by participants who felt the success of the model was dependent on the morals and ethics of the insider. For instance, some participants asserted:

> The model concept may deter the threat for a subset of insiders; it will deter the insider threat only if positive assumptions are made as to both the insider's morality and ethics.
>
> (Participant #11)

> Not sure about this—it might deter normal employees with good morals and a conscience, but not criminals who are out to steal info. :-) You are presumably talking about an environment where there is nothing like identity management or rbac (role-based access control)?—yes, then it would provide a measure of protection. I would not throw out rbac and IDM (identity management) yet. :-)
>
> (Participant #17)

Participant #14 asserted that the success of the model is contingent on the level of access granted to an insider:

> The threat will always be there. It might deter a few low level employees but sys(tem) admins or DBA's are harder to control because of access levels they have.

Participant #7, who agreed with the statement, suggested that as a means of increasing the efficacy of the model there should be penalties as well. Participant #7 provided the following disclaimer: 'It will only work if there are accompanying consequences for accessing inappropriate records.'

##### 7.1.2.2 Compliancy

In terms of the compliance effect of the model concept (question S4), 64% of the respondents agreed that the protection mechanisms, such as neutralisation mitigation, would compel the insider to comply with the established rules of behaviour in order to protect confidential information in the future. Once again, the issue of the insider's disposition and the need for penalties were highlighted. The effectiveness of neutralisation mitigation is contingent on the motivation of the insider, as exemplified by the following comments:

> A motivate(d) insider will only deterred [sic] temporarily and find alternative avenues to obtain confidential information.
>
> (Participant #3)

> It might be effective on some insider threats depending (on) their level of motivation to comply.
>
> (Participant #15)

> A criminal element consists of a motive and therefore the fraudster would have broken all rules to commit the crime. This mechanism will enhance protection of systems to a degree but it will not limit the ones with criminal intends [sic].
>
> (Participant #22)

Participant #8 considered the condition that an insider may motivated by external forces:

> Suspect it will only be partially effective since organised crime is targeting internal attacks by partnering with internal resources.

Participant #21 reflected that the model may be effective in instances where the insider is ambivalent:

> There are always some that will commit offences, it should however prevent those that may be undecided not to.

Participant #7 proposed that the success of neutralisation mitigation is contingent on penalties:

> It will only work if there are accompanying consequences for accessing inappropriate records. Just providing warnings and links to policies is not enough.

Likewise Participant #17 echoed Participant #7's sentiment that the success of the model should be complemented by other controls:

> I do think that this is s [sic] novel approach, but I would still deploy technical controls to assist with enforcing policy. It is the same as when looking at DLP products—my idea of the importance of my salary is not the same as yours—that is—if you let the human decide, and not use technology to make the decision.

Likewise Participant #11 was not convinced that cognitive dissonance is effective:

> Cognitive dissonance by itself is not compelling. It can be reduced by justifying the cognition through changing cognitions or adding new cognitions. If cognitive dissonance were compelling, there would be no smokers of (legal, health warning carrying) cigarettes in South Africa.

#### 7.1.2.3 Utility of honeypots

In terms of the utility of honeypots (question S5), 60% of the respondents agreed that honeypots would support an organisation in mitigating insider attack risk by focusing the attacker's attention on decoy assets, while the critical assets remain protected. Here the majority of the participants indicated the value of honeypots in terms of containing damage, providing insight into the attacker's methods, motives and behaviour. However, on the negative side, some participants expressed the opinion that the effectiveness of honeytokens is transitory: for instance, Participant #3 indicated: 'Once discovered, the existence of a honeypot will be disclosed. It will loose [sic] its mitigation effect.' This sentiment was also shared by Participant # 23:

> This will depend on how much does the insider attacker understands [sic] the infrastructure or system. But if [it] is someone who is new in the organisation and also is not in IT or Decision making level he/she won't be able to know about the system and how it work [sic] but therefore the system will be able to catch him. In general the system will operate well so my Answer is either YES OR NO it just depends, with the trust level.

Participant #14 highlighted the limitations of honeypots by remarking:

> It would mitigate some risk but not all. Honeypots are a good deterrent but would inevitably be circumvented.

Participant #15 commented on the limitations of the applications of honeypots:

> I have doubts honeypots will provide much of a serious distraction to a [sic] insider threat. At most it can act as a [sic] "alarm" to notify if a transgression has taken place. Also insider threats are broader than just accessing confidential information, but also could include, process abuse (i.e. stopping a critical service), adversely affecting the integrity of business data etc.

Participant #20, who raised concerns about honeypots being a form of 'security through obscurity', stated:

> I am in two minds about this one, honeypots are great for research and to establish methods used by attackers. Information learned from honeypots must then be used to design better systems. I don't think that honeypots must act as the protection because this seems to be 'security through obscurity'. Data must be protected even if the attacker knows exactly where the data is, as an insider probably would.

Participant #25 raised concerns about the intent of the insider, stating:

> This also depends on the focus of the attacker, if they require only specific information and do not stray from what they want. These attacks may not be protected.

Those participants who agreed highlighted the positive aspects of honeypots. For example, Participant #4 remarked: 'This may also be used to monitor the behavior [sic] and trends of certain suspicious inside users.' Similarly Participant #22 commented on the ability of honeypots to determine behavioral [sic] trends:

> Honeypots are a great measure to limit damage and enhance the protection of assets. Honeypots also gives [sic] insight to attacker's methods and motives.

#### 7.1.2.4 Utility of cognitive dissonance

In terms of the utility of cognitive dissonance (question S6), 96% of the respondents agreed that cognitive dissonance might be a useful technique in encouraging compliance. However, this was dependent on morality and ethics, and had to be used in conjunction with other techniques.

Participants who agreed provided the following substantiations. In support of a comprehensive approach, Participant #1 remarked:

> Yes to some degree subject to full testing and changing where needed. For now the

> technique seems to be practical but I still do recommend full POV (point of view).

Similarly, this was supported by Participant #4, who said:

> This and other technical methods would help with success rate. On its own I don't think it could work.

Participant #11 felt the success of the system depended on morality, as evidenced by the following remark:

> Agreed provided that selection criteria for personnel ensures similar value systems in terms of morality and ethics.

Participant #21 felt it would work in certain circumstances:

> To a degree I agree, if it dissuades some people that may have been potentially thinking of performing illicit actions then it has worked.

Participant #3, who disagreed with the statement, indicated: 'The implementation variation is too complex to make it practical.' On the contrary, Participant #22 who agreed with the statement observed:

> The user is confronted with warning messages which is practical. This is far better than conventional training materials e.g. video or slideshow.

#### 7.1.3 Efficacy

Efficacy was ranked at 58% overall, this being based on the overall efficacy of the model concept, the efficacy of honeytokens, the efficacy of cognitive dissonance and the efficacy of neutralisation mitigation.

##### 7.1.3.1 Efficacy of model concept

In terms of the efficacy of the model concept (question S7), 56% disagreed that other mechanisms such as a written policy document or adequate training would have been more effective than the mechanisms identified in the product concept. Participants expressed the view that this model concept would work in conjunction with other methods (i.e. policy and training).

Some participants indicated that other mechanisms such as training and segregation of duties, logging and monitoring would be more effective. For example, Participant #1 declared: 'Nothing beats effective information security training awareness'. Participant #23 stressed the importance of training in tandem with the model concept:

> Training is a very important factor when it comes to information security issues, people needs [sic] to be trained annually even if they think they know and you can also have posters in bathroom [sic] and other places. This is to make sure that people understand things, since applications alone can't defend insider threats. So training plus this application will make a great difference.

In contrast, Participant #3 professed:

> I think automated segregation of duties, tagging and tracking data records, DLP, finely tuned logging and monitoring are better techniques in enforcing security.

Even though some participants agreed that other mechanisms would be more effective, they substantiated their answers by proposing a combination of techniques. This evidenced by Participant #4:

> I don't believe it should be a choice between having one or the other as an enforcement method. Also, written policies signed and accepted by the user can also be considered neutralisation techniques.

This sentiment was shared by Participant #17:

> I agree. Policies etc. are administrative controls. We use technology as technical controls to actually enforce the policies. Cognitive dissonance should be seen as, and used as an additional weapon in the arsenal. In this sense, one cannot say that cognitive dissonance would be more effective than policies or vice versa. Only technical control—which are [sic] monitored, can with a certain degree of accuracy be seen as effective. Think of not stopping at a stop street—most people generaaly [sic] do not stop at a stop street. No matter what cognitive mechanisms you use. The [sic] WILL however stop when they know they are being watched by someone who can take action against them—such as a cop! :-)

Those participants who disagreed also proposed that a combination of all the techniques would be appropriate; for example Participant #7 indicated:

> Policies and training is [sic] critical to the success of any Information Security program, but is flawed due to the human factor. Having a mechanism in place as in the product concept in combination with the policies and training will be far superior.

Some participants who disagreed with the statement provided substantiations as to why they deemed policies and training to be ineffective. For example Participant #8 declared:

> By nature humans do not read the policies they supposed [sic] to. Even popup messages fail to be effective. Therewith adequate training is extremely difficult to implement without spending a huge amount of money which industries today reduce due to cost optimisation to a level that clear [sic] audits.

Participant #19 remarked:

> Policies teaches users what he/she may or may not do regarding the system. Training only teaches the user how the system works. Neither will have the same affect [sic] as the model concept.

Other participants were of the opinion that the model concept may be considered as entrapment, which has legal implications. For instance, Participant #21 stated:

> It is important to make sure the user is aware of policies and procedures otherwise there can be no action taken against them, however if you look at some legislation you may need to give a person an opt-out feature otherwise it may be deemed unconstitutional.

Similarly Participant #11 noted:

> Enforcement mechanisms or controls that would be more effective would be preventative. Examples are authorisation controls that would ensure that the honey token would not be displayed in the first place. Unless of course the objective is to determine the individual's propensity to default, in which case the proof of concept described would probably succeed. This view needs to be tempered with the country's labour and criminal legislation (entrapment comes to mind).

Two participants emphasised the difference between the static nature of training and policies versus the dynamism of the model concept:

> The mechanisms used in the product concept (e.g. honeypots) catch the user in the act of security transgression compared to the policy and training.
>
> (Participant #10)

> Staff members rarely reads and/or un[derstands] [sic] policy documents. Training [is] also not 100% effective in addressing security awareness. In this scenario the user is confronted with warning messages which is practical.
>
> (Participant #22)

#### 7.1.3.2 Efficacy of honeypots

In terms of the efficacy of honeypots (question S8), 68% agreed that the benefits of a honeytoken might be counteracted if insiders waste time and resources interacting with them. The participants who provided negative feedback explained that insiders may overload the system resources by interacting with the honeytokens, depending on their pervasiveness, that innocent users may waste time interacting with the honeytokens, and that a honeytoken is based on the assumption that the attacker does not know exactly what they are looking for. Those who agreed that honeytokens may affect productivity provided the following substantiations:

> [May] cause confusion to some users which might lead them to interrogate the presence of it.
>
> (Participant #1)

> If the insider spends most of their time on the honeytoken, the system will be overloaded and it might crush [sic] or start experiencing some problems.
>
> (Participant #2)

> Application users might be curious in finding all the honeytokens.
>
> (Participant #3)

> This assumes the attacker does not know exactly what he/she is looking for.
>
> (Participant #20)

> This may impact productivity and efficiency especially in highly integrated IT environments e.g. call centres.
>
> (Participant #22)

> May get false positive results, some users may just be curious. Sometimes these curious users find find [sic] systems information that may lead to other more glaring security issues, which they inform you about.
>
> (Participant #25)

Those participants who disagreed that honeytokens may affect productivity offered the following counter-arguments:

> Honeytokens can give valuable information for security in terms of viewing which user tried to violate which policy.
>
> (Participant #4)

> Malicious insiders will be wasting time irrespective of there being honeypots in place or not.
>
> (Participant #7)

> I'd rather have a user waste time than leak important information and or costing [sic] the company millions of rands.
>
> (Participant #19)

#### 7.1.3.3 Efficacy of neutralisation mitigation

In terms of the efficacy of neutralisation mitigation (question S9), 60% of the respondent disagreed that most insiders will not benefit from neutralisation mitigation, as they will ignore the process. Participants who agreed that neutralisation mitigation was not beneficial cited the following reasons:

> Insiders (similar to hackers) seek the path or (of) least resistance.
>
> (Participant #3)

> A determined insider with a reason (however trivial) to default will not benefit from neutralisation mitigation as described.
>
> (Participant #11)

> Unless you enforce it somehow, but I think helpdesk will be swamped with calls to disable it :-)
>
> (Participant #17)

Participants that felt that neutralisation mitigation may be beneficial but offered the following caveats:

- the benefits of neutralisation mitigation are transitory:

  > I believe that this is a great tool that can enhance existing security measures. It may however loose [sic] its weight when there is a process every time data is accessed
  >
  > (Participant #6)

- neutralisation mitigation should be enforced with penalties:

  > Re-enforcing a message constantly, however there is a change [sic] it will simply be ignored after a time unless it is shown to have consequences.
  >
  > (Participant #21)

- neutralisation mitigation may be appropriate for lower-end staff:

  > Users will most definitely benefit from this. IT staff and senior management will ignore this.
  >
  > (Participant #22)

- neutralisation mitigation will not thwart sophisticated/highly motivated insiders:

  > If someone wants to commit a crime they will take their time, to learn how to trick the system so I believe that the organisation will have a good benefit if the attack is not well planned and organised but if its organised the organisation will not benefit but the attacker will do so.
  >
  > (Participant #23)

#### 7.1.3.4 Efficacy of cognitive dissonance

In terms of the efficacy of cognitive dissonance (question S10), 84% disagreed that cognitive dissonance has no influence on compliant security behaviour. Here participants referred to the ethical considerations of the insider and expressed the view that cognitive dissonance would be effective if the insider was generally moral.

Participants who disagreed highlighted the morality and ethics dependency, while some participants disagreed with the order of the process:

- cognitive dissonance is ethically and morality dependent

  > it will have an effect on people with value systems that hold ethics and morality in high regard.
  >
  > (Participant #11)

  > the ethical side of the user has complied with the security requirements
  >
  > (Participant #19)

- the order of process should be reconsidered

  > I believe there is a place for this—maybe do the post-action check before they actually retrive [sic] the data—i.e. let them state upfront they understand they are being monitored/give reasons why they retrieve the data
  >
  > (Participant #17)

- it must be used in tandem with neutralisation mitigation

  > It has impact through neutralisation. Users are educated through awareness of security policy.
  >
  > (Participant #22)

Participants who agreed that cognitive dissonance had no impact on compliance offered the following substantiations for their perspective:

- cognitive dissonance may incite the insider threat

  > likely (to) increase the insider's determination to achieve their criminal objective.
  >
  > (Participant #3)

- honeypots which stimulate the cognitive dissonance may create false positives

  > a compliant person might inadvertently access a honeypot if there are multiple records returned during a search.
  >
  > (Participant #7)

#### 7.1.4 Usability

In terms of usability (question S11), 76% of participants agreed that the technical mechanisms may be distracting to a user. Participants who agreed that the technical mechanisms were distracting provided the following justifications:

- It depends how well the honeytoken is customised 'to specific set of audience [sic].' (Participant #1)
- The concept has transient benefits as users 'will eventually avoid it' (Participant #2) or 'repetitive acknowledgement may eventually just be accepted blindly with no real effect' (Participant #6).
- The concept may cause 'fear amongst users who accidently' access the honeytoken (Participant #3).
- It may hamper productivity, as it may 'distract and hinder user-computer interaction' (Participant #22)

Participants who disagreed offered the following substantiations:

> Could be streamlined. I think the principle is solid.
>
> (Participant #17)

> It's a necessity which will repeatedly remind the user of the sensitivity and importance of the data and thus will be the reminder to do good with the information.
>
> (Participant #19)

> Maybe only once/first time after logon.
>
> (Participant #24)

#### 7.1.5 Scalability

The scalability of the model was ranked at 84%, based on consideration of the practicality and applicability of the model concept.

##### 7.1.5.1 Practicality

In terms of practicality (question S12), 72% of the respondents agreed that this model would be scalable in a real-world context. Participants who agreed that the model was scalable offered the following caveats:

- change management may impact the scalability

  > ... however change management aspects of it will make or break this model.
  >
  > (Participant #1)

- dependence on the system size

  > Although dependent on the size of the system, the mechanisms doesn't [sic] have to be implemented on every transaction: only classified or high sensitivity tasks.
  >
  > (Participant #19)

- accounts for the advent of social media

  > Although a lot of research is still needed to take real world context into account. For example user behavior [sic] over social media.
  >
  > (Participant #22)

Participants who disagreed that the model concept was scalable offered the following substantiations:

- lack of variation

  > There are so many different applications. Who do you target? How do you ensure the element of surprise?
  >
  > (Participant #3)

- requires highly skilled resources to maintain it

  > The model will greatly depend on the data sensitivity and proper data classification; and the balanced implemented between real-world-context and control level. The greatest threat to its successful use will be skilled resources to maintain the model post deployment since all security controls need to be reviewed continuously.
  >
  > (Participant #8)

- subject to cultural relativism

  > The real world context presents a diverse range of value systems and beliefs. The world view on truth is an example—many people believe that truth is relative, not absolute, and justify their viewpoint.
  >
  > (Participant #11)

##### 7.1.5.2 Applicability

In terms of applicability (question S13), 96% of the participants disagreed that there are no conceivable environments in which this product concept will be applicable. Participants who mostly agreed that the model concept had potential applicability offered the following observations:

> I think this might be useful as a standalone honeypot solution, where unauthorised access to data triggers automated responses which warn the insider.
>
> (Participant #3)

> The concept was demonstrated in a real world scenario.
>
> (Participant #10)

> This concept would be applicable in an environment where people have similar value systems.
>
> (Participant #11)

> Mobile environment etc.
>
> (Participant #17)

> This product will be applicable in systems where sensitive and classified information will be handled.
>
> (Participant #19)

> This might be applicable in some situations. Especially with users that are more vulnerable to social engineering, such as telephone operators or receptionists. If they are inadvertently committing a crime, warning messages may work as indented. [sic]
>
> (Participant #20)

> This concept has relevance to user environment.
>
> (Participant #22)

It is clear that while the participants did concede the model was applicable, they felt that it would be more appropriate in specific environments and it was also dependent on the value systems of users in that specific environment.

### 7.2 Validation

The design science methodology ensures rigor by validating the abstraction, originality, justification and benefit dimensions of the model concept.

#### 7.2.1 Abstraction

In terms of abstraction (i.e. does the model concept help to solve the insider threat problem in general?), 52% of the participants partially agreed and indicated that this solution should be part of an integrated insider threat solution that also incorporates training, sanctions and auditing.

The participants who partially agreed about the pragmatism of the solution provided the following substantiations for their view:

> To a degree, awareness is the first step, aligned with a "scare tactic" and the risk of an audit this [sic] would be more effective than normal training.
>
> (Participant #2)

> The model will definitely have an impact in reducing the insider threat overall. Unfortunately there will still be the more persistent insider threat that will continue even with the mechanisms in place.
>
> (Participant #19)

> Partially. IT consists of various artifacts or elements e.g. application, database, network and it could be difficult to translate the concept into all elements. For example an IT network specialist maintains a router.
>
> (Participant #22)

> Partly yes, but careful your users might feel like 'big-brother' is always watching over their shoulder.
>
> (Participant #24)

Of the participants, 28% agreed that this solution would help to solve the insider threat problem. Some participants provided the following substantiations for their endorsement:

> Yes. The concept provides active mechanisms (e.g. pre-warning and active banner) in solving the insider threat.
>
> (Participant #10)

> It would reduce the loss of data and could provide a mechanism to measure the effectiveness of the policy, provide a maturity matrix of level of compliance.
>
> (Participant #13)

Some participants (20%) completely disagreed with this validation and reflected on the limitations. Two of the participants who disagreed with the statement commented that the success of the model is dependent on the sophistication of the insider:

> Some insiders have extensive knowledge of their control environment, so this will minimise the insider threat problem, not solve it completely.
>
> (Participant #5)

> It does to a point, as mentioned certain users have high access levels and think security policies do not apply to them.
>
> (Participant #14)

Some of the reasons cited to explain why the model would not work included the sporadic motivations of the insider, the limitations of honeytokens and the unpredictability of insiders:

> It does not solve the problem in general because the motivation of insiders varies drastically from the curious insider probing data areas of the organisation to highly motivated individuals who may be part of commercial crime syndicates. The model will address the former type of insider threat more efficiently than the latter type of insider threat.
>
> (Participant #18)

> No, because the honey-token approach assumes the attacker does not know what data he/she is looking for.
>
> (Participant #20)

> Not really as there are ma(n)y more ways that users access data as well as many more data types other than that which resides in databases.
>
> (Participant #21)

Participant #11, who also commented on the limitations of the model concept, however also highlighted the significance of the concept:

> No, it certainly does not solve the insider problem, but it assists us in providing insight into the soul of the human being and generates discussion which could lead to further enlightenment.

### 7.2.2 Originality

In terms of originality (i.e. does the model concept contribute to the advancement of the body of knowledge in information security?), all of the participants except one agreed that the concept was original. The model concept clearly is original; however, the implementation used existing technologies in an original fashion. Some participants found the social aspect of the model novel because it questions why people (insiders) default, and potentially leads to a greater understanding of human beings. It highlights the psychological/human factor and people's response to mechanisms placed at strategic points to make them rethink their behaviour.

While most participants remarked on the originality of the model concepts, some highlighted the limitations of the concept:

> Yes. If the user is made aware of security policies each time they access something then they are likely to become accustomed to it.
>
> (Participant #4)

> Yes. Honey-pots have great uses. But I think using them in terms of "security through obscurity" is not the correct way to approach the issue of insider threat.
>
> (Participant #20)

Some participants highlighted the significance of the model by reflecting on its social aspects:

> Yes. Insiders will be more aware and are more likely to stop security violations.
>
> (Participant #10)

> Yes, because it questions why people (insiders) default and potentially leads to a greater understanding of the human being.
>
> (Participant #11)

> It could be an enhancement to the data classification control objective and contribute towards the assets identification and protection of such an asset.
>
> (Participant #13)

> Yes this concept will be able to further develop the human factor in Information Security.
>
> (Participant #16)

> Yes, it raises the possibility of the types of actions that can be taken when users attempt to breach organisational access controls and could offer interesting mechanisms in data loss protection mechanisms.
>
> (Participant #18)

> Yes it does. It highlight [sic] the psychological/human factor and their reaction to mechanisms when placed at strategic points to rethink their behaviour.
>
> (Participant #19)

#### 7.2.3 Justification

In terms of justification (i.e. is the model concept justified in a comprehensible manner?), all of the participants except one were of the view that the model concept was justified. This was exemplified by the majority of participants stating that the concept was expressed in a comprehensive manner:

> Yes. The concept is clear and rational behind it [sic] is clearly linked to mitigation of the insider threat.
>
> (Participant #3)

> The content presented clearly communicates the proposed model concept. It's evident that a lot of research and thought contributed towards the model.
>
> (Participant #8)

> Yes. The model is very easily demonstrate [sic] in a real world scenario. (Participant #10)

Some participants highlighted the limitations of the model:

> It will need to be refined but the concept is valid.
>
> (Participant #2)

> Certainly, if you believe the world view that people are basicly [sic] good and that only a small subset are criminals. Refer Bruce Schneier's two last books where he expounds this sentiment.
>
> (Participant #11)

> It is, but it seems a farfetched idea.
>
> (Participant #12)

> Partly yes but as a user I will always feel like I'm being watched as a (insider) threat.
>
> (Participant #24)

Some participants explained why the model concept is justified and highlighted its relevance.

> Yes the organisation has to be able to protect your personal information.
>
> (Participant #16)

> This might be applicable in some situations. Especially with users that are more vulnerable to social engineering, such as telephone operators or receptionists. If there [sic] are inadvertently committing a crime, warning messages may work as indented [sic].
>
> (Participant #20)

#### 7.2.4 Benefit

In terms of benefit (i.e. does the model concept yield any benefit for information security, either immediately or in the future?), all of the participants responded positively.

Some participants indicated that the model concept does require refinement; however some participants (32%) indicated that they could see future potential:

> It yields? [sic] both immediately and in future. Users a [sic] likely to stop when they see an active warning while in the act of violating policy. The users may spread the word to which helps with future compliance.
>
> (Participant #10)

> Certainly in the future. It will be a hard sell in the current as most enteprises [sic] look to implement baseline controls through configuration.
>
> (Participant #12)

A smaller number (12%) indicated that they could see an immediate benefit:

> Immediately as it makes users aware of their actions and the consequences thereof.
>
> (Participant #5)

> There is immediate benefit and I am convinced that further refinements will improve the product.
>
> (Participant #6)

> It would be an immediate contributor in terms of the information security life cycle.
>
> (Participant #13)

> It holds immediate benefit as mechanisms exist to monitor the activities of users and the automation of responses to users after activating of a "honey token" is possible. Data loss protection is maturing in the industry

> and the mechanism could be included to further enhance information security efforts.
>
> (Participant #18)

The majority (58%) of participants generally accepted the statement:

> Definately [sic] as the study of why humans default is at the core of information security or any security for that matter.
>
> (Participant #11)

> To a degree yes. Especially from a compliance and security awareness perspective.
>
> (Participant #22)

> The model yields good benefits when it is combined with other defence mechanisms in information security. Remember it takes training, cultures, systems implementations technique and the ability of the IT department team and the will of the executive management.
>
> (Participant #23)

Participant #24 commented on the limitations of the honeytoken model:

> Yes, but not sure I agree with the honeypot concept, if user does not require access to specific info then have systems in place to restrict access, human factor is that people are always curious so they [sic] bound to access files/ folders they should not and that might trigger false-positives of users targeted/labelled as suspect data leakers.

#### 7.2.5 Recommendations

Some of the recommendations made by participants were based on when neutralisation mitigation should be invoked. Two participants indicated that it would better to determine the neutralisation technique before the access, while one suggested using artificial intelligence to determine whether the access should be allowed or denied, or escalated to management. Two participants indicated that neutralisation mitigation should be time-based or random rather than access-based. However, while the prototype is specific, the model concept does allow for this interpretation as well. Some participants commented on the order and timing of the events:

> I think the deterrent mechanism (i.e. the questions asking the user why he needed to access the file) and reference to policy should be before the file is accessed not after the fact. We should try the cognative [sic] dissonance method before the user is allowed to view the file.
>
> (Participant #4)

> I would ask you to consider less regular neutralisation mitigation, say time based and not accessed based.
>
> (Participant #6)

> Streamlining the process. Maybe change the tactics to have a user state his need for the data upfront, rather than after the fact. Build in some intelligence to either allow or deny based on the answer, or escalate to amangement [sic] etc.
>
> (Participant #17)

> I think if the message is displayed at random, rather when a specific record is accessed. How is it guaranteed that access of a honey token is malicious?
>
> (Participant #20)

> If possible the system should have artificial intelligence added to it, so possibly track what users should and should not be doing and if they stray then throw in neutralisation mitigation strategy or honey pot.
>
> (Participant #25)

Some participants recommended the integration of sanctions, auditing, monitoring, behavioural analysis, trend analysis and detective and corrective measures into the model concept. Others commented that the model could be improved by including accountability and audit:

> I would personally like to see it used to contribute towards policy compliance in a manner where the individual can be kept accountable for policy principles during daily operations without having to enter into separate policy awareness sessions for specific business rules [. . . ]
>
> (Participant #8)

> Perhaps I have a traditional view, but IMHO [in my humble opinion] the model concept could be greatly improved if it introduced severe consequences for defaulters and measured the response.
>
> (Participant #11)

> Insure [sic] that the user can be monitored and a full audit report can be kept for either prosecution, evidence.
>
> (Participant #16)

> Enforcement of policy. Monitoring of actions and trend analysis that can be used in behavioral [sic] analysis.

(Participant #22) Other participants commented on balancing detection, usability and social aspects with the model concept:

> The focus of this method is prevention, rather than detective or corrective. Whilst proven to be less much [sic] cheaper to implement preventative measures, it should always be considered in conjuction [sic] with detective and corrective measures.
>
> (Participant #15)

> With security there is always tradeoffs [sic], security vs usability. As long as the security is not tiresome when using the system the user will continue using the system.
>
> (Participant #19)

> Fairly agree and support the concept but I would recommend human factor is also taken into consideration as it might give an impression your users are just suspicious "human resources" and the approch [sic] might need a bit of perception vs reality to be proper interms [sic] of "people management".
>
> (Participant #24)

Finally, Participant #3 remarked on the limitations of the model concept's applicability:

> Consider this is a standalone implementation or consider it a very high risk, high value application (for example, protecting trade secrets).

As this research is based on the design science method, the recommendations will feed back into the next iteration of the model concept.

### 7.3 Limitations

In the study reported on in this article, purposive sampling was used, which could have biased the results. However, the fact that the participants were guaranteed anonymity might have offset possible bias. The quality of the results also depends on the capability of the panel; in this case, the panel had an average of 7.04 years of experience. A small sample size allowed for a more in-depth analysis of each participant's value judgements on the prototype and model concept.

## 8 DISCUSSION OF FINDINGS

With respect to the three main parts of the model concept—honeypots, neutralisation mitigation and cognitive dissonance—cognitive dissonance was the best received, while the benefits of honeypots and neutralisation mitigation were rated above average. However, honeypots were most severely criticised.

Honeypots are partially subject to 'false positives'. However if an insider directly manipulates the honeytoken by accessing, editing and saving it then it is certainly probable that this insider's intent was malicious, as such a command would not have been issued on fake data. Furthermore, an innocent user is likely to manipulate a honeytoken by accident less often than would a malicious user. The contextual factors that surround the attack of the honeytoken need to be considered before red-flagging an insider. In any event the model concept is not intended to vilify, but rather to remove excuses (i.e. neutralisation techniques).

The evaluation showed honeypots to be useless to the insider who was seeking a specific type of information, which is why the honeytoken generator needs to produce believable honeytokens. The insider will be attracted to the crime if they find the honeytoken to be believable and enticing, and to resemble what they were searching for. The possible legal risk associated with honeypots could be circumvented by having the insider sign a declaration regarding monitoring.

While neutralisation mitigation was criticised for being too narrowly focused, cognitive dissonance was seen as a positive step in compliance. In terms of neutralisation mitigation, it was felt that it would work if the user was generally moral and ethical. This supports the theory that cognitive dissonance will only work for individuals who are not psychopathic. Thus, for its efficacy to be enhanced, neutralisation mitigation clearly needs to be coupled with information security ethics. The model already compensates for this, as the 'alert conscience' technique encompasses a code of ethics. Cognitive dissonance was viewed as an improvement in information security and an important step in putting the human aspect of the insider threat into focus.

Some participants felt that induced dissonance would only work for certain types of individual. It is clear that in order for cognitive dissonance to work, several parameters need to be taken into account. Firstly, if an insider has psychopathic traits induced dissonance will not be as effective, hence it is suggested that all employees should go through a pre-screening first. It is clear if an insider ignores these cognitive dissonance interventions, then that individual is a definite threat and should be monitored more closely. Thus induced dissonance can be a means of identifying high-threat individuals. Secondly, cognitive dissonance will be ineffective in individuals with a low self-concept. Hence it is suggested that strategies that may include tactics to bolster the self-concept may be enforced prior to the induced dissonance intervention. Thirdly, the timing of the induced dissonance is crucial as it may not work on individuals who are already committing maleficence, while it may work on an insider who is misguided, for example an individual who assumed it was acceptable to break policies to meet a deadline. Hence relying on cognitive dissonance is an entirely preventative tactic and must be used in conjunction with other techniques to truly isolate the insider threat.

Some participants presumed that the model might benefit from the inclusion of penalties [84]. Barlow et al. [8] conclude that both deterrent sanctions and neutralisation should be given due attention during training, while Siponen and Vance [11] found neutralisation to be more powerful than deterrents. In terms of the model concept, this is an interesting perspective. For example, Aronson and Carlsmith [85] found milder forms of deterrent to be more effective in changing attitudes than severe deterrents. If an individual is threatened with a mild punishment if they do not comply, then this causes cognitive dissonance, particularly if the person has a strong desire not to comply. In this case, the individual will have to rely on internal justification to reduce this dissonance in order to comply, whereas the threat of severe punishment is sufficient to force someone to comply without an internal justification. This implies that once the threat is removed,

those individuals who rely on internal justification are more likely to remain compliant. This observation was also confirmed by a recent study in which it was found that a low-level threat may be more effective in changing attitudes about online gaming [86]. While cognitive dissonance is viewed as way to incorporate cultural change, it is more effective if it is supported by positive reinforcement.

In general the challenges of integrating extant systems with the model concept may be overcome through aspect-oriented programming, since aspect orientation allows developers to view neutralisation as a separable concern without compromising the extant code. This also solves the other potential problem cited by the participants, namely that over time the process will erode, and insiders will eventually ignore the process. The same technique to lure and mitigate insiders cannot be used continually and so the concept will have to be upgraded regularly with variation. If the aspect orientation is used, these variations maybe swapped in and out with minimal perturbation to the system.

It can be argued that model concept deployment will affect productivity. However the mechanism could be switched off during peak periods. Another possibility is to apply restrictions to model deployment, which involve maintaining a set of properties when the operation should not react. While some participants claimed that the model was just another form of training, others were acutely aware that the model is actually a form of dynamic training based on self-persuasion rather than static direct-persuasion training. Some participants did highlight the fact that the typical training does not work. However, the model was not intended merely to train individuals but to provide neutralisation mitigation-type training. While the prototype of the model concept did indicate in the post-neutralisation phase links to policy, the intention was to present policy in a way that promoted policy and assisted compliance. As these aspects are context-dependent, the prototype did not provide any indications of how this may be achieved. (This aspect is implied in the presentation of the prototype in Appendix B). This may have caused the participants to assume this is a weakness of the model concept, as the focus of this first cycle of the prototype was creating the intervention points for neutralisation mitigation. This is an area that does require further research in terms of how the policy should be presented at this stage. It has been suggested that it could be done by a commitment-type exercise or perhaps a questionnaire. Most participants agreed that the model is an enhancement to rather than a replacement of existing technologies.

The central theoretical proposition of this research is that inducing cognitive dissonance and simultaneously countering the resultant neutralisations (i.e. justifications) by misguided or possibly malevolent insiders via neutralisation mitigation will be linked to an increased mindfulness of the limitations of their neutralisation techniques.

The second proposition, which was proved by previous empirical research, is that this awareness will trigger a behavioural shift towards compliance. Barlow et al.'s [8] study confirmed that 'training focused around fighting neutralisation should be powerful in reducing intentions to violate policies'. They furthermore found that both neutralisation mitigation and deterrence-based statements lowered the participants' intention to violate policy. This finding was also reported by the participants in the current study—they felt there was a need for additional deterrence mechanisms. Siponen and Vance [11] found that 'neutralisation is an excellent predictor of employees' intention to violate IS security policies'. They suggested that training sessions and scenario-based exercises be established and that security policies be advertised prominently to circumvent neutralisations. However, they were not certain which methods would work best.

The research presented in this paper attempts to bridge this gap by simulating a scenario where dissonance is induced and neutralisation mitigation is used to challenge the resultant neutralisations used to commit the simulated offense. As the ITNM$^{CD}$ model was shown to be acceptable to the participants, and since inducing dissonance was suggested as a useful approach, they were neither entirely convinced by the method used to induce dissonance (i.e. honeytokens), nor by the methods to mitigate the neutralisations (i.e. techniques of Situational Crime Prevention). Hence, although the central proposition of this research was shown to have some merit, it does require further refinement—which is the objective of design science research. Perhaps, in the next iteration, the alternative neutralisation mitigation techniques proposed by Wortley [72] (i.e. 'rule setting', 'clarifying responsibility', 'clarifying consequences', 'increasing victim worth') will be incorporated. Both honeytokens and honeynets were proposed by Spitzner [34] as a means to entrap the insider threat. For the proposed model, honeytokens were favoured on account of their ease of implementation; however, honeynets may involve a more sophisticated means of entrapment. This provides two new directions for the ITNM$^{CD}$ model, firstly to revise the neutralisation mitigation techniques and secondly to re-evaluate the type of honeypot.

## 9 IMPLICATIONS FOR PRACTICE

The model concept requires highly skilled resources to implement and maintain it. Moreover, the success of the model is contingent on an environment where there is a clear assignment of duties and pertinent guidance on data classification. It is clear that the model will not be viable in an environment where neither properly defined information security policies nor continuous improvement practices are in place. Each 'breach' requires the information security policy to be reviewed on a continuous basis. The value system of the organisation will also need to be clearly defined. For example, the ITNM$^{CD}$ model will be unproductive in an organisation that does not have a clearly defined ethical code in place. Essentially an organisation

will have to be at a high level of information security maturity in order for this model to be effective. It is clear from the respondents that the model would probably be effective in a subset of cases, and hence it solves a specific subset of the insider threat problem. It would be highly effective for misguided insiders and insiders who are contemplating crime and it is proposed that this model should be one component of a comprehensive prevention strategy.

## 10 CONCLUSION

This article presented the ITNM$^{CD}$ model, which is intended to mitigate the rationalisations that insiders employ to commit maleficence. The research conducted contributes to the literature on the problem of the insider threat. First, this article considered the concept of neutralisations from three perspectives: neutralisation techniques, neutralisation mitigation and neutralisation drivers (i.e. cognitive dissonance). Second, a three-tier model based on the sociological, technical and socio-technical dimensions to resolve neutralisation was proposed. Third, a possible interpretation of the model concept was presented. Hence, the main contribution made by this article is the multidimensionality of the model concept, which provides a new solution space in which to reason about mitigating insider threat neutralisations. The ultimate goal of the model is to provide organisations with a better understanding of the concept of neutralisation and to act as an impetus for a customised neutralisation mitigation strategy. The model was prototyped and evaluated, and the model concept was generally well received.

The evaluation does give rise to new research questions. For example, are honeypots the best mechanisms to induce cognitive dissonance? Should neutralisation mitigation be coupled with sanctions, since the literature on this aspect is inconclusive? If so, to what degree should sanctions be applied? Does the success of neutralisation mitigation depend on the morality of the insider? These questions form the basis for future research in this area. The model concept may empower administrators to prevent, detect, contain and possibly counteract the insider threat. Furthermore, the model concept derived may assist in reassessing the granularity of access that a malicious insider could possibly be entrusted with. For example, an individual who is found to be tempted by the luring honeytoken should not be given access to highly classified information in the future.

## REFERENCES


[1] R. Willison and M. Siponen. "Overcoming the insider: Reducing employee computer crime through situational crime prevention". *Communications of the ACM*, vol. 52, pp. 133 – 137, September 2009. DOI http://dx.doi.org/10.1145/1562164.1562198.

[2] CSO Magazine, U.S. Secret Service, Software Engineering Institute CERT Program and PricewaterhouseCooper. "2014 US state of cybercrime survey", June 2014. URL http://www.pwc.com/en_US/us/increasing-it-effectiveness/publications/assets/2014-us-state-of-cybercrime.pdf.

[3] F. Farahmand and E. H. Spafford. "Understanding insiders: An analysis of risk-taking behavior". *Information systems frontiers*, vol. 15, pp. 5 – 15, March 2013. DOI http://dx.doi.org/10.1007/s10796-010-9265-x.

[4] J. T. Wells. *Principles of fraud examination*. Wiley, Hoboken, NJ, 2008.

[5] N. L. Beebe and V. S. Roa. "Improving organizational information security strategy via meso-level application of situational crime prevention to the risk management process". *Communications of the Association for Information Systems*, vol. 26, pp. 329 – 358, March 2010.

[6] D. Rothe. *State criminality: The crime of all crimes*. Rowman & Littlefield, Lanham, Maryland, 2009.

[7] L. Festinger. *A theory of cognitive dissonance*. Stanford University Press, Standford, California, 1962.

[8] J. B. Barlow, M. Warkentin, D. Ormond and A. R. Dennis. "Don't make excuses! Discouraging neutralization to reduce IT policy violation". *Computers & security*, vol. 39, pp. 145 – 159, November 2013. DOI http://dx.doi.org/10.1016/j.cose.2013.05.006.

[9] R. V. Clarke. "Introduction". In R. V. Clarke (editor), *Situational crime prevention: Successful case studies*, pp. 1 – 43. Harrow and Heston, Guilderland, NY, 1997.

[10] G. Sykes and D. Matza. "Techniques of neutralization: A theory of delinquency". *American sociological review*, vol. 22, pp. 664 – 670, December 1957. DOI http://dx.doi.org/10.2307/2089195.

[11] M. Siponen and M. Vance. "Neutralization: New insights into the problem of employee systems security policy violations". *MIS quarterly*, vol. 34, pp. 487 – 502, September 2010.

[12] J. Cooper. *Cognitive dissonance: 50 years of a classic theory*. SAGE, London, 2007.

[13] I. Redondo and J.-P. Charron. "The payment dilemma in movie and music downloads: An explanation through cognitive dissonance theory". *Computers in human behavior*, vol. 29, pp. 2037 – 2046, September 2013. DOI http://dx.doi.org/10.1016/j.chb.2013.04.015.

[14] L. Coles-Kemp and M. Theoharidou. "Insider threat and information security management". In C. W. Probst, J. Hunker, D. Gollmann and M. Bishop (editors), *Insider threats in cyber security*, pp. 45 – 71. Springer, US, 2010.

[15] R. Willison. "Understanding the perpetration of employee computer crime in the organisational context". *Information and organization*, vol. 16, pp. 304 – 324, January 2006. DOI http://dx.doi.org/10.1016/j.infoandorg.2006.08.001.

[16] R. Walton. "Balancing the insider and outsider threat". *Computer fraud & security*, vol. 11, pp. 8 –11, November 2006. DOI http://dx.doi.org/10.1016/S1361-3723(06)70440-7.

[17] M. Cappelli, A. P. Moore, T. J. Shimeall and R. Trzeciak. "Common sense guide to prevention/detection of insider threats", July 2006. URL https://www.cylab.cmu.edu/files/pdfs/CERT/CommonSenseInsiderThreatsV2.1-1-070118-1.pdf.

[18] S. L. Pfleeger, J. B. Predd, J. Hunker and C. Bulford. "Insiders behaving badly: Addressing bad actors and their actions". *IEEE transactions on information forensics and security*, vol. 5, pp. 169 –179, December 2010. DOI http://dx.doi.org/10.1109/TIFS.2009.2039591.

[19] M. Bishop and C. Gates. "Defining the insider threat". In *4th annual workshop on cyber security and information intelligence research: Developing strategies to meet the cyber security and information intelligence challenges ahead*. 2008. DOI http://dx.doi.org/10.1145/1413140.1413158.

[20] M. B. Salem and S. J. Stolfo. "Combining baiting and user search profiling techniques for masquerade detection". *Journal of wireless mobile networks, ubiquitous computing, and dependable applications (JoWUA)*, vol. 3, pp. 13 – 29, March 2012.

[21] M. B. Salem, S. Hershkop and S. J. Stolfo. "A survey of insider attack detection research: Beyond the hacker". In S. J. Stolfo, S. M. Bellovin, A. D. Keromytis, S. Hershkop, W. S. Smith and S. Sinclair (editors), *Insider attack and cyber security*, Advances in Information Security, pp. 69 – 90. Springer US, New York, 2008.

[22] G. B. Magklaras and S. M. Furnell. "Insider threat prediction tool: Evaluating the probability of IT misuse". *Computers & security*, vol. 21, pp. 62 – 73, February 2001. DOI http://dx.doi.org/10.1016/S0167-4048(02)00109-8.

[23] C. P. Pfleeger. "Reflections on the insider threat". In S. J. Stolfo, S. M. Bellovin, A. D. Keromytis, S. Hershkop, S. W. Smith and S. Sinclair (editors), *Insider attack and cyber security*, pp. 5 –16. Springer, US, 2008.

[24] B. J. Wood. "An insider threat model for adversary simulation". In *Research on mitigating the insider threat to information systems*, vol. 2, pp. 1 – 3. SRI International, 2000.

[25] B. M. Bowen, M. B. Salem, S. Hershkop, A. D. Keromytis and S. J. Stolfo. "Designing host and network sensors to mitigate the insider threat". *IEEE security & privacy*, vol. 7, pp. 22 – 29, November-December 2009. DOI http://dx.doi.org/10.1109/MSP.2009.109.

[26] D. F. Anderson, D. Cappelli, J. J. Gonzalez, M. Mojtahedzadeh, A. Moore, R. Elliot and J. M. Sarriegu. "Preliminary system dynamics maps of the insider cyber-threat problem". In *22nd International conference of the system dynamics society*. 2004.

[27] S. Zeadally, Y. Byunggu, H. J. Dong and L. Liang. "Detecting insider threats: Solutions and trends". *Information security journal: A global perspective*, vol. 21, pp. 183 – 192, June 2012. DOI http://dx.doi.org/10.1080/19393555.2011.654318.

[28] G. J. Silowash and C. King. "Insider threat control: Understanding data loss prevention (DLP) and detection by correlating events from multiple sources", January 2013. URL http://repository.cmu.edu/sei/708.

[29] S. Liu and R. Kuhn. "Data loss prevention". *IT professional*, vol. 12, pp. 10 – 13, March-April 2010. DOI http://doi.ieeecomputersociety.org/10.1109/MITP.2010.52.

[30] M. D. Guido and M. W. Brooks. "Insider threat program best practices". In *2013 46th Hawaii international conference on system sciences (HICSS)*. 2013. DOI http://dx.doi.org/10.1109/HICSS.2013.279.

[31] A. Joch. "Why you can't stop insider threats: You can only hope to contain them", February 2011. URL http://fcw.com/articles/2011/02/28/feat-cybersecurity-insider-threats.aspx.

[32] E. Cole and S. Ring. *Insider threat: Protecting the enterprise from sabotage, spying, and theft*. Syngress, Rockland, Maine, 2006.

[33] E. E. Schultz. "A framework for understanding and predicting insider attacks". *Computers & security*, vol. 21, pp. 526 – 531, October 2002. DOI http://dx.doi.org/10.1016/S0167-4048(02)01009-X.

[34] L. Spitzner. "Honeypots: Catching the insider threat". In *19th Annual computer security applications conference (ACSAC 2003)*. 2003. DOI http://dx.doi.org/10.1109/CSAC.2003.1254322.

[35] R. McGrew, B. Rayford and J. R. Vaughn. "Experiences with honeypot systems: Development, deployment, and analysis". In *2006 39th Hawaii international conference on system sciences (HICSS06)*. 2006. DOI http://dx.doi.org/10.1109/HICSS.2006.172.

[36] L. Spitzner. "Honeypots: Are they illegal?", June 2010. URL http://www.symantec.com/connect/articles/honeypots-are-they-illegal.

[37] M. E. Kabay. "Honeypots, part 4: Liability and ethics of honeypots", May 2003. URL http://www.networkworld.com/newsletters/2003/0519sec2.html.

[38] K. Padayachee. "A conceptual opportunity-based framework to mitigate the insider threat". In *Information security for South Africa*. 2013. DOI http://dx.doi.org/10.1109/ISSA.2013.6641060.

[39] N. Nykodym, R. Taylor and J. Vilela. "Criminal profiling and insider cyber crime". *Digital investigation*, vol. 2, pp. 261 – 267, December 2005. DOI http://dx.doi.org/10.1016/j.diin.2005.11.004.

[40] J. Hunker and C. W. Probst. "Insiders and insider threats-an overview of definitions and mitigation techniques". *Journal of wireless mobile networks, ubiquitous computing, and dependable applications*, vol. 2, pp. 4 – 27, March 2011.

[41] C. Colwill. "Human factors in information security: The insider threat–Who can you trust these days?" *Information security technical report*, vol. 14, pp. 186 – 196, November 2009. DOI http://dx.doi.org/10.1016/j.istr.2010.04.004.

[42] Ponemon Institute. "Privileged user abuse & the insider threat", May 2014. URL http://www.trustedcs.com/resources/whitepapers/Ponemon-RaytheonPrivilegedUserAbuseResearchReport.pdf.

[43] D. D'Arcy and A. Hovav. "Does one size fit all? Examining the differential effects of IS security countermeasures". *Journal of business ethics*, vol. 89, pp. 57 – 71, May 2009. DOI http://dx.doi.org/10.1007/s10551-008-9909-7.

[44] E. Schultz. "Security training and awareness–Fitting a square peg in a round hole". *Computers & security*, vol. 23, pp. 1 – 2, February 2004. DOI http://dx.doi.org/10.1016/j.cose.2004.01.002.

[45] M. Kajzer, J. D'Arcy, C. R. Crowell, A. Striegel and D. V. Bruggen. "An exploratory investigation of message-person congruence in information security awareness campaigns". *Computers & security*, vol. 43, pp. 64 – 76, June 2014. DOI http://dx.doi.org/10.1016/j.cose.2014.03.003.

[46] O. Brdiczka, J. Liu, B. Price, J. Shen, A. Patil, R. Chow, E. Bart and N. Ducheneaut. "Proactive insider threat detection through graph learning and psychological context". In *IEEE symposium on security and privacy workshops (SPW)*. 2012. DOI http://dx.doi.org/10.1109/SPW.2012.29.

[47] T. L. Powers and E. P. Jack. "The influence of cognitive dissonance on retail product returns". *Psychology & marketing*, vol. 30, pp. 1520 – 6793, July 2013. DOI http://dx.doi.org/10.1002/mar.20640.

[48] L. Law, S. H. Ting and C. Jerome. "Cognitive dissonance in dealing with plagiarism in academic writing". *Procedia–Social and behavioral sciences*, vol. 97, pp. 278 – 284, November 2013. DOI http://dx.doi.org/10.1016/j.sbspro.2013.10.234.

[49] Z. I. Latheef and S. Werner. "Organizational dissonance: Development of a new construct". In *Academy of management proceedings*. 2013. DOI http://dx.doi.org/10.5465/AMBPP.2013.10448abstract.

[50] S. W. Chun. "Change that attitude: The ABCs of a persuasive security awareness program". In H. F. Tipton and M. Krause (editors), *Information security management handbook*, pp. 521 – 530. CRC Press, Boca Raton, 2007.

[51] M. Workman. "Wisecrackers: A theory-grounded investigation of phishing and pretext social engineering threats to information security". *Journal of the American Society for Information Science and Technology*, vol. 59, pp. 662 – 674, December 2008. DOI http://dx.doi.org/10.1002/asi.20779.

[52] S. P. Lawrence and D. D. Caputo. "Leveraging behavioral science to mitigate cyber security risk". *Computers & security*, vol. 31, pp. 597 – 611, June 2012. DOI http://dx.doi.org/10.1016/j.cose.2011.12.010.

[53] M. A. Freeman, E. V. Hennessy and D. M. Marzullo. "Defensive evaluation of antismoking messages among college-age smokers: The role of possible selves". *Health psychology*, vol. 20, pp. 424 – 433, November 2001. DOI http://dx.doi.org/10.1037/0278-6133.20.6.424.

[54] E. Aronson, C. Fried and J. Stone. "Overcoming denial and increasing the intention to use condoms through the induction of hypocrisy". *American journal of public health*, vol. 81, pp. 1636 – 1638, December 1991. DOI http://dx.doi.org/10.2105/AJPH.81.12.1636.

[55] V. Fointiat. "Saying, but not doing: Induced hypocrisy, trivialization, and misattribution". *Social behavior and personality: An international journal*, vol. 39, pp. 465 – 475, May 2011. DOI http://dx.doi.org/10.2224/sbp.2011.39.4.465.

[56] B. A. Morrongiello and M. Landa. "'Practice what you preach': Induced hypocrisy as an intervention strategy to reduce children's intentions to risk take on playgrounds". *Journal of pediatric psychology*, vol. 33, pp. 1117 – 1128, February 2008. DOI http://dx.doi.org/10.1093/jpepsy/jsn011.

[57] E. Aronson. "Dissonance, hypocrisy, and the self-concept". In E. Harmon-Jones and J. Mills (editors), *Cognitive dissonance: Progress on a pivotal theory in social psychology*, pp. 103 – 126. American Psychological Association, Washington, DC, 1999.

[58] C. A. Dickerson, R. Thibodeau, E. Aronson and D. Miller. "Using cognitive dissonance to encourage water conservation". *Journal of applied social psychology*, vol. 22, pp. 841 – 854, June 1992. DOI http://dx.doi.org/10.1111/j.1559-1816.1992.tb00928.x.

[59] A. J. McClurg. "Good cop, bad cop: Using cognitive dissonance theory to reduce police lying". *UC Davis law review*, vol. 32, p. 389453, Winter 1999.

[60] A. A. Murray, J. M. Wood and S. O. Lilienfeld. "Psychopathic personality traits and cognitive dissonance: Individual differences in attitude change". *Journal of research in personality*, vol. 46, pp. 525 – 536, October 2012. DOI http://dx.doi.org/10.1016/j.jrp.2012.05.011.

[61] L. Festinger and J. M. Carlsmith. "Cognitive consequences of forced compliance". *The journal of abnormal and social psychology*, vol. 58, pp. 203 – 210, March 1959. DOI http://dx.doi.org/10.1037/h0041593.

[62] T. Freijy and E. J. Kothe. "Dissonance-based interventions for health behaviour change: A systematic review". *British journal of health psychology*, vol. 18, pp. 310 – 337, September 2013. DOI http://dx.doi.org/10.1111/bjhp.12035.

[63] E. Aronson and D. R. Mettee. "Dishonest behavior as a function of differential levels of induced self-esteem". *Journal of personality and social psychology*, vol. 9, pp. 121 – 127, June 1968. DOI http://dx.doi.org/10.1037/h0025853.

[64] J. Gilbert. "ID governance: Bridging compliance and security", May 2014. URL http://old.wallstreetandtech.com/it-infrastructure/id-governance-bridging-compliance-and-se/229625468.

[65] W. W. Minor. "Techniques of neutralization: A reconceptualization and empirical examination". *Journal of research in crime and delinquency*, vol. 18, pp. 295 – 318, July 1981. DOI http://dx.doi.org/10.1177/002242788101800206.

[66] C. Klockars. *The professional fence*. Free Press, New York, 1974.

[67] R. Willison and M. Warkentin. "Beyond deterrence: An expanded view of employee computer abuse". *MIS quarterly*, vol. 37, pp. 1 – 20, March 2013.

[68] S. Curley and S. Zamoon. "IT influences on moral intensity in ethical decision-making", March 2009. URL http://misrc.umn.edu/workingpapers/fullpapers/2009/ZamoonCurley2009-03.pdf.

[69] M. G. Piacentini, A. Chatzidakis and E. N. Banister. "Making sense of drinking: the role of techniques of neutralisation and counter-neutralisation in negotiating alcohol consumption". *Sociology of health & illness*, vol. 34, pp. 841 – 857, July 2012. DOI http://dx.doi.org/10.1111/j.1467-9566.2011.01432.x.

[70] D. B. Cornish and R. V. Clarke. "Opportunities, precipitators and criminal decisions: A reply to Wortley's critique of situational crime prevention". In M. J. Smith and D. B. Cornish (editors), *Theory for practice in situational crime prevention (Crime prevention studies, vol. 16)*, pp. 41 – 96. Criminal Justice Press, New York, 2003.

[71] N. L. Beebe and V. S. Roa. "Using situational crime prevention theory to explain the effectiveness of information systems security". In *2005 SoftWars conference*. 2005.

[72] R. K. Wortley. "Guilt, shame and situational crime prevention". In R. Homel (editor), *The politics and practice of situational crime prevention*, pp. 115 – 132. Criminal Justice Press, New York, 1996.

[73] R. K. Wortley, E. McDonagh and R. Homel. "Perceptions of physical, psychological, social and legal deterrents to joyriding". *Crime prevention and community safety: An international journal*, vol. 4, pp. 7 – 25, January 2002. DOI http://dx.doi.org/10.1057/palgrave.cpcs.8140111.

[74] R. Agnew and A. A. Peters. "The techniques of neutralization: An analysis of predisposing and situational factors". *Criminal justice and behavior*, vol. 13, pp. 81 – 97, March 1986. DOI http://dx.doi.org/10.1177/0093854886013001005.

[75] S. Hinduja and B. Kooi. "Curtailing cyber and information security vulnerabilities through situational crime prevention". *Security journal*, vol. 26, pp. 383 – 402, June 2013. DOI http://dx.doi.org/10.1057/sj.2013.25.

[76] L. Spitzner. "Honeytokens: The other honeypot", July 2003. URL http://www.securityfocus.com/infocus/1713.

[77] E. Turban and J. E. Aronson. *Decision support systems and intelligent systems*. Prentice Hall, USA, 1998.

[78] D. Andrews and J. Bonta. *The psychology of criminal conduct*. Anderson Publishing Co, Cincinnati, 1998.

[79] J. Park and R. Sandhu. "The UCON ABC usage control model". *ACM transactions on information and system security (TISSEC)*, vol. 7, pp. 128 – 174, February 2004. DOI http://dx.doi.org/10.1145/984334.984339.

[80] K. Peffers, T. Tuunanen, C. E. Gengler, M. Rossi, W. Hui, V. Virtanen and J. Bragge. "The design science research process: A model for producing and presenting information systems research". In *Proceedings of the first international conference on design science research in information systems and technology (DESRIST 2006)*. 2006.

[81] M. T. March and G. F. Smith. "Design and natural science research on information technology". *Decision support systems*, vol. 15, pp. 251 – 266, December 1995. DOI http://dx.doi.org/10.1016/0167-9236(94)00041-2.

[82] P. Offerman, O. Levina, M. Schonherr and Bub. "Outline of a design science research process". In *Proceedings of the 4th international conference on design science research into information systems and technology (DESRIST)*. 2009. DOI http://dx.doi.org/10.1145/1555619.1555629.

[83] H. Österle, J. Becker, U. Frank, T. Hess, D. Karagiannis, H. Krcmar, P. Loos, P. Mertens, A. Oberweis and E. J. Sinz. "Memorandum on design-oriented information systems research". *European journal of information systems*, vol. 20, pp. 7 –10, January 2010. DOI http://dx.doi.org/10.1057/ejis.2010.55.

[84] A. McIlwraith. *Information security and employee behaviour: how to reduce risk through employee education, training and awareness*. Gower Publishing, Ltd., United Kingdom, 2006.

[85] E. Aronson and J. M. Carlsmith. "Effect of the severity of threat on the devaluation of forbidden behavior". *Journal of abnormal and social psychology*, vol. 66, pp. 584 – 588, June 1963. DOI http://dx.doi.org/10.1037/h0039901.

[86] C. S. Wan and W. B. Chiou. "Inducing attitude change toward online gaming among adolescent players based on dissonance theory: The role of threats and justification of effort". *Computers & education*, vol. 54, pp. 162 –168, January 2010. DOI http://dx.doi.org/10.1016/j.compedu.2009.07.016.

## APPENDIX A: VALUE JUDGEMENTS

Table 1: Results of the value judgements

| Sample items | Statements | Value judgements |
|---|---|---|
| **Viability** (feasibility) | **S1**: The model concept could easily be translated into an implementable product. | 80% |
| | **S2**: The model concept can be integrated into existing systems. | 92% |
| **Utility** (value) | **S3**: The model concept will deter the insider threat from abusing their privileges. | 60% |
| | **S4**: The protection mechanisms, such as neutralisation mitigation, will compel the insider threat to comply with the established rules of behaviour in order to protect confidential information in the future. | 64% |
| | **S5**: The luring honeypots will support an organisation in mitigating the insider attack risk by focusing the attackers attention on decoy assets, while the critical assets are protected. | 60% |
| | **S6**: Cognitive dissonance may be a useful technique in encouraging compliance. | 96% |
| **Efficacy** (effectiveness) | **S7**: In terms of the enforcement of security, other mechanisms such as a written policy document or adequate training would have been more effective than the mechanisms identified in the product concept. | 44%† |
| | **S8**: The benefits of a luring honeytoken may be outweighed if insiders waste time and resources interacting with them. | 68%† |
| | **S9**: Most insiders will NOT benefit from neutralisation mitigation, as they will ignore the process. | 40%† |
| | **S10**: Cognitive dissonance has NO impact on compliant security behaviour. | 16%† |
| **Usability** | **S11**: The technical mechanisms may be distracting to a user. | 76%† |
| **Scalability** | **S12**: This model will be scalable in a real-world context | 72% |
| | **S13**: There are NO conceivable environments in which this product concept will be applicable. | 4%† |

(Judgments marked with † reversed scored in analysis)

## APPENDIX B: THE PROTOTYPE

The demo application begins with typical authentication and authorisation (Fig. 7).

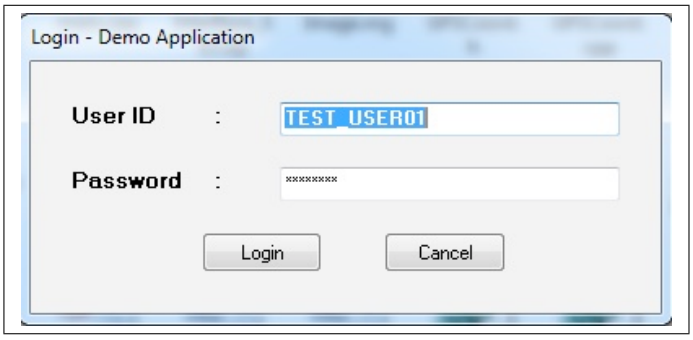


Figure 7

**User action**: User logs in.
The demo application is a typical database application with real data and honeytokens embedded (Fig. 8):

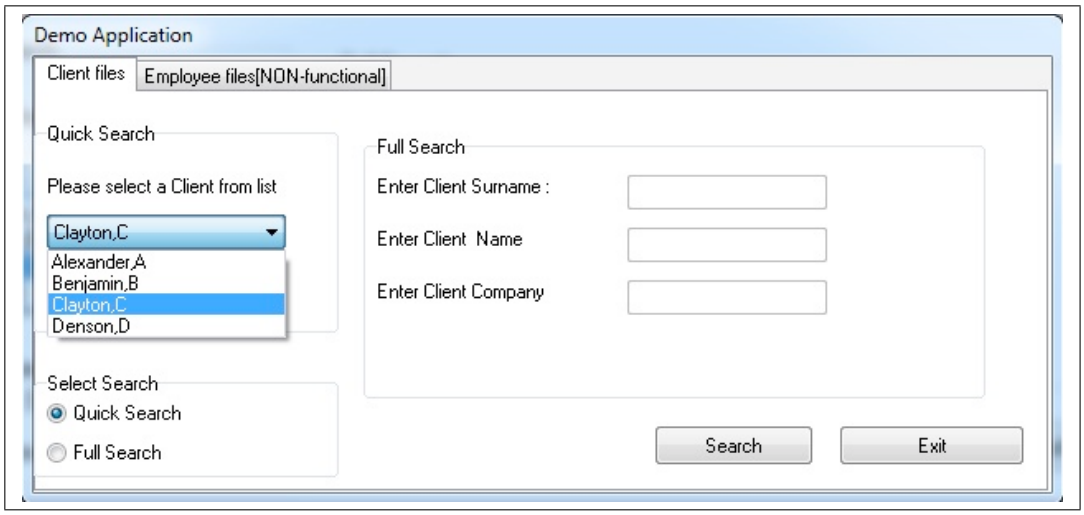


Figure 8

**User action**: Selects 'Clayton, C' (which is a honeytoken).
**Pre-Neutralisation**: Alert Conscience before access if a honeytoken is selected (e.g. , 'Clayton, C' in this case) (Fig. 9). **User action**: Selects 'Yes'.

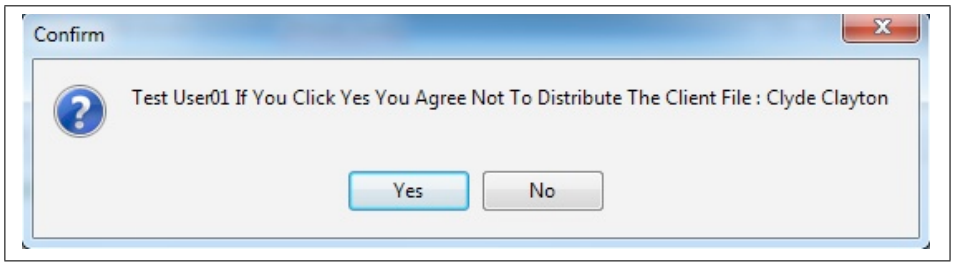


Figure 9

**Ongoing neutralisation**: Display data alongside banner ('Posting instructions' on use of data) (Fig. 10)

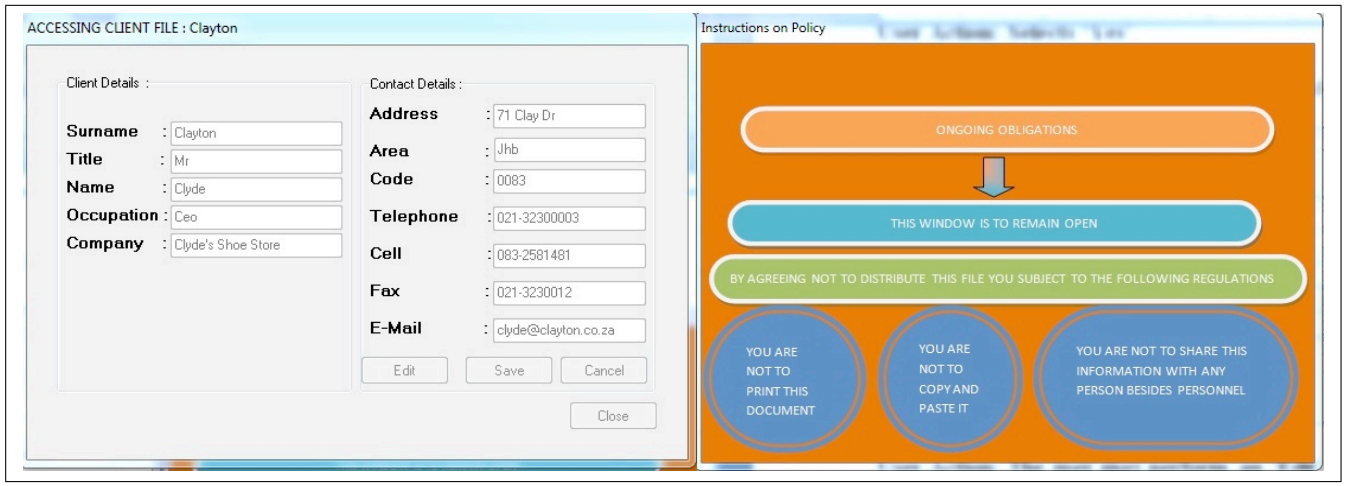


Figure 10

**User action**: The user may perform an 'Edit' or 'Save' on this data (this clearly shows that the user is performing actions on the data that was not instructed).

**Post-neutralisation**: Alert Conscience (this involves determining the nature of neutralisations used to justify access) (Fig. 11)

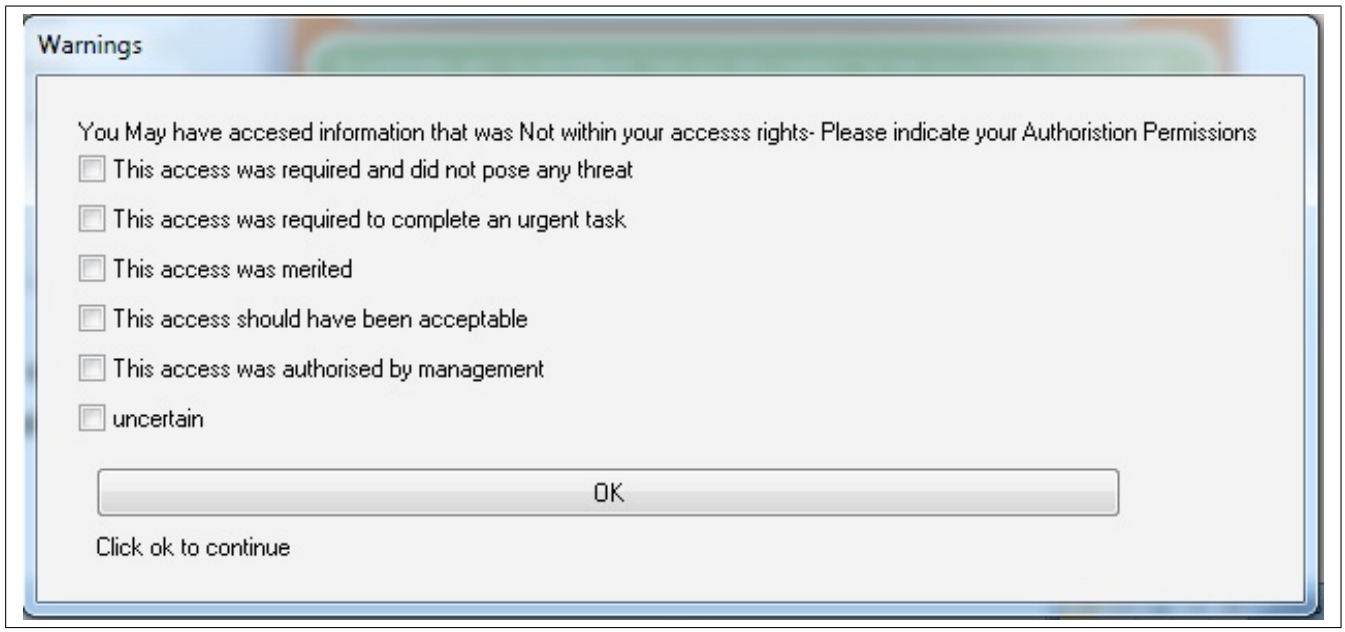


Figure 11

**User action**: Selects 'This action was required and did not pose any threat'.

**Post-neutralisation**: Assist Compliance and Promote Policy (Fig. 12)


Figure 12

**Post-neutralisation**: Assist Compliance and Promote Policy by demonstrating the hypocrisy of their neutralisations (i.e. rationalisations) (Fig. 13)

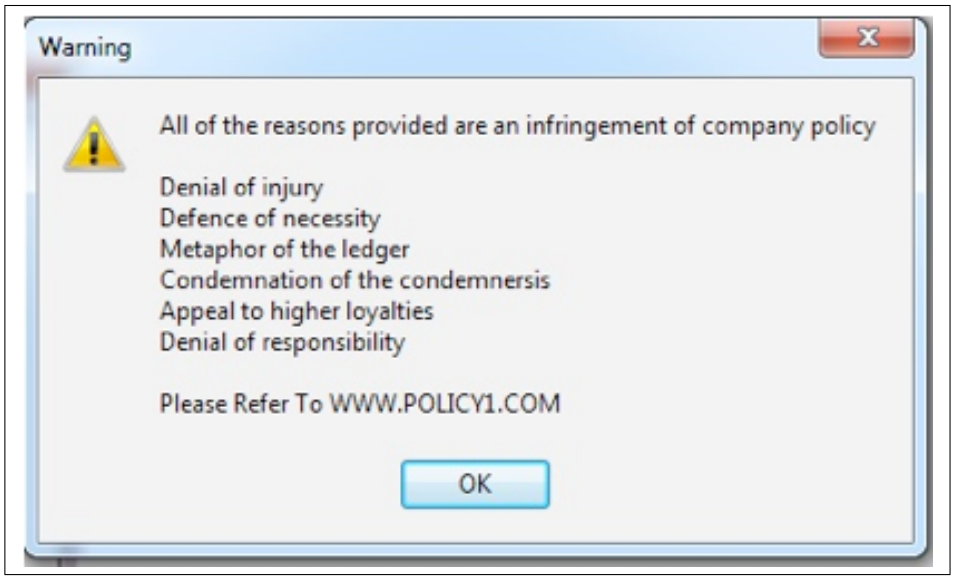


Figure 13

At this juncture, the user should be either shown an online demonstration or be given targeted training establishing the hypocrisy of their neutralizations (i.e. rationalizations).