6-18-2022

Who gets phished? Insights from a Contextual Clustering Analysis Across Three Continents

Muriel Frank
Institute for Information Systems, frank@wiwi.uni-frankfurt.de

Niklas Wagner
Goethe University Frankfurt, nw.niklas.wagner96@gmail.com

Lukas Manuel Ranft
Institute for Information Systems, lranft@wiwi.uni-frankfurt.de

WHO GETS PHISHED? INSIGHTS FROM A CONTEXTUAL CLUSTERING ANALYSIS ACROSS THREE CONTINENTS

Research Paper

Muriel Frank, Goethe University Frankfurt, Frankfurt, Germany, frank@wiwi.uni-frankfurt.de

Niklas Paul Wagner, Goethe University Frankfurt, Frankfurt, Germany, nw.niklas.wagner96@gmail.com

Lukas Manuel Ranft, Goethe University Frankfurt, Frankfurt, Germany, lranft@wiwi.uni-frankfurt.de

Abstract

Phishing attacks are one of the most prevalent cybersecurity threats to modern organizations. As a result, researchers and practitioners alike have pooled their strengths to understand who is most at risk of falling for phishing attacks. Since recent work calls for consideration of discrete context dimensions when examining phishing susceptibility, we use cluster analysis in conjunction with a large-scale phishing experiment to identify and scrutinize highly deceivable employees across three continents based on contextual factors. The results reveal salient similarities between employee groups in Europe, Australia, and North America. Consequently, our findings underscore the importance of classifying employees based on discrete contextual characteristics impacting their phishing susceptibility. Furthermore, the identified clusters have important implications for policymakers, awareness programs, and anti-phishing interventions, as they allow to better target individuals based on contextual attributes.

Keywords: Phishing susceptibility, cluster analysis, k-means algorithm, discrete context.

1 Introduction

A leading cause of information security incidents is people's susceptibility to deception (Goel *et al.*, 2017). Cybercriminals exploit this human weakness by orchestrating phishing messages and tricking e-mail recipients into revealing confidential information (Tambe Ebot, 2018; Wright *et al.*, 2014). In doing so, the attackers circumvent the established security and malware detection controls (Tambe Ebot, 2019). Not surprisingly, the Anti-Phishing Working Group (2021) reported a 50 percent increase in phishing attacks in 2020.

The increase in phishing attacks and their sophistication is prompting researchers around the world to better understand individuals' susceptibility to phishing and provide recommendations to prevent e-mail recipients from falling for fraudulent messages (Abbasi *et al.*, 2021; Moody *et al.*, 2017; Sarno *et al.*, 2020). However, studies using the same theoretical lenses reach conflicting conclusions (Wright *et al.*, 2020), and recommendations prove to be not very effective (Alsharnouby *et al.*, 2015; Goel *et al.*, 2017), especially when they promote one-size-fits-all approaches (Tambe Ebot, 2019). This suggests that phishing susceptibility varies across contexts, and contextual factors need to be considered to fully understand this phenomenon as well as to identify who is likely to fall victim to deceptive e-mails.

Recent work in the field of information security presents salient confirmation of the relevance of omnibus and discrete context factors for explaining and understanding phishing susceptibility (Wright *et al.*, 2020). The authors present a multi-level model of phishing susceptibility considering several social and task contextual elements, like centrality in informal IT advice networks and the work-task network. However, their findings are limited to the United States only and since previous literature suggests that individuals' phishing susceptibility differs by location (Butavicius *et al.*, 2017; Tembe *et al.*, 2014), accounting for cultural differences seems essential. Building on John's (2006) contextual framework, the present research intends to augment our understanding of contextual factors pertaining to phishing susceptibility and reveal (dis)similarities between groups of workers at risk of becoming phishing victims in Europe, Australia, and North America. Hence, we seek to answer the following research question:

Do discrete contextual factors help to identify employee groups susceptible to phishing across three continents?

Invoked by scholars who call for greater consideration of contextual influences (Sarker, 2016; Wright *et al.*, 2020), we used a contextual approach to study individual's phishing susceptibility across three continents. To gather the data for our analysis, we conducted a large-scale field experiment among employees of a pharmaceutical company with sites in Europe, Australia, and North America and then used cluster analysis. Our findings indicate that contextual factors can help identify groups of employees who are at risk of falling for phish. Consequently, our results may help individualize training approaches regarding safe e-mail practices for organizations with employees on different continents. This is consistent with the work of Dincelli and Chengalur-Smith (2020) and Tambe Ebot (2019), showing that individuals targeted by phishing e-mails exhibit different security behaviors and therefore need different awareness training.

The paper proceeds as follows: The next section presents related work and the research framework. We then describe the data collection procedure and the methodological approach, including the four phases required to conduct a clustering analysis. Ensuing, Section 4 discusses similarities and differences between at-risk groups across the three continents, delineates the implications following these results, and points to future research endeavors. Finally, Section 5 concludes the paper.

2 Related Work

Over the past twenty years, numerous researchers have studied the behavioral, economic, and technical aspects related to phishing (Abbasi, Zahedi, *et al.*, 2016; Chen *et al.*, 2020). They examined individuals' phishing susceptibility (e.g., Li *et al.*, 2020), behavioral interventions focusing on aiding users to detect malicious e-mails (e.g., Dincelli and Chengalur-Smith, 2020; Tambe Ebot, 2018) as well as technology-based approaches centering around reducing software vulnerabilities and warning subjects before clicking on deceptive e-mails (e.g., Nguyen *et al.*, 2021). Nevertheless, the phenomenon of phishing is not yet solved, and its influence on businesses and individuals remains significant (Greene *et al.*, 2018). Wright et al. (2020) see the main reason for this as a lack of consideration of context – and they are not alone in this opinion. Several IS scholars note that consideration of context in IS research is generally rare (Cheng *et al.*, 2016; Davison and Martinsons, 2016; Sarker, 2016), and even if contextual conditions are accounted for, they are most often inadequately conceptualized (Avgerou, 2019).

In the phishing literature, we find plenty of confirmation that researchers either acknowledge context to a limited extent (e.g., Goel *et al.*, 2017; Kim *et al.*, 2020; Li *et al.*, 2020) or relegate it to the role of control variables (e.g., Jensen *et al.*, 2017 for instance control for the job status, i.e., faculty staff and students). However, this contradicts the multifaceted nature of context and its impact on behavior (Johns, 2006). According to Gary Johns (2006), context comprises two different levels: the first is termed omnibus context, and the second is discrete context. Omnibus context refers to context broadly considered and informs about the who, the when, the where, and the why (Johns, 2006). The discrete context is nested within the omnibus context and can be differentiated into several subdimensions, such as task, social, and informational (Sarker, 2016). The task dimension encompasses contextual

influences with respect to organizational tasks, job roles, and responsibilities. The social context relates to social influences and interactions with peers that affect an individual's behavior. And the last dimension, the informational context, includes individual contextual characteristics that shape their behavior. Figure 1 displays the context-sensitive framework that we adapted from Johns (2006). In the following subsections, we review pertinent literature related to the three contextual dimensions to build the theoretical foundation of this study.

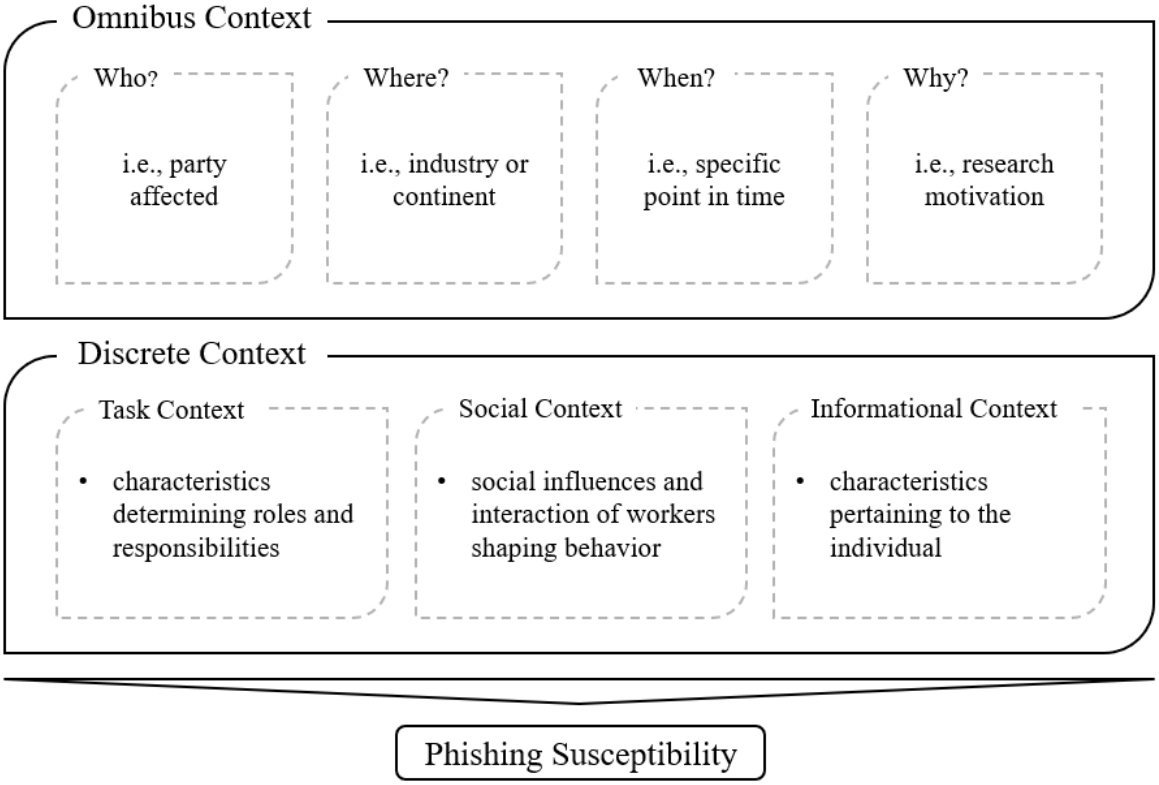

Figure 1. Contextual framework adapted from Gary Johns (2006).

2.1 Task Context & Phishing Susceptibility

The task context relates to contextual influences concerning organizational tasks, responsibilities, and roles (Johns, 2006). In modern organizations, workers' tasks differ according to their job levels which in turn influence their phishing susceptibility (Greene *et al.*, 2018). For instance, employees working at higher job levels, such as management, need specialized knowledge and have more responsibilities (Deloitte, 2015), which usually translates into higher salaries. Therefore, several scholars studied the influence of job levels on phishing susceptibility and found managers to be more vulnerable to phishing (Alwanain, 2019; Kim *et al.*, 2020). Interestingly, phishing failures of superiors tend to correspond to subordinates' posture, meaning that those working with superiors who got phished are more likely to be phishing victims (Coronges *et al.*, 2012). However, when performing routine tasks and monotonous work, employees are also more likely to get phished (Hanus *et al.*, 2021).

Because phishing is still evolving (Hanus *et al.*, 2021), many organizations pit on security education training and awareness programs to ensure that their workforce is aware of the threats and to train them in phishing detection (Dodge *et al.*, 2012). As shown in various research experiments (Jensen *et al.*, 2017; Kumaraguru *et al.*, 2010), phishing exercises and interventions significantly reduce the likelihood of clicking on malicious e-mails. The relationship is even more pronounced in more individualistic countries like the US (Rocha Flores *et al.*, 2015). Prior findings also indicate that whether employees receive little or much cybersecurity training depends on their employment status.

Generally, part-time employees have less experience with security practices. This is why they have different perceptions of vulnerabilities and behave differently than full-time employees (Anwar *et al.*, 2016).

Among the less studied task attributes that have a bearing on phishing susceptibility is job experience. In contrast to human capital theory (Becker, 1962) espousing that more experience leads to better performance, however, employees tend to pose a higher security risk the longer they work for a company (Sebescen and Vitak, 2017).

2.2 Social Context & Phishing Susceptibility

Johns (2006) identifies social structures as another essential set of contextual factors. These structures encompass the interactions between different organizational actors. So far, social context factors have not found much consideration in studies of phishing susceptibility. To the best of our knowledge, Wright et al. (2020) are among the first who explicitly incorporate social context factors into a multi-level model to explain phishing susceptibility. They show that individuals who frequently seek IT advice through official help desk channels are more vulnerable to clicking on a link embedded in a fraudulent e-mail.

Apart from advice and help networks, interactions between team members also affect their phishing vulnerability (Rajivan and Cooke, 2017) and serve as the foundation for taking appropriate actions (Salas *et al.*, 1995). Findings by Champion et al. (2012) suggest that collaboration and communication in teams foster cognitive processing and help increase security awareness. However, engagement in sharing knowledge and collaboration with colleagues is contingent on team size; members of larger teams tend to be less cooperative (Powers and Lehmann, 2017).

2.3 Informational Context & Phishing Susceptibility

The third context refers to contextual influences related to the individual workers. Informational context factors include attributes such as age or gender; both have been studied extensively by numerous phishing researchers, however with inconsistent results (e.g., Jagatic *et al.*, 2007; Li *et al.*, 2020; Sheng *et al.*, 2010). For instance, Jagatic et al. (2007) found younger users to be more susceptible to phishing. Three years later, Sheng et al. (2010) came to similar results showing that students aged 18 to 25 are more likely to click on phishing e-mails. In contrast, a recent survey among almost 7000 participants shows that older workers are most at risk of becoming victims of social engineering (Li *et al.*, 2020).

Prior investigations related to the impact of gender also point out that females and males differ regarding their likelihood of falling for phish. Some scholars like Sheng et al. (2010) show that females are more susceptible to phishing attacks, while Li et al. (2020) find males to be slightly more likely to click on deceptive e-mails.

3 Methodological Approach

Throughout this section, we provide an overview of the data collection procedure and the cluster analysis – a method used to identify patterns in large data sets (Halkidi *et al.*, 2001) and partition similar objects into homogeneous clusters (Ahmad and Dey, 2007). To collect the data needed, we performed a large-scale phishing experiment in an international pharmaceutical company. Our analysis focussed on employees working on three different continents – Australia, Europe, and North America. These countries differ with regard to their individualism score (Hofstede *et al.*, 2010), which has a bearing on individuals' phishing susceptibility (Butavicius *et al.*, 2017). In the course of three months, more than 17,000 employees received one of four different phishing e-mails. These four phishing e-mails were part of a company-wide awareness campaign and contained different phishing messages – ranging from generic to more targeted phishing e-mails. Following Hanus et al. (2021), we collapsed the victims of all four e-mails into one group labeled "phished". Subsequent to the awareness

campaign, we collected contextual data on all phishing e-mail recipients. The data provided by the pharmaceutical company were anonymized prior to being processed by the authors.

In the following, we will elaborate more on the four different phases required for clustering (see Figure 2). The first phase – feature selection – deals with the choice of distinguishing features (Xu and Wunsch, 2005). This is followed by selecting a clustering algorithm that will fit the data set. The third phase – cluster validation – refers to verifying the clustering algorithm results (Halkidi *et al.*, 2001). The final phase – cluster interpretation – is concerned with drawing meaningful insights from the data under examination (Xu and Wunsch, 2005).

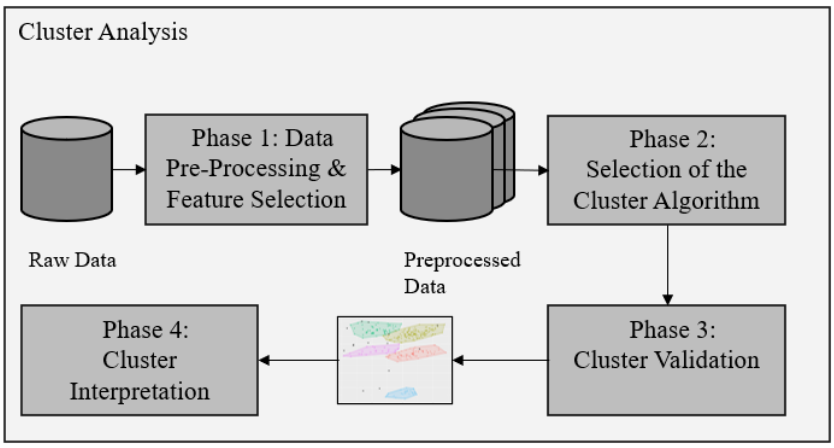

Figure 2. Four phases of the clustering process.

3.1 Phase 1 – Data Pre-Processing and Feature Selection

As stated above, we conducted a phishing experiment among several thousand employees and collected anonymized contextual information on all our study participants. The contextual information came from various internal resources, such as human resources databases and the learning management system, which affected the quality and completeness of the extracted data. Therefore, we used pre-processing techniques to improve our raw data efficiency and determine features suitable for further analysis (Alasadi and Bhaya, 2017). In a first step, for instance, we scanned the data set for missing values (Jadhav *et al.*, 2019) and found that information on age, gender, and training attributes was missing from the records of all external employees. Consequently, we excluded these datasets. In addition, we relied on clipping to remove outliers from the data set (Bagnall and Janacek, 2005). After having applied all pre-processing techniques, a total of 12,815 data sets (Australia: 1,247 (phished subjects: 313); Europe: 5,248 (phished subjects: 802); North America: 6,320 (phished subjects: 913)) remained for further analysis. Statistical analyses were conducted in R (R version 4.0.2).

Regarding the feature selection, we draw on Johns (2006) and Sarker (2016) and selected various discrete contextual variables, each of which had been proven respectively are assumed to influence phishing susceptibility. The data is mixed data containing numeric attributes like age or job experience and categorical attributes like gender and training compliance or training non-compliance. A complete list of the discrete context variables used for the cluster analysis can be found in the Appendix (Table 5). It should be noted that the selection is not exhaustive (Johns, 2006) but instead serves as a basis for a better understanding of which employee groups are universally at risk of getting phished.

3.2 Phase 2 – Selection of Clustering Algorithm

Clustering algorithms can be broadly categorized into either partitional or hierarchical algorithms (Jain and Dubes, 1988). Here, we focus on partitional clustering algorithms, in particular k-means. K-means is one of the most (cost)efficient clustering algorithms (Ahmad and Dey, 2007). The algorithm partitions data sets into k-clusters that are both as compact as possible and at the same time as distinct

as possible (MacQueen, 1967). Since some variables are continuous and others categorical, we used factorial analysis for mixed data (FAMD) to deconstruct the original data and retain only meaningful factors (Han *et al.*, 2021; Josse and Husson, 2016). The k-means cluster analysis was then performed based on the FAMD-transformed matrix. The parameters for maximum iterations and randomized initial configurations were set to 100.

3.3 Phase 3 – Cluster Validation

To determine the number of clusters, we used the SD Validity Index. The index considers both compactness and separation: compactness measures the similarity of objects within a cluster and separation measures the distinctiveness of objects in different clusters (Liu *et al.*, 2010). More importantly, the index determines the optimal number of clusters almost independently of the maximum number of clusters (Halkidi *et al.*, 2001). As shown in Table 1, the SD Validity Index suggests six clusters for Australia, four for Europe, and three for North America.

# of Clusters	Australia	Europe	North America
2	1.4753	1.6572	1.7775
3	1.2603	1.2547	**1.4821**
4	1.2418	**1.1428**	1.7428
5	1.1837	1.1801	1.6712
6	**1.1823**	1.3248	1.6630
7	1.2619	1.5489	1.7310
8	1.2052	1.4910	1.6746
9	1.2027	1.5508	1.7631
10	1.1876	1.5370	1.7832

Table 1. Partitions based on the SD Validity Index.

3.4 Phase 4 – Cluster Interpretation

For interpretation purposes, we compared each cluster of at-risk employees to the respective reference cluster (non-phished). Significances of each attribute were calculated using Mann-Whitney-U tests or Chi-Square tests. Effect sizes indicate the strength of the phenomenon. Please note that we only report significant attributes with small, medium, and strong effect sizes.

Table 2 presents the clusters generated based on the Australian sample set. In total, six unique clusters were generated. Cluster 1 includes individuals with an average age of 37.7 (SD:6.60). They have comparatively little job experience (4.61 years; SD:3.39), show little help desk reliance, and are counted among the low-wage earners (3.08; SD:0.92). Accordingly, we label this cluster "Unexperienced low-wage earners". Cluster 2 contains significantly more females (67.2%), all in full-time employment. They have little job experience (4.67 years; SD:4.24), but their security knowledge is relatively up to date. Given this, we term the cluster "Unexperienced but up-to-date security knowlegde". The third cluster, "Old low-wage earners", consists of mostly males (70.5%), old of age (52.7 years, SD:6.99) and with low help desk reliance. They work in large teams (11.4; SD:4.9) and earn less than the non-phished reference group (3.52; SD:1.02). Cluster 4 encompasses females (96.6%) with high job experience (11.2 years: SD:6.52). They are mostly working part-time (89.7%) and show high training compliance, meaning that the vast majority do the required security training in time. However, their security knowledge tends to be outdated. Accordingly, we label the cluster "Experienced part-timers". The fifth cluster is mostly superiors (63,2%) with medium job experience (6.12 years; SD:4.09) and medium income (5.32; SD:1,23). Interestingly, the superior's security knowledge seems to be outdated; their last security training was almost ten months ago. Given these descriptions, we label the cluster "Superiors with outdated security knowledge". The last cluster, "Old

high wage earners", encompasses the by far oldest individuals (53.6; SD:6.38). They are mostly males (65.8%) and have managerial responsibility except for one. Moreover, they occupy the highest pay grades (6.71; SD:1.71).

Australia				
#	**Cluster Size**	**Label**	**Significant Attributes**	**Effect Sizes**
1	87	Unexperienced low-wage earners	Male	+
			Age	-
			Job experience	--
			No managerial responsibilities	+
			Salary	---
			Help desk reliance	--
			Team size	+
2	58	Unexperienced but up-to-date security knowledge	Female	+
			Age	-
			Job experience	--
			Salary	-
			Security up-to-dateness	+
			Help desk reliance	+++
			Team size	-
3	43	Old low-wage earners	Age	+++
			Job experience	+++
			No managerial responsibilities	+
			Salary	--
			Help desk reliance	--
			Team size	+++
4	29	Experienced part-timers	Males	+
			Job experience	++
			Part-time	++
			Security up-to-dateness	-
			Training non-compliance	+
5	57	Superiors with outdated security knowledge	Job experience	-
			Managerial responsibilities	+
			Salary	++
			Security up-to-dateness	--
			Help desk reliance	-
			Team size	-
6	38	Old high-wage earners	Age	+++
			Job Experience	++
			Managerial responsibilities	++
			Salary	+++
			Help desk reliance	-
+/++/+++;-/--/--- small, medium, strong effect sizes				

Table 2. Australian clusters with significant attributes and effect sizes.

Table 3 presents the four clusters generated based on the European subsample. The first cluster contains younger employees (37.4 years; SD:11.0) with medium job experience (6.74; SD:6.10) and low incomes. Interestingly, the individuals have comparatively up-to-date security knowledge. Given this, we label the cluster "Young low-wage earners". Cluster 2 "Experienced part-timers" are mainly females (86.6%) with high job experience (17.3 years; SD:10.7), who work primarily part-time (81.3%). The members of this cluster tend to have a higher help desk reliance (11.9 tickets; SD:10.2), meaning that they regularly interact with official help channels. The third cluster contains individuals with relatively little job experience (4.56 years; SD:3.71), but medium income. They tend to have outdated security knowledge since the last training is way back. Accordingly, we label the cluster "Unexperienced but medium-income earners". Cluster 4, "Old high-wage earners", consist of individuals with the highest job experience (19.9 years; SD:12.0) and a higher help desk reliance. They are significantly older (52.2 years; SD:6.54) than the non-phished reference group and work in smaller teams (6.59; SD:3.42).

Europe				
#	**Cluster Size**	**Label**	**Attributes**	**Effect Sizes**
1	285	Young low-wage earners	Males	+
			Age	-
			Job experience	-
			Salary	--
			Security up-to-dateness	+
			Team size	-
2	134	Experienced part-timers	Females	+
			Age	++
			Job experience	++
			Part-time	+
			Salary	+
			Help desk reliance	++
			Team size	-
3	198	Unexperienced but medium-income earners	Age	-
			Job experience	--
			Salary	+++
			Security up-to-dateness	+
			Help desk reliance	+++
			Team size	--
4	185	Old high-wage earners	Age	+++
			Job experience	+++
			Managerial responsibilities	++
			Full-time	+
			Salary	+++
			Help desk reliance	++
			Team size	--
+/++/+++;-/--/--- small, medium, strong effect size				

Table 3. European clusters with significant attributes and effect sizes.

As depicted in Table 4, three at-risk clusters were generated for the North American subsample. The first cluster encompasses mainly males (57.9%) that are not only significantly older than the references group (52.5 years; SD:9.21) with plenty of job experience (9.4 years; SD:9.21) but also show a higher pay grade (7.35; SD:2.15). Moreover, they tend to let their security training slide and show a higher help desk reliance. Given the above, we label the cluster "Old high-wage earners". Cluster 2, labeled "Experienced superiors", contains mostly superiors (85.5%) with medium job experience (7.05 years; SD:6.05) and the highest help desk reliance (24.1 tickets; SD:9.01). Cluster 3 eventually includes mainly females (75.9%) with little job experience (3.53 years; SD:3.03) and no managerial responsibilities. In addition, they tend to work in smaller teams (9.62 persons; SD:4.58). Accordingly, we label this cluster "Unexperienced, with no managerial responsibilities".

North America				
#	**Cluster Size**	**Label**	**Attributes**	**Effect Sizes**
1	214	Old high-wage earners	Age	+++
			Job experience	++
			Managerial responsibilities	+
			Salary	+++
			Security up-to-dateness	-
			Training non-compliance	+
			Help desk reliance	++
			Team size	---
2	235	Experienced superiors	Job experience	+
			Managerial responsibilities	++
			Salary	+++
			Help desk reliance	+++
3	482	Unexperienced, with no managerial responsibilities	Job experience	-
			No managerial responsibilities	+
			Team size	-
+/++/+++; -/--/--- small, medium, strong effect size				

Table 4. North American clusters with significant attributes and effect sizes.

4 Discussion

The study's purpose was to uncover whether discrete context factors help to identify at-risk clusters of employees across three continents. We deployed cluster analysis coupled with an elaborate phishing study among employees of an internationally operating pharmaceutical company. The most salient cluster we detected across all three continents is labeled "Old high-wage earners" (North America: cluster 1; Australia: cluster 6; Europe: cluster 4). Members of this cluster were more likely to be older, with higher job experience and a higher salary. It seems that employees who have more job experience become negligent of phishing threats. This is consistent with studies proving that more experienced employees are not as vigilant anymore, which increases their likelihood of falling prey to cybercriminals (Sebescen and Vitak, 2017).

Furthermore, our results indicate that older workers seem to have difficulties detecting deceptive e-mail communication – an issue found in prior studies (see, for instance, Bullee *et al.*, 2017; Li *et al.*, 2020). The best explanation for this finding may be that older employees rely too heavily on their acquired skills and become inattentive when clicking on suspicious e-mails. Apart from this, the members of this cluster earn more than their peers in the non-phished reference group. Since salary is usually linked to job tasks and roles, we can assume that employees at higher pay grades have more

responsibilities and non-routine tasks, eventually resulting in higher workloads and stress levels (Sarno *et al.*, 2021; Sato *et al.*, 2009). Higher stress levels, in turn, increase the likelihood of failing to carefully scrutinize incoming mails (Greitzer *et al.*, 2021).

The "Experienced part-timers" (cluster 2 in Europe and cluster 4 in Australia) also show similarities. Members of these clusters were far more likely to be female, working part-time and showing higher job experience. These findings suggest that working part-time leaves employees with fewer possibilities to train their security knowledge. This corresponds with Li et al. (2020) and Greitzer et al. (2021), confirming that part-time employees are more susceptible to phishing. For the American subsample, we cannot identify a similar cluster which may be attributed to the small share of part-time workers in the sample. Another explanation for why female part-timer workers are vulnerable to phishing is their lack of computer and cyber security skills (Anwar *et al.*, 2019).

Clusters containing superiors (cluster 2 in North America and cluster 5 in Australia) also display similarities. They consist of individuals who work in full-time employment and show medium job experience. In addition, the American superiors show high help desk reliance, indicating that they may feel less in charge of taking care of IT or security issues. Another explanation is related to the concept of risk homeostasis, which states that people take more risks if they perceive measures to be in place to protect them from harm (Renaud and Warkentin, 2017). Following this perspective, individuals may see the help desk as their safety net leading to more risky behavior (Wright *et al.*, 2020). Noteworthy is that the Australian superiors show relatively outdated security knowledge – their last training was more than nine months ago. This is an issue because superiors and those who work higher job levels usually have access to sensitive data, which cybercriminals aim for (Vance *et al.*, 2018).

4.1 Implications for Theory and Practice

The present study offers contributions to both theory and practice. For academics, our research underscores the consideration of contextual determinants when investigating an individual's likelihood of falling for phish. Previous phishing studies have primarily focussed on intervention measures, individual characteristics, or technological means, thereby mostly disregarding the context (Wright *et al.*, 2020). Building on Johns (2006), however, the present work put contextual influences at the heart of its investigations, demonstrating that social, task, and informational context factors enable the identification of at-risk groups across continents, which is a new and significant finding. With this empirical evidence, we extend the importance of contextual factors in explaining security behaviors beyond what Johns (2006) and Wright et al. (2020) have already done.

Another contribution is related to the methodological approach. Following the call of researchers like Abbasi et al. (2016), we used predictive analytics to examine the interplay of contextual factors and phishing susceptibility. The identified clusters provide insight into cross-national similarities and differences in the characteristics of employees at risk of falling for fraudulent e-mails. Hence, our study underscores the importance of accessing knowledge in large data sets and profile users using clustering algorithms.

Consistent with previous work (D'Arcy and Hovav, 2009; Tambe Ebot, 2018, 2019; Tembe *et al.*, 2014), our results suggest that the likelihood of employees becoming phishing victims varies across continents. Therefore, a one-size-fits-all approach to anti-phishing is not a solution. At the same time, our findings allow managers to better target vulnerable groups of workers on different continents who share similarities. In addition, this approach facilitates cost-reduction of organizational phishing interventions. For instance, the cluster "Old high-wage earners" contains over-average old employees with high job experience and higher wages. Since the cluster is present in Europe and North America as well as Australia, it could be beneficial to design training sessions for elderly and experienced workers and consider that they might have different issues when it comes to information security. Furthermore, individualizing training programs may contribute to phishing resilience (Tambe Ebot, 2019).

Our results indicate that outdated security knowledge seems to be an issue for superiors. Like every other employee, they regularly have to do web-based security training – which is common practice in

modern organizations (Alshaikh *et al.*, 2018). However, as information security is constantly evolving, the training interval of one year may be too long. They may need more frequent reminders of the seriousness of phishing threats (Tambe Ebot, 2019). Other than that, our findings show that part-time workers are among the most vulnerable employees. Hence, we advise paying them particular attention when developing information security measures. In addition, since part-timers are working less than full-timers, it would help provide them more training to keep their knowledge updated. Added together, organizations that favour part-time over full-time contracts may have serious security challenges and need to implement costly prevention measures.

Another absorbing finding is that help desk reliance plays a significant role for phishing susceptibility across all continents. While Australians show below-average help desk reliance in most clusters, we observe a responsibility shift for Europeans and Americans: they tend to rely too heavily on official support channels. Hence, executives might want to promote help desk reliance in Australia, while they need to train Europeans and Americans to turn IT help desk advice into actionable knowledge.

A final point to mention is that processing e-mails and then identifying phish will remain in the hands of the employees. So, identification of at-risk groups in order to specifically train them makes economic sense, but at the same time organizations are well-advised to foster security awareness among all personnel and even encourage them to report suspicious e-mails to official help channels. In doing so, they may prevent others from falling prey to cybercriminals (Jaeger and Eckhardt, 2018).

4.2 Limitations and Future Research

As with any research endeavor, results need to be evaluated considering potential limitations. One of the issues that need to be addressed is the choice of contextual determinants. As stated above, the choice was not exhaustive (Johns, 2006). Therefore, it might be worthwhile also to consider physical context characteristics. Physical context may capture the device's characteristics used for e-mail communication (e.g., the installed software versions) or the users' primary worksite (office or home office). Previous work indicates that awareness campaigns are less influential on employees that spend more time outside the office (D'Arcy and Hovav, 2009). Hence, more research is needed to shed light on the influence of physical context on phishing susceptibility. Furthermore, it would be interesting to integrate prior phishing experiences. Recent findings indicate that those who fell for phishing once are more likely to become victims again (Greitzer *et al.*, 2021).

Another point to mention is that the present study concentrates solely on individualistic countries (Hofstede *et al.*, 2010). However, we find confirmation that phishing susceptibility varies across different cultures (Tembe *et al.*, 2014). Hence, we recommend considering collectivistic countries in future work and investigating whether segmentation across individualistic and collectivistic countries reveals similar at-risk groups. This may also help to move away from the one-size-fits-all approaches often used in awareness training, which have been shown to be less effective (Tambe Ebot, 2019). In a next step, researchers could also test whether gamified awareness programs, such as those proposed by Dincelli and Chengalur-Smith (2020), reduce the likelihood that the at-risk groups mentioned above will fall victim to phishing.

Our study relied on data collected in the course of a large-scale phishing campaign in a pharmaceutical company. Hence, differences between firms within a country, such as size or industry, are not considered here. Therefore, future research could rerun the study with different firms and in different industries, such as healthcare or energy, where the costs of a data breach are even higher than in pharma (IBM, 2020).

5 Conclusion

Modern organizations are at high risk of getting attacked by cybercriminals who target the weakest link: employees. Despite all efforts to train the workforce to combat security threats, social engineering is still one of the most prevalent threats. Recent work suggests that contextual factors help understanding who is likely to fall for phishing. This study, therefore, intended to segment at-risk

groups based on social, task, and informational context characteristics. With the help of more than 12,000 employees of a pharmaceutical company, we conducted a large-scale phishing experiment. The following cluster analysis revealed groups of individuals showing striking similarities across all or at least two continents. Our findings provide researchers and practitioners with new insights into individuals' susceptibility to phishing attacks and help to target similar at-risk groups individually with intervention programs.

References

Abbasi, A., Dobolyi, D., Vance, A. and Zahedi, F.M. (2021), "The Phishing Funnel Model: A Design Artifact to Predict User Susceptibility to Phishing Websites", *Information Systems Research*, Vol. 32 No. 2, pp. 410–436.

Abbasi, A., Sarker, S. and Chiang, R.H.L. (2016), "Big Data Research in Information Systems: Toward an Inclusive Research Agenda", *Journal of the Association for Information Systems*, Vol. 17 No. 2, pp. 1–32.

Abbasi, A., Zahedi, F.M. and Chen, Y. (2016), "Phishing Susceptibility: The Good, the Bad, and the Ugly", *IEEE International Conference on Intelligence and Security Informatics: Cybersecurity and Big Data, ISI 2016*, pp. 169–174.

Ahmad, A. and Dey, L. (2007), "A k-mean clustering algorithm for mixed numeric and categorical data", *Data & Knowledge Engineering*, Vol. 63, pp. 503–527.

Alasadi, S.A. and Bhaya, W.S. (2017), "Review of Data Preprocessing Techniques in Data Mining", *Journal of Engineering and Applied Sciences*, Vol. 12 No. 16, pp. 4102–4107.

Alshaikh, M., Maynard, S.B., Ahmad, A. and Chang, S. (2018), "An Exploratory Study of Current Information Security Training and Awareness Practices in Organizations", *51st Hawaii International Conference on System Sciences*, Vol. 9, pp. 5085–5094.

Alsharnouby, M., Alaca, F. and Chiasson, S. (2015), "Why phishing still works", *International Journal of Human-Computer Studies*, Vol. 82, pp. 69–82.

Alwanain, M.I. (2019), "Effects of user-awareness on the detection of phishing emails: A case study", *International Journal of Innovative Technology and Exploring Engineering*, Vol. 8 No. 4, pp. 480–484.

Anwar, M., He, W. and Ash. (2019), "Gender Difference and Employees ' Cybersecurity Behaviors", *Computers in Human Behavior*, Vol. 69, pp. 1–9.

Anwar, M., He, W. and Yuan, X. (2016), "Employment Status and Cybersecurity Behaviors", *IEEE International Conference on Behavioral, Economic and Socio-Cultural Computing*, pp. 1–2.

Avgerou, C. (2019), "Contextual explanation: Alternative approaches and persistent challenges", *MIS Quarterly*, Vol. 43 No. 3, pp. 977–1006.

Bagnall, A. and Janacek, G. (2005), "Clustering Time Series with Clipped Data", *Machine Learning*, Vol. 58, pp. 151–178.

Becker, G.S. (1962), "Investment in Human Capital: A Theoretical Analysis", *The Journal of Political Economy*, Vol. 70 No. 5, pp. 9–49.

Bullee, J.W., Montoya, L., Junger, M. and Hartel, P. (2017), "Spear phishing in organisations explained", *Information and Computer Security*, Vol. 25 No. 5, pp. 593–613.

Butavicius, M., Parsons, K., Pattinson, M., Mccormac, A., Calic, D. and Lillie, M. (2017), "Understanding Susceptibility to Phishing Emails : Assessing the Impact of Individual Differences and Culture", *Proceedings of the Eleventh International Symposium on Human Aspects of Information Security & Assurance*.

Champion, M.A., Rajivan, P., Cooke, N.J. and Jariwala, S. (2012), "Team-Based Cyber Defense Analysis", *2012 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support*, pp. 218–221.

Chen, R., Gaia, J. and Rao, H.R. (2020), "An examination of the effect of recent phishing encounters on phishing susceptibility", *Decision Support Systems*, Vol. 133, p. 113287.

Cheng, Z., Dimoka, A. and Pavlou, P.A. (2016), "Context may be King, but generalizability is the Emperor!", *Journal of Information Technology*, Vol. 31 No. 3, pp. 257–264.

Coronges, K., Dodge, R., Mukina, C., Radwick, Z., Shevchik, J. and Rovira, E. (2012), "The Influences of Social Networks on Phishing Vulnerability", *45th Hawaii International Conference on System Sciences*, pp. 2366–2373.

D'Arcy, J. and Hovav, A. (2009), "Does One Size Fit All? Examining the Differential Effects of IS Security Countermeasures", *Journal of Business Ethics*, Vol. 89, pp. 59–71.

Davison, R.M. and Martinsons, M.G. (2016), "Context is king! Considering particularism in research design and reporting", *Journal of Information Technology*, Vol. 31 No. 3, pp. 241–249.

Deloitte. (2015), *Job Architecture. Laying the Building Blocks of Effective Human Capital Management*, available at: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/human-capital/us-cons-job-architecture-041315.pdf.

Dincelli, E. and Chengalur-Smith, I. (2020), "Choose your own training adventure: designing a gamified SETA artefact for improving information security and privacy through interactive storytelling", *European Journal of Information Systems*, Taylor & Francis, Vol. 29 No. 6, pp. 669–687.

Dodge, R., Coronges, K. and Rovira, E. (2012), "Empirical benefits of training to phishing susceptibility", *IFIP Advances in Information and Communication Technology*, Vol. 376 AICT, pp. 457–464.

Goel, S., Williams, K. and Dincelli, E. (2017), "Got Phished? Internet Security and Human Vulnerability", *Journal of the Association for Information Systems*, Vol. 18 No. 1, pp. 22–44.

Greene, K.K., Steves, M.P., Theofanos, M.F. and Kostick, J. (2018), "User Context: An Explanatory Variable in Phishing Susceptibility", *Proceedings of the Network and Distributed Systems Security (NDSS) Symposium*, pp. 1–14.

Greitzer, F.L., Li, W., Laskey, K.B., Lee, J. and Purl, J. (2021), "Experimental Investigation of Technical and Human Factors Related to Phishing Susceptibility", *ACM Transactions on Social Computing*, Vol. 4 No. 2, pp. 1–48.

Group, T.A.-P.W. (2021), *Phishing Activity Trends Report*.

Halkidi, M., Batistakis, Y. and Vazirgiannis, M. (2001), "On Clustering Validation Techniques", *Journal of Intelligent Information Systems*, Vol. 17 No. 2–3, pp. 107–145.

Han, L., Shen, P., Yan, J., Huang, Y., Ba, X., Lin, W. and Wang, H. (2021), "Exploring the Clinical Characteristics of COVID-19 Clusters Identified Using Factor Analysis of Mixed Data-Based Cluster Analysis", *Frontiers in Medicine*, Vol. 8, pp. 1–16.

Hanus, B., Wu, Y.A. and Parrish, J. (2021), "Phish Me, Phish Me Not", *Journal of Computer Information Systems*, available at:https://doi.org/10.1080/08874417.2020.1858730.

Hofstede, G., Hofstede, G.J. and Minkov, M. (2010), *Cultures and Organizations - Software of the Mind: Intercultural Cooperation and Its Importance for Survival*, 3rd Editio., McGraw-Hill Education, available at: https://geerthofstede.com/hofstedes-globe/.

IBM. (2020), *Cost of a Data Breach Report*, Armonk, NY, available at: https://www.capita.com/sites/g/files/nginej291/files/2020-08/Ponemon-Global-Cost-of-Data-Breach-Study-2020.pdf.

Jadhav, A., Pramod, D. and Ramanathan, K. (2019), "Comparison of Performance of Data Imputation Methods for Numeric Dataset", *Applied Artificial Intelligence*, Taylor & Francis, Vol. 33 No. 10, pp. 913–933.

Jaeger, L. and Eckhardt, A. (2018), "When colleagues fail: Examining the role of information security awareness on extra-role security behaviors", *Proceedings of the 26th European Conference on Information Systems (ECIS) 2018*, No. 2015.

Jagatic, T.N., Johnson, N.A., Jakobsson, M. and Menczer, F. (2007), "Social phishing", *Communications of the ACM*, Vol. 50 No. 10, pp. 94–100.

Jain, A.K. and Dubes, R.C. (1988), *Algorithms for Clustering Data*, Prentice Hall, Englewood Cliffs, New Jersey.

Jensen, M.L., Dinger, M., Wright, R.T. and Thatcher, J.B. (2017), "Training to Mitigate Phishing

Attacks Using Mindfulness Techniques", *Journal of Management Information Systems*, Vol. 34 No. 2, pp. 597–626.

Johns, G. (2006), "The Essential Impact of Context on Organizational Behavior", *Academy of Management Review*, Vol. 31 No. 2, pp. 386–408.

Josse, J. and Husson, F. (2016), "missMDA: A Package for Handling Missing Values in Multivariate Data Analysis", *Journal of Statistical Software*, Vol. 70 No. 1, pp. 1–31.

Kim, B., Lee, D.-Y. and Kim, B. (2020), "Deterrent effects of punishment and training on insider security threats: a field experiment on phishing attacks", *Behaviour & Information Technology*, Taylor & Francis, Vol. 39 No. 11, pp. 1156–1175.

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L.F. and Hong, J. (2010), "Teaching Johnny Not to Fall for Phish", *ACM Transactions on Internet Technology*, Vol. 10 No. 2, pp. 1–31.

Li, W., Lee, J., Purl, J., Greitzer, F., Yousefi, B. and Laskey, K. (2020), "Experimental Investigation of Demographic Factors Related to Phishing Susceptibility", *Proceedings of the 53rd Hawaii International Conference on System Sciences*, available at:https://doi.org/10.24251/hicss.2020.274.

Liu, Y., Li, Z., Xiong, H., Gao, X. and Wu, J. (2010), "Understanding of Internal Clustering Validation Measures", *2010 IEEE International Conference on Data Mining*, pp. 911–916.

MacQueen, J. (1967), "Some Methods for Classification and Analysis of Multivariate Observations", *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297.

Moody, G.D., Galletta, D.F. and Dunn, B.K. (2017), "Which phish get caught? An exploratory study of individuals′ susceptibility to phishing", *European Journal of Information Systems*, Vol. 26 No. 6, pp. 564–584.

Nguyen, C., Jensen, M.L., Durcikova, A. and Wright, R.T. (2021), "A comparison of features in a crowdsourced phishing warning system", *Information Systems Journal*, Vol. 31 No. 3, pp. 473–513.

Powers, S.T. and Lehmann, L. (2017), "When is bigger better? The effects of group size on the evolution of helping behaviours", *Biological Reviews*, Vol. 92 No. 2, pp. 902–920.

Rajivan, P. and Cooke, N. (2017), "Impact of Team Collaboration on Cybersecurity Situational Awareness", in Liu, P., Jajodia, S. and Wang, C. (Eds.), *Cyber Situation Awareness*, 10030th ed., Springer, pp. 203–226.

Renaud, K. and Warkentin, M. (2017), "Risk Homeostasis in Information Security: Challenges in Confirming Existence and Verifying Impact", *Proceedings of the 2017 New Security Paradigms Workshop*, pp. 57–69.

Rocha Flores, W., Holm, H., Nohlberg, M. and Ekstedt, M. (2015), "Investigating personal determinants of phishing and the effect of national culture", *Information & Computer Security*, Vol. 23 No. 2, pp. 178–199.

Salas, E., Prince, C., Baker, D.P. and Shrestha, L. (1995), "Situation Awareness in Team Performance: Implications for Measurement and Training", *Human Factors*, Vol. 37 No. 1, pp. 123–136.

Sarker, S. (2016), "Building on Davison and Martinsons' concerns: a call for balance between contextual specificity and generality in IS research", *Journal of Information Technology*, Vol. 31 No. 3, pp. 250–253.

Sarno, D.M., Carolina, S. and Neider, M.B. (2021), "So Many Phish, So Little Time: Exploring Email Task Factors and Phishing Susceptibility", *Human Factors*, available at:https://doi.org/10.1177/0018720821999174.

Sarno, D.M., Lewis, J.E., Bohil, C.J. and Neider, M.B. (2020), "Which Phish Is on the Hook? Phishing Vulnerability for Older Versus Younger Adults", *Human Factors*, Vol. 62 No. 5, pp. 704–717.

Sato, Y., Miyake, H. and Thériault, G. (2009), "Overtime work and stress response in a group of Japanese workers", *Occupational Medicine*, Vol. 59 No. 1, pp. 14–19.

Sebescen, N. and Vitak, J. (2017), "Securing the Human: Employee Security Vulnerability Risk in Organizational Settings", *Journal of the American Society for Information Science and Technology*, Vol. 68 No. 9, pp. 2237–2247.

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L.F. and Downs, J. (2010), "Who falls for phish? A Demographic Analysis of Phishing Susceptibility and Effectiveness of Interventions", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 373–382.

Tambe Ebot, A. (2018), "Using stage theorizing to make anti-phishing recommendations more effective", *Information & Computer Security*, Vol. 26 No. 4, pp. 401–419.

Tambe Ebot, A. (2019), "How stage theorizing can improve recommendations against phishing attacks", *Information Technology & People*, Vol. 32 No. 4, pp. 828–857.

Tembe, R., Zielinska, O., Liu, Y., Hong, K.W., Murphy-Hill, E., Mayhorn, C. and Ge, X.I. (2014), "Phishing in international waters exploring cross-national differences in Phishing conceptualizations between Chinese, Indian and American samples", *Proceedings of the 2014 Symposium and Bootcamp on the Science of Security*, pp. 1–7.

Vance, A., Jenkins, J.L., Anderson, B.B., Bjornn, D.K. and Kirwan, C.B. (2018), "Tuning Out Security Warnings: A Longitudinal Examination of Habituation Through fMRI, Eye Tracking, and Field Experiments", *MIS Quarterly*, Vol. 42 No. 2, pp. 355–380.

Wright, R., Johnson, S. and Kitchens, B. (2020), *A Multi-Level Contextualized View of Phishing Susceptibility*, available at:https://doi.org/10.2139/ssrn.3622310.

Wright, R.T., Jensen, M.L., Thatcher, J.B., Dinger, M. and Marett, K. (2014), "Influence Techniques in Phishing Attacks: An Examination of Vulnerability and Resistance", *Information Systems Research*, Vol. 25 No. 2, pp. 385–400.

Xu, R. and Wunsch, D.C. (2005), "Survey of Clustering Algorithms", *IEEE Transactions on Neural Networks*, Vol. 16 No. 3, pp. 645–678.

Appendix

Context	Attribute	Variable Type
Informational Context	Age (in years)	Continuous
	Gender	Binary
Task Context	Job experience (in years)	Continuous
	Employment status	Binary
	Training compliance	Binary
	Up-to-dateness of security knowledge (in days since last security training)	Continuous
	Salary	Categorical
Social Context	Team size	Continuous
	Help desk reliance	Continuous

Table 5. Selected discrete context variables.