Explainable Information Security: Development of a Construct and Instrument

Completed research paper

Duy Dang-Pham
School of Science and Technology
RMIT Vietnam
Ho Chi Minh City, Vietnam
Email: duy.dangphamthien@rmit.edu.vn

Ai-Phuong Hoang
School of Business and Management
RMIT Vietnam
Ho Chi Minh City, Vietnam
Email: phuong.hoangai@rmit.edu.vn

Diem-Trang Vo
School of Business and Management
RMIT Vietnam
Ho Chi Minh City, Vietnam
Email: trang.vothidiem1@rmit.edu.vn

Karlheinz Kautz
School of Accounting, Information Systems and Supply Chain
RMIT University
Melbourne, Australia
Email: karlheinz.kautz@rmit.edu.au

Abstract

Despite the increasing efforts to encourage information security (InfoSec) compliance, employees' refusal to follow and adopt InfoSec remains a challenge for organisations. Advancements in the behavioural InfoSec field have recently highlighted the importance of developing usable and employee-centric InfoSec that can motivate InfoSec compliance more effectively. In this research, we conceptualise the theoretical structure for a new concept called explainable InfoSec and develop a research instrument for collecting data about this concept. Data was then collected from 724 office workers via an online survey. Exploratory and confirmatory factor analyses were performed to validate the theoretical structure of the explainable InfoSec construct, and we performed structural equation modelling to examine the construct's impact on intention to comply with organisational InfoSec. The validated theoretical structure of explainable InfoSec consists of two dimensions, fairness and transparency, and the construct was found to positively influence compliance intention.

Keywords: explainable information security, fairness, accountability, transparency, information security compliance

1 Introduction

Employees are the most critical link in the organisational information security (InfoSec) chain (Sasse et al. 2001). Studies in the behavioural InfoSec field have examined the factors that influence different types of InfoSec behaviours which range from malicious to benevolent behaviours. For instance, organisational factors such as empowerment, security education training awareness (SETA) programs, and management's leadership have been found to motivate the employees' intention to comply (Dhillon et al. 2020). Social influences, such as subjective norms and InfoSec climate, are also positively associated with intention to comply (Bulgurcu et al. 2010; Chan et al. 2005). Moreover, rewards and sanctions have been consistently found to affect both desirable and malicious InfoSec behaviours (D'Arcy and Herath 2011).

Although several studies have identified effective ways to persuade employees to adopt organisational InfoSec, employees' resistance to following InfoSec directives remains a challenge for organisations (Boss et al. 2009). This puts emphasis on the employees' acceptance of InfoSec measures, where one of the goals of organisational InfoSec should be to empower the employees as principal agents who make their own decision to accept and adopt organisational InfoSec (Kirlappos and Sasse 2014; Viganò and Magazzeni 2018). We examine and validate the theoretical structure of a concept called explainable InfoSec (XSec). The term XSec was originally coined in research by Viganò and Magazzeni (2018), who presented a framework of questions for what to consider to achieve XSec, such as who provides/receives the explanations, what is explained, when the explanation should be given, and how to explain InfoSec. Our research complements Viganò and Magazzeni's study (2018) and researches on usable InfoSec approaches (Kirlappos and Sasse 2014), by proposing a theoretical structure for the XSec construct which consists of specific dimensions.

In our study, the concept of XSec measures the quality of organisational InfoSec that triggers employees' acceptance of InfoSec and their compliance intention. Our study is inspired by an emerging research stream that focuses on a related concept called explainable artificial intelligence (XAI), which puts forward that although intelligent systems may be viewed as a "black box", their internal mechanisms must be explainable to users in order to gain their trust in the systems' algorithmic decisions (Shin et al. 2020). Similarly, we apply the black box analogy to describe organisational InfoSec, since failing to explain and understand the importance of InfoSec can impact the employees' trust towards InfoSec measures and their compliance (Pieters 2011; Viganò and Magazzeni 2018). Our paper is structured as followed. We review the literature in the domains of behavioural InfoSec and XAI in the next section. Then, we conceptualise the theoretical structure of the XSec construct and describe its dimensions. The research design to test the construct's validity is presented in the following section, followed by our analysis and findings, discussions of results, and conclusion.

2 Literature Review

2.1 Behavioural InfoSec

Researchers in the behavioural InfoSec field have investigated employees' InfoSec perceptions and behaviours, especially their compliance with InfoSec policies. Early research in the field drew on deterrence theory to explain how sanctions and rewards influence employees' intention to comply with InfoSec policies (D'Arcy and Herath 2011). However, sanctions may create job stress and privacy invasion which in turn lead to undesirable InfoSec behaviours (D'Arcy et al. 2014). Another research stream focused on organisational control factors to explain employees' intention to comply. When employees have access to InfoSec resources, they tend to believe in their efficacy to accomplish the InfoSec-related tasks, which increases intention to comply with InfoSec policies (Dhillon et al. 2020).

Beyond the socio-organisational factors, recent studies have started to examine the characteristics of InfoSec itself. Bulgurcu et al. (2010) found that perceived benefits and costs of compliance and noncompliance affected employees' attitudes, which in turn, influence their intentions to perform these behaviours. Previous studies also found that perceived usefulness, perceived response cost and response efficacy influence employees' intention to comply (Herath and Rao 2009). If employees perceive InfoSec measures as restricting their freedom, they may develop the intention to commit computer abuses (Lowry et al. 2015). In contrast, when employees can make situational adjustments of how to comply with InfoSec policies, then compliance is more likely (Kirlappos and Sasse 2014).

The employees' perceived justice of sanction policies was found to affect their intention to comply. Employees desire to have a voice in the sanction process, in which it will be explained to them why they receive a sanction and how to avoid being sanctioned in the future (Xue et al. 2011). Similarly, Stanton

et al. (2005) discussed that successful InfoSec policies provide acceptable and reasonable monitoring and enforcement systems. The communication of InfoSec policies also influences employees' intention to comply. Siponen et al. (2010) emphasised the visibility of InfoSec actions, campaigns, advertisements, as well as formal and informal communications to increase InfoSec awareness. Lowry et al. (2015) analysed the adequacy of how InfoSec measures were explained and discussed that such explanation should be thorough, reasonable, and timely.

Despite the increasing number of studies on InfoSec compliance, getting employees to perform desirable InfoSec behaviours remains a challenge and requires the cooperation between top management and the employees (Soomro et al. 2016). In this paper, we aim to examine the characteristics of organisational InfoSec that exemplify the misalignment of perceptions and expectations between top management and employees. Based on the notion that individuals have their own unique perceptions of their environments, we use the concept of a black box to examine organisational InfoSec. We argue that the misalignment of InfoSec perceptions and expectations between top management and the employees is caused by these actors' different interpretations of the InfoSec measures and policies as a black box. Consequently, the remedy to reduce such misalignment is to open the black box and render it 'explainable'. This line of thought led us to look beyond the behavioural InfoSec literature and examine a similar, relevant, and timely concept: explainable artificial intelligence (XAI) to inspire our study.

2.2 Explainable Artificial Intelligence (XAI)

Algorithmic decisions are widely applied in different domains including job allocation, loans, or insurance. However, people are concerned about the justice, arbitrariness, and (in)dignity of these decisions (Binns et al. 2018). The main reason for such concerns is that people in general do not have a thorough understanding about the algorithmic decision-making process (Hagras 2018). Decision-making processes driven by artificial intelligence (AI) are usually considered as a black box, in which the outputs are provided with few indications of the internal mechanisms, logic, or how the input data is handled (Brunk et al. 2019; Hagras 2018). As the outputs of these algorithmic decision-making systems have direct impacts on their users and other stakeholders, the impacted parties want to understand how such decision-making processes function (Brunk et al. 2019). In this context, recent research has proposed the concept of "explainable AI" (XAI) that refers to the degree to which algorithmic decisions can be understood, interpreted and analysed by human actors (Hagras 2018). On one hand, the goal of explanations is for the impacted individuals to understand the system and thus increase their perception of transparency (Brunk et al. 2019). On the other hand, precise explanations lead to accountable AI agents that satisfy the standards of fairness (Binns et al. 2018).

Previous research on human-AI interaction has studied various methods and models on how to develop good explanations (Mueller et al. 2019, Halpern and Pearl 2005, Bostrom and Yudkowsky 2014). The concept of explanation is related to the cognitive process, in which causal analyses provide explanatory value (Halpern and Pearl 2005). In this sense, explanation is referred to as the process of generating the casual analysis that focuses on understanding past events, e.g. why something happened or how something worked (Wilkenfeld and Lombrozo 2015). Explanation is also related to the notion of counterfactual reasoning. It involves asking hypothetical counterfactual questions of "what if?". This approach suggests that the explanator provides information about how effects or outcomes might change or be different if they have different inputs (Wilkenfeld and Lombrozo 2015).

Another aspect of explainable AI systems concerns on social aspects and ethical issues regarding how AI systems impact society (Bostrom and Yudkowsky 2014). An extensive review by Mueller et al. (2019) identified four critical social and ethical aspects of explainable algorithms, which are: transparency (Hayes and Shah 2017), accountability (Kroll et al. 2016), safety and privacy, fairness and bias (Hajian et al. 2015). Likewise, Hagras (2018) argued that the concept of XAI should draw on the key aspects of transparency, compliance, fairness, accountability and ethics. On this background, our study takes a social-ethical approach to understand the misalignment between top management and employees.

In general, people are hesitant to adopt new technologies that are not trustworthy, interpretable and controllable, which has led to an increasing demand for ethical AI systems (Barredo Arrieta et al. 2020; Kumar et al. 2020). Therefore, most XAI studies that focus on social and ethical aspects emphasise the interpretability, understandability, transparency, accountability and trust when explaining the acceptance of algorithmic decision-making models and their outcomes (Barredo Arrieta et al. 2020; Binns et al. 2018; Brunk et al. 2019; Lepri et al. 2018). While understandability denotes a model's capability to support human understanding of how the model works without any need for explaning its internal algorithm, explainability refers to explanation as the interface between humans and an algorithmic decision maker that is, at the same time, both an accurate proxy of the decision maker and

comprehensible to humans (Guidotti et al. 2018). Interpretability then emphasises the ability to explain or to provide meanings in understandable terms to humans (Binns et al. 2018). AI systems are furthermore expected to meet the public's expectation of moral, justice and non-bias behaviours, while ensuring responsibility and accountability for their outcomes, both before and after development, deployment and use (Kumar et al. 2020). Barredo Arrieta et al. (2020) argued that opaque algorithmic decision-making systems could be considered as complex black-box models. In that sense, transparency, which is the opposite of blackboxness, is a prerequisite and the communication means for a direct understanding of the mechanism by which model works (Barredo Arrieta et al. 2020).

3 Fairness-Accountability-Transparency Model

Consistent with XAI, XSec refers to the quality of the explanations about or 'explainability' of organisational InfoSec, which is necessary for persuading employees to adopt InfoSec. Such explainability should be perceived by the employees during the development and implementation of InfoSec measures and policies. Drawing on the notion of explainable InfoSec, Pieters (2011) used the black box analogy to describe both AI and InfoSec and argued for the importance of providing explanations for InfoSec mechanisms to gain users' and impacted stakeholders' trust. A relevant but different concept to XSec is usable InfoSec, which focuses on the guidelines and designs for InfoSec systems, especially their user interfaces, that make them easy to use and trustworthy (Kirlappos and Sasse 2014). More recently, Viganò and Magazzeni (2018) presented a framework of questions for what to consider to achieve XSec, such as who provides/receives the explanations, what is explained, when explanations should be given, and how InfoSec explanations should be stated. Our research builds upon the work of Pieters (2011) and complements Viganò and Magazzeni (2018) by proposing a theoretical structure for the XSec construct, which consists of specific dimensions, and a set of measurement items.

Recent studies suggested that employees are motivated to comply with InfoSec policies when they understand InfoSec risks and accept the reasonable cost of compliance (Bulgurcu et al. 2010; Kirlappos and Sasse 2014). For example, Lowry et al. (2015) found that if employees are notified in advance of changes in the InfoSec policies, their perception of explanation adequacy increases and thus they are less likely to perform undesirable behaviours. Similarly, Doherty and Tajuddin (2018) suggested that when users perceive high value of their information, they are more willing to comply with InfoSec policies. In contrast, if employees feel personally invaded by organization's excessive monitoring, they are more likely to commit computer abuses (Lowry et al. 2015).

In accordance with the factors related to XAI identified above, recent studies used the fairness-accountability-transparency (FAT) model in different areas of computer science, social science and artificial intelligence (Lepri et al. 2018; Shin and Park 2019; Shin et al. 2020; Sokol et al. 2019). XAI scholars use the model as it provides key constructs for ethical AI and XAI (Mueller et al. 2019). In the model, perceived fairness refers to an assessment of whether an algorithm is fair and does not discriminate any party, and perceived accountability focuses on who should be held liable for the results of the algorithms (Shin and Park 2019). Perceived transparency refers to the provision of a clear and uncontroversial purpose, the structure, and the underlying actions of the algorithms which are deployed in the decision-making processes (Shin and Park 2019).

The importance of fairness is based on the assessment that algorithms do not always function fairly, especially as algorithms are developed by humans who are prone to bias (Shin and Park 2019). Thus, fairness of an algorithm has to be judged based on its accuracy, precision, non-bias and the ability to find related results (Lepri et al. 2018). Likewise, transparency is essential due to the complicated working of AI systems, and the general public's lack of sufficient technical knowledge to understand them. Thus, if stakeholders can interpret and understand the operations of an AI system easily, they might abandon and not feel the need to access the underlying algorithm (Barredo Arrieta et al. 2020). Accountability is also necessary to build users' and impacted stakeholders' trust, as it highlights the responsibility of the operator in case algorithms err in various possible ways (Shin and Park 2019). In their research, Shin and Park (2019) found that the perceptions of these three factors constitute algorithmic credibility, which had a positive influence on trust, and subsequently promoted the acceptance of algorithmic decisions. Similarly, Brunk et al. (2019) highlighted that the detailed explanations of algorithmic decisions improve perceived transparency, and thus increase both trust in the decisions and their acceptance. In line with our previous deliberations, we argue that InfoSec measures and policies are similar to algorithmic decision-making systems since they both might represent black boxes that are interpreted differently by different perceivers. We therefore conceptualise XSec through these dimensions and discuss their relevance for InfoSec as follow.

3.1 Fairness

In an organisational context, fairness refers to the employees' perception of being treated fairly by their organisation and their colleagues (Niehoff and Moorman 1993). Fairness is assessed when employees compare their inputs to work (e.g., time, effort) to the outcomes they receive (e.g., salary, allowance), which is specifically referred to as distributive fairness (Niehoff and Moorman 1993). Thibaut et al. (1973) suggested that fairness in an organisation also concerns procedural justice, which focuses on the employees' evaluation of work-related procedures and decision-making processes. Bies and Moag (1986) discussed the concept of interactional fairness, which refers to the treatment that an employee receives in terms of explanations for decisions with compassion and respect.

The concept of fairness in the organisational context is applied around algorithmic decision-making (Binns et al. 2018). For instance, users may demand organisations to explain the logic behind their algorithmic decision-making systems (i.e., informational justice), which enables people to assess whether the logic of the systems is just (i.e., procedural justice). This in turn moderates their assessment of fairness of the decision outcomes (i.e., distributive justice). In Shin et al.'s (2020) study, fairness is defined when the algorithm provides no favouritism and does not discriminate against people and the sources of data, and its data sources are identified, logged, and benchmarked.

On this basis, we argue that fairness in explainable InfoSec refers to the employees' assessment of justice in the organisation's development and implementation of InfoSec. For instance, they would expect InfoSec requirements to be the same across organisational levels and there is no favouritism to any department or position. Employees may also perceive fairness when they feel that rewards and punishments for InfoSec behaviours are appropriate and reasonable. Fairness in InfoSec-related decisions and processes is displayed when top management shows considerations for the employees' cost of compliance (e.g., work impediment, inflexibility).

3.2 Accountability

Accountability is about individuals' perception of how much responsibility they need to hold for a certain action. In the AI context, an organisation deploying an AI system should be held accountable for the system's actions (Garfinkel et al. 2017). Shin et al. (2020) argued that accountability of algorithmic decision-making consists of three facets: auditability, controllability, and responsibility. Intelligent systems should be open for audit to determine whether the algorithmic decision-making processes are compliant (Shin et al. 2020). Controllability refers to an actor's ability to influence conditions and processes that are conducive to goal fulfilment (Shin et al. 2020). Controllability is important for automated AI systems, as it gives their users and other impacted parties the confidence that the systems behave in manners that can be anticipated.

In the context of InfoSec, accountability refers to the employees' perception of the extent to which they are responsible for their compliance and noncompliance. Moreover, Lowry et al. (2015) emphasised that organisations should be held accountable for the loss of freedom caused by unreasonable InfoSec initiatives. In line with accountability in algorithmic decision-making, the design and implementation of organisational InfoSec should be auditable by the employees. Top management should pursue a participatory approach to the design of organisational InfoSec and consider the employees' inputs (Kirlappos and Sasse 2014). By doing so, the employees would perceive a higher level of controllability over the organisational InfoSec that they need to adopt.

3.3 Transparency

Transparency is necessary for increasing potential users' adoption of technology, by providing explanations about how the technology operates and about its behaviours (Garfinkel et al. 2017). Regardless of algorithms' complexity, their logic and outputs need to be communicable and explainable in layman's terms (Garfinkel et al. 2017). In the context of algorithmic decision-making, transparency involves not only the understandable explanations about the systems' recommendations, but also about the internal states, evaluation process, and the criteria that are employed by the algorithms during the decision-making process (Shin et al. 2020).

We argue that transparency of explainable InfoSec refers to the clear explanations for why InfoSec measures and policies need to be implemented and adopted in the work environment. Previous studies have emphasised the adequate communication of InfoSec, including aspects such as visibility of internal advertisements, explanations of InfoSec incidents, and quality of InfoSec policies (Lowry et al. 2015). The concept of transparency in InfoSec focuses on the quality of the explanations that are observable and delivered through the design and implementation of organisational InfoSec, rather than their communication. Organisational InfoSec can be considered as transparent when there are visible

indicators and mechanisms that let the employees know whether the InfoSec measures are effective and used properly, thereby justifying and support their importance.

4 Research Design

In the previous section, we adapted the FAT model to describe the theoretical construct of XSec. Consequently, for our research design we contextualised the measurement items used by Shin et al. (2020) for measuring the XAI construct, by re-wording the questions to make them relevant in the organisational InfoSec context. The questionnaire we developed consisted of 18 Likert scale items that measure different facets of the three key dimensions fairness, accountability, and transparency of XSec. We used six-point Likert scales, which ranged from "Strongly Disagree" to "Strongly Agree" where a neutral position was made unavailable. This design reduces the social desirability bias which may result from respondents choosing the midpoint that represents uncertainty (Leung 2011; Garland 1991).

We included additional questions to collect data about the respondents' demographics and their intention to comply with InfoSec. When collecting self-reported data via surveys, respondents might provide answers that are socially desirable which can impact the validity of results (D'Arcy et al. 2014). Consequently, we added questions to measure social desirability bias and performed common method bias tests in the later stage. The measurement items for social desirability bias and intention to comply were adopted from the studies of D'Arcy et al. (2014) and Bulgurcu et al. (2010) respectively. The measurement items are summarised in the appendix.

A local market research agency was employed to administer the online survey. We provided the agency with our sampling frame, which focused on employees working in Vietnam who used computers and the Internet for their daily work. We also specified that the sample would have to have balanced ratios of genders and company sizes. Data collection was performed from April 2020 to June 2020, returning a total of 2,036 responses, of which 1,616 responses were complete, resulting in a response rate of 79 per cent. Data cleaning was performed to treat missing data and responses provided by speeders (i.e., those who completed the survey under five minutes) and straight liners (i.e., sequences of the same responses for every question). We retained a sample of 724 valid responses in total. Of 724 respondents, 329 are males (45.4 per cent) and 395 are females (54.6 per cent); the majority of the respondents falls within the age range of 23-35 years old (51.4 per cent). Most respondents hold a bachelor's or graduate-level degree (80.4%). Forty-two per cent of respondents work in companies that have 50 employees or below. The full description of the sample's characteristics is omitted due to page limitation.

5 Analysis and Findings

To examine the theoretical structure of the XSec construct, we performed exploratory (EFA) and confirmatory factor (CFA) analyses, followed by structural equation modelling (SEM) to analyse the effect of the newly developed XSec construct on the employees' intention to perform InfoSec compliance. SEM is further performed to test the usefulness of the XSec construct in explaining InfoSec compliance, which is a common and important aim of behavioural InfoSec research.

EFA was performed to explore the natural theoretical structures available within the data set. We determined the number of factors by using the Kaiser's criterion i.e., any additional factor's eigenvalue must be larger than 1, rather than fixing the number of factors to be extracted. For the EFA, the maximum likelihood method was used for extraction to ensure consistency with the CFA and SEM analyses that were performed later. We employed the 'direct oblimin' rotation method which allows the factors to be correlated. In contrast to our anticipation, the EFA revealed that there were only two significant dimensions for the construct XSec. There were many cross loadings between several items of the accountability set and those of transparency and fairness, indicating that the accountability factor may be redundant. We further tested this result by performing separate EFAs for the pairs of accountability-transparency and accountability-fairness. Both EFAs failed to produce separate factors, i.e., only one factor was extracted, which thus reinforced the result that the construct XSec should have two dimensions only i.e., fairness and transparency.

We then re-ran the EFA with only the two sets of items for fairness and transparency and the other two sets of items for intention to comply and social desirability by default, which produced a clear pattern of four sets of items loaded into their respective factors. We further removed item FAI2 (i.e., item for measuring fairness) due to its cross loadings on both factors. Items SD3, SD4, and SD5 (i.e., items for measuring social desirability) were also removed due to poor loadings below 0.35. The items' factor loadings ranged from 0.542 to 0.950, which confirmed convergent validity. The correlation between transparency and fairness was 0.717, indicating that the two factors had a shared variance of 52 per

cent. In terms of discriminant validity, such high shared variance suggested that these two factors were not too different theoretically; this finding also reinforced our initial intention to model XSec as a second-order construct that is reflected by two lower-order constructs, i.e., fairness and transparency.

We carried out CFA to validate the theoretical structures that were identified from the EFA. The objectives of performing CFA are to achieve model fit, i.e., the theoretical structures accurately reflect the data, and good validity and reliability of the model. During the CFA, further items were removed, i.e., TRA6, TRA7, TRA8 (i.e., items for measuring transparency). The final model achieved excellent fit: Chi-square/df = 1.251, p-value = 0.077, CFI = 0.997, GFI = 0.983, AGFI = 0.974, SRMR = 0.021, RMSEA = 0.021, PCLOSE = 1.000. The acceptance criteria for these goodness-of-fit statistics can be found in Hu and Bentler's (1999) study. The model also achieved good convergent and discriminant validity, as well as reliability of each construct (see the statistics in table 1).

	CR	AVE	MSV	SD	INT	XSec
SD	0.715	0.556	0.181	0.746		
INT	0.885	0.720	0.199	0.426***	0.848	
XSec	0.870	0.770	0.199	0.381***	0.445***	0.878
Threshold	>0.7	>0.5	<AVE			

Note: SD = social desirability; INT = intention to comply; XSec = explainable InfoSec;
CR = composite reliability; AVE = average variance extracted; MSV = maximum shared variance

Table 1. Model validity results

We performed the zero and equal constraints tests with social desirability as the common marker variable to detect the potential common method bias issue. The detailed results of these tests are not reported due to the page limitation. Nonetheless, with the p-values of both tests lower than the statistical significance threshold of 0.05, we concluded that social desirability bias existed in our data i.e., the respondents might have provided socially desirable responses. Consequently, we included social desirability as a control variable in our SEM step.

For SEM, we included age, gender, education, tenure, company size, and industry as control variables for intention to comply, along with social desirability. The model fit during this stage remained excellent: Chi-square/df = 1.441, p-value = 0.001, CFI = 0.991, GFI = 0.974, AGFI = 0.959, SRMR = 0.024, RMSEA = 0.025, PCLOSE = 1.000. Figure 1 illustrates the structural model and results of the SEM analysis. It is worth noting that the dotted arrows from XSec to fairness and transparency denote both the impacts of XSec on the two constructs and XSec's nature as a higher/second-order construct with fairness and transparency as its properties or dimensions. The construct XSec does not have its own set of measurement items (see table 2 in the appendix) and can be represented by the lower/first-order constructs of fairness and transparency.

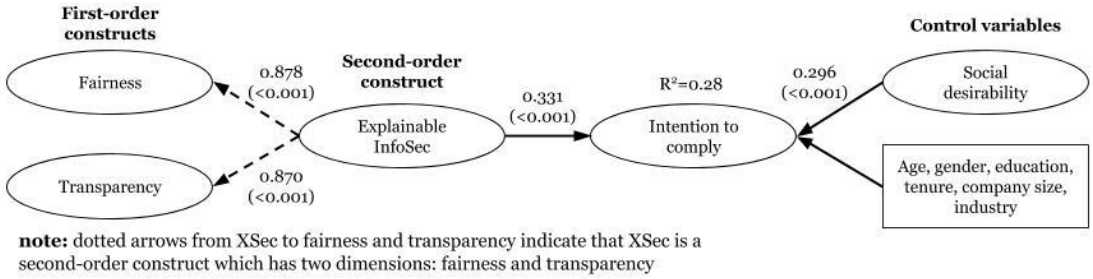

Figure 1. Structural model in SEM analysis (figures in brackets are p-values)

Perceived XSec has a positive and statistically significant impact on intention to comply, after the employees' demographics and social desirability have been controlled for. Company size is one of the demographical variables that has the p-value closest to the statistical significance threshold (p-value = 0.069) and a positive effect on intention to comply. This finding suggests that larger companies usually have more sophisticated and formal InfoSec mechanisms, which may motivate employees' compliance more. Overall, the model explained 28 per cent of the variance in the intention to comply.

6 Discussion

Organisational InfoSec is explainable if the employees perceive the development and implementation of InfoSec measures and policies as transparent and fair. Perceived fairness refers to the employees' perceptions of organisational justice in the processes of developing and implementing InfoSec. The concept of organisational justice has been examined by behavioural InfoSec studies, for instance, to explain the employees' intention to perform computer abuse when they feel their rights are violated (Willison and Warkentin 2013). Researchers also explored fairness of InfoSec as explanation adequacy, i.e., how reasonable, thorough, and candid the explanations about InfoSec are, and as the quality of InfoSec policies (Lowry et al. 2015). Our definition of perceived fairness expands the above concepts, as it focuses on the impartiality of how InfoSec measures and requirements are decided. This impartiality is also observed and evaluated by the employees based on how InfoSec is actioned in the work environment rather than reflected in policies. Our concept also includes perceived fairness in top management's considerations for the employees' scope of work when making InfoSec-related decisions, which refers to InfoSec stress and overload that affect InfoSec behaviours (D'Arcy et al. 2014).

Perceived transparency refers to the quality of explanations about the importance of and reasons for implementing organisational InfoSec, thereby motivating employees' compliance. Another antecedent of InfoSec compliance, which is related to perceived transparency, is perceived InfoSec climate. InfoSec climate includes employees' perceptions of different aspects in the work environment that highlight the importance of InfoSec, such as the InfoSec behaviours performed by peers and direct supervisors, or the sophistication of InfoSec mechanisms implemented in the workplace (Chan et al. 2005). We argue that a work environment where the level of InfoSec climate is high, i.e., priority is given to InfoSec as reflected by visible InfoSec behaviours performed by peers and management, provides observable indicators that promote transparency of organisational InfoSec. Similarly, response effectiveness is another antecedent of InfoSec compliance that is relevant to perceived transparency. Response effectiveness refers to the employees' perceptions of usefulness of InfoSec measures in protecting information assets and preventing InfoSec threats (Herath and Rao 2009). Perceived transparency also focuses on the explanations about such usefulness (e.g., see items TRA1–4 in appendix). Overall, we argue that perceived transparency is conceptually different from perceived InfoSec climate and response effectiveness.

In terms of theoretical contributions, this research presents the new theoretical construct XSec for measuring the explainability of organisational InfoSec, or the extent to which InfoSec is perceived as fair and transparent in the work environment. The SEM analysis confirmed that XSec had a positive impact on the employees' intention to comply, after social desirability and other demographical factors were accounted for (see figure 1). As such, we contribute a novel antecedent of compliance intention to the current list of widely examined factors in the behavioural InfoSec area, such as attitude, self-efficacy, perceived vulnerability, to name a few. The construct XSec was developed by adapting the concept of XAI, which is another emerging concept in both information systems and computer science disciplines. This opens up opportunities for theory development in the behavioural InfoSec field. For instance, future studies may explore the factors that increase employees' perceptions of InfoSec's explainability from the behavioural perspective. On the other hand, technical InfoSec research may focus on designing user interfaces and mechanisms that reinforce InfoSec's explainability. Research into the concept of XSec and its applications can be extended beyond the work environment to the context of InfoSec in private and home environments as well.

In terms of practical contributions, the items of the fairness and transparency constructs to measure XSec that were not dropped as presented in the appendix, can be used to benchmark the level of explainability of InfoSec in a work environment. The two facets of XSec, fairness and transparency, indicate the specific areas that need attention. In particular, management should review the organisational factors which impact on employees' perception of whether InfoSec decisions are fair and just, such as exaggerating InfoSec roles and requirements, the possible lack of employees' involvement and participation in designing InfoSec mechanisms, or the possible inconsistent assignment of InfoSec requirements to organisational members. When transparency is perceived as low, communication about InfoSec within the workplace needs to be intensified. Although the usefulness and importance of InfoSec measures have consistently been found to be motivations for InfoSec compliance (Sommestad et al. 2014), it is vital to communicate such information to the employees to achieve their compliance.

7 Conclusion

In this research, we proposed a new antecedent of the employees' intention to comply with InfoSec called explainable InfoSec (XSec) which we developed based on the concept of XAI. Empirical data was collected from 724 office workers in Vietnam who use IT, computers, and the Internet daily for their work. We performed EFA, CFA, and SEM to validate the theoretical structure of this new XSec concept. In doing so, we further analysed the impact of the XSec concept on the employees' intention to comply. We found that the new XSec concept has a statistically significant and positive effect on intention to comply, after demographical factors and social desirability bias were controlled for. A set of measurement items derived from the literature were used and refined throughout our analysis to collect data about XSec. The XSec concept contributes new explanations for InfoSec compliance and has the potential to do so as well for other InfoSec behaviours that are studied by the current literature.

The main limitation of our research lies in the nature of our sample which only consists of office workers who work in Vietnam. Therefore, the generalisability of the findings is limited to this specific context. We encourage future studies to collect empirical data from other countries and contexts to validate the theoretical structure of our proposed XSec construct. The theorising process can be enhanced, by conducting further research that follows other rigorous approaches for theory development, such as grounded theory or ethnography. Additional investigations of the new XSec concept will also lead to new and practical implications that will be beneficial for people-centric InfoSec workplaces, where InfoSec compliance becomes the result of conviction rather than of enforcement.

8 References

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., and Herrera, F. 2020. "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible Ai," *Information Fusion* (58), pp. 82-115.

Bies, R. J., and Moag, J. S. 1986. "Interactional Communication Criteria of Fairness," *Research in Organizational Behaviour* (9), pp. 289-319.

Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., and Shadbolt, N. 2018. "'It's Reducing a Human Being to a Percentage' Perceptions of Justice in Algorithmic Decisions," in *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1-14.

Boss, S. R., Kirsch, L. J., Angermeier, I., Shingler, R. A., and Boss, R. W. 2009. "If Someone Is Watching, I'll Do What I'm Asked: Mandatoriness, Control, and Information Security," *European Journal of Information Systems* (18:2), pp. 151-164.

Bostrom, N., and Yudkowsky, E. 2014. "The Ethics of Artificial Intelligence," *The Cambridge handbook of artificial intelligence* (1), pp. 316-334.

Brunk, J., Mattern, J., and Riehle, D. M. 2019. "Effect of Transparency and Trust on Acceptance of Automatic Online Comment Moderation Systems," in *2019 IEEE 21st Conference on Business Informatics (CBI)*: IEEE, pp. 429-435.

Bulgurcu, B., Cavusoglu, H., and Benbasat, I. 2010. "Information Security Policy Compliance: An Empirical Study of Rationality-Based Beliefs and Information Security Awareness," *MIS Quarterly* (34:3), pp. 523-548.

Chan, M., Woon, I., and Kankanhalli, A. 2005. "Perceptions of Information Security in the Workplace: Linking Information Security Climate to Compliant Behavior," *Journal of Information Privacy and Security* (1:3), pp. 18-41.

D'Arcy, J., and Herath, T. 2011. "A Review and Analysis of Deterrence Theory in the Is Security Literature: Making Sense of the Disparate Findings," *European Journal of Information Systems* (20:6), pp. 643-658.

D'Arcy, J., Herath, T., and Shoss, M. K. 2014. "Understanding Employee Responses to Stressful Information Security Requirements: A Coping Perspective," *Journal of Management Information Systems* (31:2), pp. 285-318.

Dhillon, G., Talib, Y. Y. A., and Ng Picoto, W. 2020. "The Mediating Role of Psychological Empowerment in Information Security Compliance Intentions," *Journal of the Association for Information Systems* (21:1), pp. 152-174.

Doherty, N. F., and Tajuddin, S. T. 2018. "Towards a User-Centric Theory of Value-Driven Information Security Compliance," *Information Technology & People* (31:2), pp. 348-367.

Garfinkel, S., Matthews, J., Shapiro, S., and Smith, J. 2017. "Toward Algorithmic Transparency and Accountability," *Communications of the ACM* (60:9), pp. 5-5.

Garland, R. 1991. "The mid-point on a rating scale: is it desirable?" *Marketing bulletin* (2):66-70.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. 2018. "A Survey of Methods for Explaining Black Box Models," *ACM computing surveys (CSUR)* (51:5), pp. 1-42.

Hagras, H. 2018. "Toward Human-Understandable, Explainable AI," *Computer* (51:9), pp. 28-36.

Hajian, S., Domingo-Ferrer, J., Monreale, A., Pedreschi, D., and Giannotti, F. 2015. "Discrimination- and Privacy-Aware Patterns," *Data Mining and Knowledge Discovery* (29:6), pp. 1733-1782.

Halpern, J. Y., and Pearl, J. 2005. "Causes and Explanations: A Structural-Model Approach. Part Ii: Explanations," *The British journal for the philosophy of science* (56:4), pp. 889-911.

Hayes, B., and Shah, J. A. 2017. "Improving Robot Controller Transparency through Autonomous Policy Explanation," *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*: IEEE), pp. 303-312.

Herath, T., and Rao, H. R. 2009. "Protection Motivation and Deterrence: A Framework for Security Policy Compliance in Organisations," *European Journal of Information Systems* (18:2), pp. 106-125.

Hu, L.T. and Bentler, P.M. 1999. "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives." *Structural Equation Modeling: a Multidisciplinary Journal* (*6:1)*, pp.1-55.

Kirlappos, I., and Sasse, M. A. 2014. "What Usable Security Really Means: Trusting and Engaging Users," in *International Conference on Human Aspects of Information Security, Privacy, and Trust*: Springer, pp. 69-78.

Kroll, J. A., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., and Yu, H. 2016. "Accountable Algorithms," *U. Pa. L. Rev.* (165), p. 633.

Kumar, A., Braud, T., Tarkoma, S., and Hui, P. 2020. "Trustworthy AI in the Age of Pervasive Computing and Big Data," *arXiv preprint arXiv:2002.05657*.

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., and Vinck, P. 2018. "Fair, Transparent, and Accountable Algorithmic Decision-Making Processes," *Philosophy & Technology* (31:4), pp. 611-627.

Leung, S.-O. 2011. "A Comparison of Psychometric Properties and Normality in 4-, 5-, 6-, and 11-Point Likert Scales." *Journal of social service research* 37 (4):412-21. doi: 10.1080/01488376.2011.580697.

Lowry, P. B., Posey, C., Bennett, R. J., and Roberts, T. L. 2015. "Leveraging Fairness and Reactance Theories to Deter Reactive Computer Abuse Following Enhanced Organisational Information Security Policies: An Empirical Study of the Influence of Counterfactual Reasoning and Organisational Trust," *Information Systems Journal* (25:3), pp. 193-273.

Mueller, S. T., Hoffman, R. R., Clancey, W., Emrey, A., and Klein, G. 2019. "Explanation in Human-Ai Systems: A Literature Meta-Review, Synopsis of Key Ideas and Publications, and Bibliography for Explainable Ai," *arXiv preprint arXiv:1902.01876*.

Niehoff, B., and Moorman, R. 1993. "Justice as a Mediator of the Relationship between Methods of Monitoring and Organizational Citizenship Behavior," *Academy of Management Journal* (36:3), p. 527.

Pieters, W. 2011. "Explanation and Trust: What to Tell the User in Security and AI?," *Ethics and Information Technology* (13:1), pp. 53-64.

Sasse, M. A., Brostoff, S., and Weirich, D. 2001. "Transforming the 'Weakest Link'—a Human/Computer Interaction Approach to Usable and Effective Security," *BT Technology Journal* (19:3), pp. 122-131.

Shin, D. 2020. "How Do Users Interact with Algorithm Recommender Systems? The Interaction of Users, Algorithms, and Performance," *Computers in Human Behavior*.

Shin, D., and Park, Y. J. 2019. "Role of Fairness, Accountability, and Transparency in Algorithmic Affordance," *Computers in Human Behaviour* (98), pp. 277-284.

Shin, D., Zhong, B., and Biocca, F. A. 2020. "Beyond User Experience: What Constitutes Algorithmic Experiences?," *International Journal of Information Management* (52).

Siponen, M., Pahnila, S., and Mahmood, M. A. 2010. "Compliance with Information Security Policies: An Empirical Investigation," *Computer* (43:2), pp. 64-71.

Sommestad, T., Hallberg, J., Lundholm, K., and Bengtsson, J. 2014. "Variables Influencing Information Security Policy Compliance : A Systematic Review of Quantitative Studies," *Information Management & Computer Security* (22:1), pp. 42-75.

Soomro, Z. A., Shah, M. H., and Ahmed, J. 2016. "Information Security Management Needs More Holistic Approach: A Literature Review," *International Journal of Information Management* (36:2), pp. 215-225.

Stanton, J., Yamodo-Fagnot, I., and Stam, K. 2005. "The Madness of Crowds: Employees Beliefs About Information Security in Relation to Security Outcomes," in *The Security Conference*.

Thibaut, J., Walker, L., LaTour, S., and Houlden, P. 1973. "Procedural Justice as Fairness," *Stan. L. Rev.* (26), p. 1271.

Viganò, L., and Magazzeni, D. 2018. "Explainable Security," *arXiv preprint arXiv:1807.04178*.
Wilkenfeld, D. A., and Lombrozo, T. 2015. "Inference to the Best Explanation (Ibe) Versus Explaining for the Best Inference (Ebi)," *Science & Education* (24:9-10), pp. 1059-1077.
Willison, R., and Warkentin, M. 2013. "Beyond Deterrence: An Expanded View of Employee Computer Abuse," *MIS Quarterly*, pp. 1-20.
Xue, Y., Liang, H., and Wu, L. 2011. "Punishment, Justice, and Compliance in Mandatory It Settings," *Information Systems Research* (22:2), pp. 400-414.

Appendix

Code	Item	Reference
FAI1	There is no favouritism or difference in InfoSec requirements for organisational members across departments, positions, or seniority levels.	Adapted from Shin et al. (2020)
FAI2*	There is clarity about the information sources and references (e.g., standards, regulations, expert opinions etc.) that my company used to come up with the InfoSec measures.	
FAI3	I believe the design and implementation of InfoSec requirements of my company follow due process of impartiality with no prejudice.	
FAI4	Overall, the InfoSec requirements and responsibilities for all employees are fair and reasonable in my company.	
FAI5	When decisions are made about InfoSec measures, my company considers our scope of work with respect and dignity.	
ACC1*	In my company, it is clear who will be accountable for the outcomes and consequences of using the InfoSec measures.	Adapted from Shin et al. (2020)
ACC2*	There are mechanisms in my company to hold employees accountable for InfoSec-related issues.	
ACC3*	The InfoSec measures in my company are audited adequately by third party bodies (e.g., IT auditing firms and government agencies) and/or by internal organisational members.	
ACC4*	Employees in my company are able and/or encouraged to provide inputs regarding how organisational InfoSec should be maintained.	
ACC5*	I can flexibly make local and situational adjustments when using the InfoSec measures in my company.	
TRA1	It is clear and understandable why my company prefers the current InfoSec measures.	Adapted from Shin et al. (2020)
TRA2	It is clear and understandable why employees are recommended or required to use InfoSec measures in my company.	
TRA3	It is explained to me why the recommended InfoSec measures are effective and useful in my company.	
TRA4	It is explained to me or I can tell whether the use of InfoSec measures in my company have been effective or not.	
TRA5	There are visible indicators and mechanisms in my workplace to let employees know whether InfoSec measures are being used properly.	
TRA6*	Overall, decisions and efforts related to organisational InfoSec are transparent, visible, and/or clearly communicated.	
TRA7*	It is explained to me how the recommended InfoSec measures can help to protect computers and information systems in my company.	
TRA8*	I understand the benefits and limitations of the InfoSec measures in my company, as well as the outcomes and impacts that result from its use.	
INT1	I intend to comply with the requirements of the InfoSec policy) (ISP) of my organisation in the future.	Bulgurcu et al. (2010)
INT2	I intend to protect information and technology resources according to the requirements of the ISP of my organisation in the future.	
INT3	I intend to carry out my responsibilities prescribed in the ISP of my organisation when I use information and technology in the future.	
SD1	I am always courteous even to people who are disagreeable.	D'Arcy et al. (2014)
SD2	No matter who I'm talking to, I'm always a good listener.	
SD3*	I'm always willing to admit it when I make a mistake.	
SD4*	I have never intensely disliked anyone.	
SD5*	I would never think of letting someone else be punished for my wrong doings	

Note: FAI = Fairness; ACC = Accountability; TRA = Transparency; INT = Intention to comply; SD = Social desirability; asterisk (*) denotes items that were dropped

Table 2. Measurement items