UNIVERSITÉ DU LUXEMBOURG

PhD-FHSE-2025-023

Faculty of Humanities, Education and Social Sciences

DISSERTATION

Defence held on 30 September 2025 in Esch-sur-Alzette

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG EN PSYCHOLOGIE

by

Xiaowei CHEN

Born on 8 March, 1987 in Jiangsu (China)

An Interdisciplinary Approach to Improve Security and Privacy Intervention Design: Motivation Theories, User Experience, and Field Experiments

Dissertation defence committee

Dr. Christine Schiltz, Dissertation Supervisor
Full Professor, Université du Luxembourg

Dr. Verena Distler, Vice Chairman
Assistant Professor, Aalto University

Dr. Gabriele Lenzini, Chairman
Associate Professor, Université du Luxembourg

Dr. Verena Zimmermann
Assistant Professor, ETH Zürich

Dr. Yixin Zou
Faculty Member, Max Planck Institute for Security and Privacy

Affidavit / Statement of originality

I declare that this thesis:

- is the result of my own work. Any contribution from any other party, and any use of generative artificial intelligence technologies have been duly cited and acknowledged;
- is not substantially the same as any other that I have submitted, and;
- is not being concurrently submitted for a degree, diploma or other qualification at the University of Luxembourg or any other University or similar institution except as specified in the text.

With my approval I furthermore confirm the following:

- I have adhered to the rules set out in the University of Luxembourg's Code of Conduct and the Doctoral Education Agreement (DEA)[1], in particular with regard to Research Integrity.
- I have documented all methods, data, and processes truthfully and fully.
- I have mentioned all the significant contributors to the work.
- I am aware that the work may be screened electronically for originality.

I acknowledge that if any issues are raised regarding good research practices based on the review of the thesis, the examination may be postponed pending the outcome of any investigation of such issues. If a degree was conferred, any such subsequently discovered issues may result in the cancellation of the degree.

Approved on 2025-07-08

[1] If applicable (DEA is compulsory since August 2020)

Abstract

We live in a society characterized by ubiquitous computing and pervasive digital services. Individuals who are unaware of the security and privacy risks that arise during interactions with digital technologies need to be informed of these risks. Therefore, security and privacy interventions remain necessary—not only to advise individuals of risks but also to empower them to address these risks. Nevertheless, many security and privacy interventions proposed by practitioners are perceived as neither engaging nor practical by their target users. This dissertation adopts an interdisciplinary approach to improve the design of security and privacy interventions. It addresses how autonomous motivation influences individuals' security-related behaviors and how motivation theories can be employed to guide the design of security and privacy interventions.

I address these objectives using a mixed-methods approach, including a systematic literature review, empirical data collection with focus groups, a user study with a qualitative survey, a mixed-design field experiment, and a longitudinal randomized controlled trial.

The first research objective is to examine how autonomous motivation influences individuals' security behaviors. We conducted a systematic literature review of relevant empirical studies in organizational contexts. By systematically analyzing the definitions, measurements, and referred theoretical frameworks, we identified 17 unique autonomous motivators and three types of related security behaviors. We not only developed a refined taxonomy of autonomous motivation related to security behaviors but also charted a path forward for conducting theory-informed research in human-centered security.

The second objective is to explore how motivation theories can be employed to design intervention programs for specific demographic groups. We first conducted two user studies: (a) seven focus groups in a workplace setting, and (b) a qualitative survey in family contexts. With insights from the focus groups and propositions from Self-Determination Theory, we developed group discussion and role-playing trainings for the organizational context. Combining findings from the qualitative survey and propositions from the Expectancy-Value framework, we created a short video intervention program for parents to empower them to support their children in addressing security and privacy concerns in family settings.

The third objective is to evaluate the effectiveness of proposed interventions in real-world settings. Specifically, we conducted a mixed-design experiment for the anti-phishing trainings, incorporating repeated measures across three time points and three in-situ phishing tests. We found that both trainings enhanced employees' anti-phishing self-efficacy and support-seeking intention in

within-group analyses. Only the role-playing training significantly improved support-seeking intention when compared to the control group. Participants in both trainings reported more phishing tests and demonstrated heightened vigilance to phishing attacks compared to the control group. To evaluate the short video intervention program, we used a 14-week longitudinal randomized controlled trial. We revealed that short videos enhanced parents' security awareness and their conversation strategies. Notably, parents who initially exhibited lower levels of these measurements benefited the most from the intervention. Moreover, short videos were effective in enhancing parents' self-efficacy in protecting their children from online risks.

Overall, this doctoral dissertation contributes to the field of human-centered security and privacy. The refined taxonomy of autonomous motivation facilitates future research to examine and develop human-centered security policies and interventions. The contextualization of the Expectancy-Value framework lays the foundation for future scholars who wish to further examine the framework in the security context. Our findings highlight the value of grounding intervention design in established theories to improve user acceptance and engagement. Further, the interventions proposed in this dissertation can be scaled up and further improved to enhance security and privacy for various demographic groups. The field experiments and proposed measurements in this dissertation are useful for future empirical investigations. Last but not least, this dissertation exemplifies and charts a path for conducting theory-informed research in human-centered security and privacy.

Pursuing a PhD is like running a marathon;
thank everyone who shielded me from the headwinds.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

We live in a society where digital devices and online services are both ubiquitous and pervasive. A quick inventory of popular internet-connected devices in a smart home illustrates this *ubiquity*: smartphones, smartwatches, laptops, robot vacuums, smart speakers, and internet televisions. Technology companies continuously collect, analyze, and store personal data through their services. These *pervasive* practices raise concerns about user privacy and data monetization, as personal information is increasingly treated as a tradable commodity within these digital ecosystems. When personal data is misused, it can pose various security and privacy (*S&P*) risks, including tailored advertisements to manipulate user attention, phishing emails to steal users' digital assets, and scams to deceive users for monetary gain. Therefore, individuals need to be informed about these risks and learn how to protect their personal data from exploitation. Moreover, they must be empowered to defend themselves against evolving cyber threats that target them directly.

Researchers have called for understanding the interplay between security measures and human factors [322], examining the behavioral, psychological, and societal aspects of S&P, and valuing users' positive contributions in the field of human-centered security and privacy. A few research streams have emerged in the past two decades, including (a) "making security and privacy usable," which focuses on designing more usable, secure, and accessible interfaces or mechanisms for S&P (e.g., [1, 105]); (b) "engaging users in S&P activities," which creates more engaging interventions to develop users' security skills and safe responses, such as identifying and reporting phishing emails [392, 55]; and (c) examining users' positive role in organizational and societal S&P, instead of viewing humans are the "weakest link," researchers explore how individuals can be empowered and contribute to S&P (e.g., [414, 304]). Research on *S&P intervention* has remained a consistent theme across these various research streams [133]. In this dissertation, I focus on a specific type of S&P intervention, i.e., *knowledge and skill-oriented training aimed at supporting users to protect themselves and others from security and privacy risks*. There are, of course, other types of interventions, such as informative interfaces to aid users in decision-making [416]; however, these are outside the scope of this dissertation.

Most individuals who are not specialized in computer science learn and update their S&P knowledge through informal sources. As a result, individuals can hold various misconceptions regarding the security and privacy aspects of the digital technologies with which they interact [173]. Thus, S&P interventions are still necessary to inform users of risks associated with technology usage and empower them to address these potential risks. However, S&P interventions are not always

embraced by their target users. Various reasons can lead to this mismatch between S&P practitioners and target users, for example, when the proposed training was neither practical nor engaging. In the following section, I briefly outline how motivation theories can guide S&P interventions for more engaging designs, discuss why User Experience should be anticipated, and highlight the challenges involved in evaluating the effectiveness of interventions.

1.1 Background

1.1.1 Guiding S&P intervention design with motivation theories

What drives a user to engage with an intervention? Researchers might respond with distinctive answers, depending on the theoretical lens that guided their intervention design. For instance, the Behavior Change Model (BCM) posits that behavior occurs when an individual's motivation, ability, and a trigger converge at a target behavior [126]; an intervention grounded in BCM, therefore, seeks to satisfy a wide range of user motivators, lower the ability requirements of interaction, or deliver triggers to create moments of high receptibility [91, 71]. By contrast, Protection Motivation Theory (*PMT*) [309] suggests that individuals act when they appraise a threat as both severe and relevant and believe that the prescribed action is both effective (response efficacy) and feasible for them (self-efficacy). Thus, an intervention grounded in PMT combines communicating threat severity and vulnerability while assuring users' confidence in, and perceived benefits of, the protective behavior [374, 416]. From the perspective of the Theory of Planned Behavior (*TPB*), an individual's S&P attitude, normative beliefs, and perceived behavioral control over performing an intervention influence their intention to adopt it [234]. Despite discrepancies in the drivers each theory emphasizes, most behavioral theories agree that human motivation is a core driver of behavior change. Therefore, motivation should be carefully considered during the design of interventions.

Self-Determination Theory (*SDT*) has been applied to improve interaction designs by many scholars in Human-Computer Interaction (HCI) [369]. SDT proposes that humans are "inherently curious, physically active, and deeply social beings. Individual human development is characterized by proactive engagement, assimilating information and behavioral regulations, and finding integration within social groups" [315, p.4]. Within the framework of SDT, *intrinsic motivation* refers to the interest, enjoyment, and satisfaction that arise from an individual's engagement with an activity [97] and *autonomous motivation* refers to engaging in an activity with volition, arising from one's sense of self and the well-internalized value and personal importance of the activity [352]. Integrating intrinsic motivation into the intervention design has been examined as an effective approach to engage users with S&P interventions. For example, Silic and Lowry [338] developed a gamified security training

that intentionally fostered joy and curiosity to engage employees with learning. In a six-month field study (N = 420) [338], they demonstrated that satisfying users' motivational and coping needs within a gamified system produced statistically significant improvements in security behavior. While an increasing number of studies have explored how intrinsic motivation can be applied to improve the design of S&P interventions, some works exhibit conceptual confusion between intrinsic motivation and autonomous motivation [280, 235]. In this dissertation, I further investigate how autonomous motivation influences individuals' security-related behaviors in organizations (see Chapter 2) and apply propositions grounded in SDT to design anti-phishing trainings in a workplace setting (see Chapter 4).

Security and privacy interventions can be considered as a form of education. Accordingly, I argue that theories from educational psychology offer a valuable lens for understanding users' engagement with interventions. Expectancy–Value Theory (*EVT*) [111], an influential theory in educational psychology, proposes that an individual's engagement in an activity is directly related to two factors: their expectation of how successfully they can perform the task and the subjective value they assign to it. While theories like TPB and PMT have already been commonly examined in security research [234], EVT has received little attention in this domain. To address this research gap, I investigate the motivating and discouraging factors that influence employees' engagement with phishing interventions through the lens of EVT (see Chapter 3). A core proposition of EVT asserts that increasing perceived task value while reducing perceived cost optimizes the benefit-to-cost ratio, thereby enhancing individuals' engagement with the activity. Building on this proposition, we design an intervention program targeting parents in family settings (see Chapter 5).

1.1.2 Examining User Experience to inform intervention design

Why is user experience vital to an intervention's success? Mandating individuals to attend training sessions may satisfy compliance rules, but this rarely produces the focused attention or behavior change that such interventions aim for. Instead, users judge an intervention through their perceptions of relevance, cognitive effort, and enjoyment; when the presented content is perceived as enjoyable, immersive, and satisfying, users are more likely to internalize and enact it afterward [190, 338]. By contrast, jargon-filled instructions and authoritative tones can lead users to ignore or even resist the S&P recommendations. In short, a positive user experience in S&P interventions engages users and can potentially drive positive security behavior change.

User Experience (UX) examines beyond the instrumental needs of technology and views individuals' interaction with them "as a subjective, situated, complex and dynamic encounter" [166, p.95]. The user experience of a service can be examined through its anticipated use or actual use

[381]. UX scholars [381] highlight a few key steps critical for designing a service, including surveying the needs of target users prior to design, examining users' subjective experience of a prototype, and iteratively pretesting until all critical issues are addressed. Among the commonly applied methods for UX research, focus groups offer rich, synergistic insights through group interaction yet can be subject to dominant voices and social-desirability effects [401, 402]. Qualitative surveys are praised for their scalability and cost-effectiveness, but criticized for potential bias and limited depth [230]. Thus, researchers need to weigh these advantages and limitations when they select UX methods for specific cases.

Studies on UX and Meaning Making of ubiquitous computing were highlighted as the third wave in HCI development by Bødker [44]. Mekler and Hornbæk [256] proposed a framework of meaning in interaction, in which they highlighted that the experience of meaning *connects beyond the immediate experience*, *has a sense of direction*, and *views life as a whole*, denoting both *the immediate, unreflected experience* and *enduring value and importance*. Further, Renaud and Flowerday [304] observed a similar occurrence of UX and Meaning Making studies in the human-centered security and privacy community, which explored the influence of context, situational impact, or Meaning Making in S&P topics. I follow this line of research [101, 105] and consider UX essential for creating engaging S&P interventions. A good S&P intervention should carefully consider the social context and physical environment of users, and it should encompass both immediate utility values and the creation of meaningful experiences for target users. To achieve these, I conduct focus groups (see Chapter 3) and qualitative surveys (see Chapter 5) to inquire into target users' experiences with previous interventions and expectations for future interventions.

1.1.3 Evaluating intervention in the field poses challenges

How can the effectiveness of interventions be evaluated? Studies evaluating the effectiveness of S&P interventions can be categorized into three types, differentiated by the timing of data collection [189]: interaction assessment, post-intervention assessment, and pre- and post-intervention assessment. Interaction assessment focuses on studying how well users can interact with an intervention during the process [381]. Logs [66], task performance [118], and observations of users' interactions [102] are frequently analyzed for this purpose. For example, Hart et al. [165] examined the ease of understanding the mechanics of a serious game for raising cybersecurity awareness. Post-intervention assessment collects users' performance and feedback after the deployment of an intervention. Researchers commonly adopt a between-subjects design to compare training effects with alternative interventions, e.g., [397]. For pre- and post-intervention comparison, it can reveal the difference between users' performance before and after the intervention. Researchers use within-subject experiments to compare

metrics before and after an intervention [65], e.g., [354]. The majority of S&P intervention studies conduct their assessments in controlled environments; very few studies [38, 302, 340] have chosen to measure training effects over time in the field.

Scholars have argued that field experiments offer higher ecological validity than those conducted in controlled laboratory settings [411]. The observed behaviors from field experiments are more likely to reflect how users act in the real world. This is particularly relevant for S&P research [102, 28], where individuals make decisions amid everyday distractions and trade-offs. As a result, findings from controlled environments may not accurately reflect individuals' real-world S&P decision-making. Considering that the ultimate goal of interventions is to facilitate the transfer of learned knowledge or practices into security and S&P actions, without deployment in natural settings, researchers can only assume that target users would behave similarly to how they perform in the laboratory. Lastly, even when an intervention is well-designed and effective, individuals often require time to internalize new knowledge, adjust existing habits, and integrate new behaviors into their routines. Field experiments allow researchers to observe users' adaptation to new practices over time and to assess whether behavior changes are sustained longitudinally [38].

Nevertheless, evaluating S&P interventions in the field presents several challenges. Ethically, researchers must balance the need for ecological validity with participant protection, particularly when the study design may cause negative impacts [392, 50]. From a privacy standpoint, behavioral data collected in the real world raises concerns about re-identification and long-term data stewardship [104, 249]. Technically, field deployments demand robust and context-resilient designs capable of functioning across unpredictable user environments, often requiring sophisticated infrastructure [66]. Legal and regulatory constraints further complicate field experimentation, as researchers must comply with data protection frameworks such as GDPR and institutional regulations, and these require well-designed data processing plans. Logistically, field studies require sustained participant engagement and coordination with multiple stakeholders [308], all of which can be hindered by dropout, invalid input, or external disruptions. Addressing these challenges requires carefully designed and ethically sound approaches to ensure meaningful and responsible evaluation of S&P interventions in the field.

Following the background, I present the overarching research objectives of this dissertation (Section 1.2) and provide an overview of its structure (Section 1.3). Subsequently, I outline the publications associated with this dissertation (Section 1.4).

1.2 Overarching research objectives

This dissertation investigates approaches to improve the design and evaluation of security and privacy interventions. It addresses two central questions: (1) how autonomous motivation influences individuals' security-related behaviors, and (2) how motivation theories can inform the design of effective security and privacy interventions. Further, to empirically evaluate the effectiveness of the proposed interventions, we conduct field experiments comparing the effects of the intervention programs with control groups.

The following high-level research objectives guide the design and implementation of this dissertation:

- **Objective 1** is to examine how autonomous motivation influences individuals' security behaviors. To this end, we conduct a systematic literature review of relevant empirical studies in organizational contexts.
- **Objective 2** is to explore how motivation theories can be applied to design training programs for specific demographic groups. We first conduct two user studies: (a) seven focus groups in a workplace setting, and (b) a qualitative survey in family contexts. We then develop two security training programs (group discussion and role-playing as hackers) and one security and privacy intervention program (a series of six short videos).
- **Objective 3** is to evaluate the effectiveness of proposed interventions in real-world settings. Specifically, we conduct a mixed-design experiment for the security training programs, incorporating repeated measures across three time points and three in-situ phishing simulations. For the short video intervention program, we employ a 14-week longitudinal randomized controlled trial.

1.3 Structure of the dissertation

Chapter 1 presented a concise overview of existing work on the design and evaluation of security and privacy interventions and outlined the overarching research objectives of this dissertation. In this subsection, I will outline the dissertation's structure, followed by a list of the associated publications.

In Chapter 2, we systematically review 45 empirical studies examining autonomous motivation in organizational security contexts. Recent studies suggest that autonomous motivators hold untapped potential in promoting security behaviors without relying on controlled motivation alone. However, prior work has used a variety of theoretical frameworks from various disciplines to study autonomous motivators, leading to fragmented and heterogeneous literature. To reconcile fragmented findings, we

systematically analyze the measurements, definitions, and references of reviewed papers. We propose a refined taxonomy of autonomous motivation related to organizational security behaviors and chart a path for conducting theory-informed studies on autonomous motivation in human-centered security.

In Chapter 3, we explore the motivating and discouraging factors that influence employees' engagement with phishing interventions. Through seven focus groups with 34 employees at a Western European university, we inquire about their perceived benefits/costs, goal setting, confidence, identity, and discouraging factors in the context of phishing interventions. These questions were adapted from the core concepts of EVT that affect an individual's choices and performance. We reveal a spectrum of factors that influence employees' intentions to report phishing emails and engage with phishing awareness campaigns. Additionally, we propose to include new concepts—such as the behaviors of colleagues and supervisors, and organizational culture—into the EVT framework. Further, leveraging insights from both the employee-generated suggestions and the EVT framework, we propose several improvements to phishing interventions at organizations.

In Chapter 4, we design two phishing training approaches ("group discussion" and "role-playing as hackers") to foster social interaction among employees—supporting their *psychological need for relatedness*—and incorporated "interest" and "fun" elements into the role-playing training. Using a mixed-design field experiment with 105 university employees, we evaluate the effectiveness of these two trainings by comparing them with a control group. Our findings reveal that both trainings were effective in enhancing perceived self-efficacy and support-seeking intention in the Day 7 assessment. However, only role-playing significantly enhanced support-seeking intention compared to the control group. Both trainings contribute to an increase in reporting simulated phishing emails and safe responses to phishing emails. Our study demonstrates the feasibility of obtaining informed consent from research participants for simulated phishing tests while still gaining valuable insights from the results.

In Chapter 5, to examine our theory-informed intervention in a broader context, we investigate the longitudinal development of parental competence and coping strategies in S&P topics. Parents often rely on online resources to support and guide their children in S&P related topics. However, the abundance of online resources makes it challenging for parents to find high-quality and relevant resources that align with their S&P needs. Through a formative study with 210 U.S. parents, we investigate the challenges parents face in educating children about online S&P topics and inform the design of a remote intervention program (six short videos). Short videos were selected as the intervention format to provide diverse content in a condensed and flexible manner for parents at minimal cost. In the main study, we evaluated this intervention's efficacy using a 14-week longitudinal randomized controlled trial, which consisted of 201 U.S. parents, with 113 assigned to the control

group and 88 to the intervention group. This study provides valuable insights into various challenges parents face and respective coping strategies that could be implemented to address S&P concerns in family settings. The design and evaluation of the intervention program serve as a foundation for future S&P researchers and educational stakeholders.

Finally, Chapter 6 concludes the dissertation by reflecting on the key findings and contributions, and by outlining potential avenues for future research.

1.4 Associated publications

Three chapters have been published in the top computer science conference (ACM CHI) and a premium conference on usable security and privacy (SOUPS), while Chapter 5 will be presented at a top security conference (ACM CCS). See Table 1.1 for an overview of the chapters. All papers have been reformatted to fit the structure and requirements of this dissertation.

Table 1.1: Dissertation Chapters and their associated publications.

Chapter	**Associated Publication**	**Publication Status**
Chapter 2	Chen, X., Schöni, L., Distler, V., & Zimmermann, V. (2025). Beyond Deterrence: A Systematic Review of the Role of Autonomous Motivation in Organizational Security Behavior Studies. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1–28).	Published *Honorable Mention Award*
Chapter 3	Chen, X., Doublet, S., Sergeeva, A., Lenzini, G., Koenig, V., & Distler, V. (2024). What Motivates and Discourages Employees in Phishing Interventions: An Exploration of Expectancy-Value Theory. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)* (pp. 487–506).	Published
Chapter 4	Chen, X., Sacré, M., Lenzini, G., Greiff, S., Distler, V., & Sergeeva, A. (2024). The Effects of Group Discussion and Role-playing Training on Self-efficacy, Support-seeking, and Reporting Phishing Emails: Evidence from a Mixed-design Experiment. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1–21).	Published
Chapter 5	Chen, X., Distler, V., Gordon, C., Yao, Y., & Teuber, Z. (2025). Empowering Parents to Support Children's Online Security and Privacy: Findings from a Randomized Controlled Trial. In *Proceedings of ACM Conference on Computer and Communications Security (CCS '25)* (pp. 1–15).	Accepted

There are four related publications that I completed during my PhD training. These are not included in this dissertation (* indicates shared first authorship):

- Chen, X., Hedman, A., Distler, V., & Koenig, V. (2023). Do Persuasive Designs Make

Smartphones More Addictive?-A Mixed-methods Study on Chinese University Students. *Computers in Human Behavior Reports*, 10, 100299.

- Chen, X., Doublet, S., Distler, V. (2024). Making Motivation Theories Accessible: Introducing Motivation Cards to Map Motivators for Security and Privacy Education. In *S&PEI Workshop of the Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*.

- Wang, Z., Wu, Y., Yang, S., Chen, X., Rohles, B., & Fjeld, M. (2024). Exploring Intended Functions of Indoor Flying Robots Interacting With Humans in Proximity. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-16).

- Metreveli, A.*, Chen, X.*, Hedman, A., & Sergeeva, A. (2025). "Who Will be Left Behind?": A Swedish Case of Learning AI in Vocational Education. *International Journal of Educational Research*, 133, 102697.

Chapter 2

A Systematic Review of the Role of Autonomous Motivation in Organizational Security Behavior Studies

Published as: Chen, X., Schöni, L., Distler, V., & Zimmermann, V. (2025). Beyond Deterrence: A Systematic Review of the Role of Autonomous Motivation in Organizational Security Behavior Studies. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1–28).

Abstract What drives employees to ensure security when handling information assets in organizations? There is growing interest from the security behavior community in how autonomous motivators shape employees' security-related behaviors. To reconcile the scattered viewpoints on *autonomous motivation* and synthesize findings from studies utilizing various theoretical frameworks, we systematically reviewed relevant publications. We present a preregistered literature review that investigated (a) what forms of autonomous motivation have been examined in organizational security contexts, (b) which behaviors/behavioral intentions are related to autonomous motivators, and (c) how autonomous motivation affects employees' security behaviors. Based on an initial set of 432 papers, filtered down to 45 studies, we identified 17 unique autonomous motivators and three types of related security behaviors. This review not only develops a refined taxonomy of autonomous motivation related to security behaviors but also charts a path forward for future research on autonomous motivation in human-centered security.

2.1 Introduction

Organizations face the critical challenge of securing their information systems against cyber threats that target humans. Various lenses can be applied to improve our understanding of why people behave the way they do. Prior work has highlighted important obstacles to information security such as prescribed security policies not being feasible and too cumbersome [193], inappropriate cost-benefit trade-offs of complex security advice [175], or situational factors making secure responses challenging [102]. Recent conceptualizations of security behavior change [321, 176] highlight that security behaviors need to be met with low friction and to be within the compliance budget. People need to understand the behaviors, agree that these behaviors matter, believe that they are able to implement them, acquire the skills to implement the behavior, and then embed them in their everyday life. Hence, promoting secure behaviors in an organization is a challenging, multi-step process, involving various stakeholders

[107].

The majority of security behavior studies focused on the roles of deterrence and sanctions in guiding employees' security behaviors [92, 13]. A recent review [92] found that out of 49 studies concerned with cybersecurity behavior, 36 measured compliance exclusively. While compliance is crucial for maintaining organizational information security, attacks are becoming increasingly sophisticated. Employees need to flexibly cope with new threats that are not yet prescribed by existing information security policies (*ISP*s), thus going beyond compliance when necessary. Recently, there has been a shift of focus from "threats and sanctions" approaches, which do not always lead to the desired security behaviors [58], toward exploring employees' motivation to engage in "self-driven" protective security behaviors [257, 70, 131]. Correspondingly, a growing number of recent publications have investigated how employees' autonomous motivation shapes their security behaviors [277, 92, 186].

"What is autonomous motivation?" In the framework of Self-Determination Theory (*SDT*), human motivation can be categorized into three types [97, 316]: *amotivation* (lack of intention and motivation), *controlled motivation* (driven by external reward/punishment and pressure), and *autonomous motivation*. *Autonomous motivation* comprises both *intrinsic motivation* (including interest, enjoyment, and satisfaction) and the internalized extrinsic motivation in which people have identified with an activity's value and integrated it into their sense of self (including values, commitment, and ethics) [97]. Satisfying an individual's psychological needs for competence, autonomy, and relatedness creates autonomous motivation [352]. When performing tasks autonomously, employees experience a sense of choice and enjoyment; they do not feel compelled by outside forces [98]. Autonomous motivation is enduring and sustainable in driving employee performance [352].

Security-related research shows that autonomous motivators are positively correlated with certain security behaviors in the workplace [344, 131] and might make compelling contributions to explain and foster employees' protective cybersecurity behaviors [13, 257]. Despite this potential, autonomous motivation is a construct derived from psychology which necessitates adaptation and empirical validation in the context of security behaviors. It is worth noting that in a widely cited taxonomy of security behavior grounded in SDT [280], autonomous motivators have been categorized as intrinsic motivation. Moreover, researchers have applied constructs from other theories to investigate autonomous motivators, frequently without clarifying their theoretical foundation [332, 204] or the definition of "autonomous/intrinsic" [344, 235]. These inconsistencies in the conceptualization and application of autonomous motivation (e.g.,[280, 204]) pose challenges for future research, particularly in terms of synthesizing prior findings from studies that examined autonomous motivation in relation to security behaviors.

To consolidate the theoretical foundation and reconcile scattered findings on autonomous motivation, we conducted a preregistered literature review using the Scopus and ACM Digital Library databases to investigate (a) what forms of autonomous motivation have been examined in organizational security contexts, (b) the behaviors/behavioral intentions that are related to autonomous motivators, and (c) how autonomous motivation affects security behaviors in the workplace. The contributions of our review are three-fold:

- We developed a refined taxonomy of autonomous motivation including 17 motivators clustered into five categories by reviewing theoretical frameworks and comparing measurements of autonomous motivators in the security domain. The taxonomy and the accompanying toolbox of existing measurements provide relevant and timely support for researchers and practitioners who aim to examine and develop user-centered security policies and interventions.

- While previous studies aiming to explain security behavior understood as compliance often made use of theories focusing on *threats* and *deterrence*, our review suggests that there is a shift towards Self-Determination Theory as the most frequently applied theory in studies that focused on what *motivates* employees to ensure security. This shift in theory mirrors a paradigm shift in the security domain from viewing the human as the weakest link towards viewing the human as a valuable resource that can be enabled and motivated to contribute to security.

- We provide an overview of suitable avenues to extend the study of autonomous motivation in the domain of organizational information security and provide practical suggestions for researchers who want to conduct theory-informed studies on autonomous motivation in human-centered security.

2.2 Background

2.2.1 Motivation in security behavior studies

The interdependent relationship and distinctive contributions of emotion, cognition, and motivation to human behavior have sparked extensive discussions in psychology [232]. In the security context, emotions influence the degree of attention individuals direct toward cybersecurity tasks and their adherence to guidelines [385]. Cognitive processes such as perception, attention, and elaboration affect how a person interprets and responds to cyber attacks [56]. Motivation is critical in both initiating and maintaining behaviors [228]. Whereas initial changes are often driven by the expectation of long-term benefits, maintaining these changes relies more on the regular satisfaction derived from the behavior itself [228]. While we acknowledge the critical roles of emotion, cognition, and other

factors in behavior change, this review narrows its scope to autonomous motivation.

A prevalent approach for studying employees' security behaviors involves deductive methodology [233], where motivation is often included as one of the independent variables, and security behaviors are the dependent variables in research models. In this approach, researchers examine motivational factors alongside other variables to explain or influence employees' compliance intentions [170, 236]. Established theories from other disciplines have been frequently introduced into security behavior research [233] to explain the relationships between motivational factors and behaviors. Theory-based literature reviews [234, 17] indicate that the Theory of Planned Behavior [343], Protection Motivation Theory [158], and Deterrence Theory [226] have been the most frequently utilized to examine employees' security behaviors. Below, we describe how each of these theories conceptualizes the role of motivation in changing security behaviors.

The Theory of Planned Behavior proposes that an individual's behavioral intentions are determined by their *attitudes toward the behavior*, *subjective norm* (e.g., perception of others' expectations), and *perceived behavioral control* (e.g., one's ability to perform the behavior) [4]. Individuals' beliefs in their ability to perform a behavior are crucial in determining their choice of action [264]. Thus, an individual's motivation to perform a security behavior is influenced by their attitude and social influences, but their ability (perceived behavioral control) determines whether they can successfully carry it out [264]. Though the Theory of Planned Behavior has frequently been applied to examine employees' compliance behaviors, researchers have raised concerns about missing variables in the framework [343]. Kranz and Haeussinger [218] proposed integrating the Theory of Planned Behavior and the Organismic Integration Theory — a subtheory of SDT. They empirically tested their research model in a sample of 444 employees [218]. They found that when employees' personal values and principles aligned with their employer's information security prescriptions and goals, employees' intentions to comply increased significantly [218]. This finding suggests that autonomous motivation is a relevant factor in fostering information security compliance.

Protection Motivation Theory, initially developed in the health management domain [309], posits that individuals' protective behaviors are influenced by their evaluations of the *severity* and *certainty* of a threat and assessments of the efficacy and their ability to perform protective behavior. In the framework of Protection Motivation Theory, situations involving threats (e.g., health, intrapersonal and interpersonal, economic [158]) motivate people to choose protective solutions. In the context of information security, threats represent events with potentially harmful consequences [158]. A review of 67 studies applying Protection Motivation Theory showed that most studies examined threats and coping appraisal constructs, specifically *self-efficacy* (91.0%), severity (89.6%), *vulnerability* (88.1%), and *response efficacy* (83.6%) [158]. Menard et al. [257] argued that if a threat is perceived

as irrelevant, the appeal will not evoke fear and will fail to connect with the individual. After comparing three competing research models in their study [257], they suggested that intrinsic motivators could be a powerful factor that influences organizational security behaviors.

Deterrence Theory, rooted in criminology, proposes that "certain controls can serve as deterrent mechanisms by increasing the perceived threat of punishment for information system misuse" [85, p.1]. It has frequently been applied to understand employees' ISP compliance and violation behaviors [353]. According to Deterrence Theory, motivation is linked to an individual's perception of the certainty, swiftness, and severity of the sanctions that may result from their actions [353]. In a survey study of 602 U.S. employees, Son compared the intrinsic and extrinsic motivation models integrated into the framework of Deterrence Theory [344]. Their study revealed that "intrinsic" motivators (*perceived legitimacy* and *value congruence*) played a more substantial role in explaining employee compliance than extrinsic motivations (*perceived deterrent certainty* and *severity*) [344]. Son [344] suggested that exploring intrinsic motivation-based approaches, rather than only sanction-based methods, is likely to enhance employees' compliance with ISPs.

To summarize, research involving the most frequently used theories suggested that autonomous motivation provides an alternative approach for explaining and fostering security behaviors. However, systematic examinations of how autonomous motivation is related to organizational security behaviors are lacking. This research gap thus led us to our first research question:

> **RQ1**. What forms of autonomous motivation have been found to influence employees' information security behaviors?

2.2.2 Paradigm shift from compliance to extra-role security behavior

Maintaining information security in organizations is not only a technical challenge but also needs to consider human actions [243]. The complexity of human behavior positions individuals as key influencers in securing information systems, and research on behavioral factors in cybersecurity is critical [243]. Lahcen et al. [243] analyzed *insider threats* (risks caused by an employee or any other individual with authorized access to the information system) in the workplace and categorized them into three types: unintentional, intentional, or malicious. Unintentional errors are caused by a lack of knowledge or skill (e.g., accessing confidential information through public Wi-Fi), while intentional errors arise from knowingly risky behavior (e.g., leaving passwords on a sticky note), and malicious actions are deliberate with the intent to cause significant harm (e.g., stealing confidential data) [243]. Most organizations have an information security policy (ISP) that describes insiders' responsibilities and prescribes actions to protect the organization [344].

Different taxonomies have been proposed for understanding employees' *security compliance*,

focusing on factors such as intentionality, technical expertise, and intrinsic/extrinsic motivation [348, 280]. Via interviews (n = 110) and surveys (n = 1167), Stanton et al. [348] categorized employees' behaviors into six categories with two parameters: intentionality (including malicious, neutral, and benevolent intentions) and technical expertise (including novice and expert). They suggested that employees might exhibit behaviors from different categories at different points in time [348]. Organizations can enhance their security status through motivational interventions that promote benevolent intent among employees [348], including effective security management, strong leadership, clear role designations, and training programs. In another taxonomy, Padayachee [280] categorized research findings on security compliance and deterrent control into a taxonomy predicated on SDT. They linked motivational factors with security-compliant behaviors and suggested that organizations apply the framework to understand employees' motivations, thereby assessing and promoting security compliance [280].

Recent research has highlighted a proactive approach towards employees that extends beyond mere compliance with ISP guidelines. Posey et al. [287] proposed a taxonomy of protection-motivated behaviors that indicates which protection-motivated behaviors are critical, which are difficult to promote, and which are considered common sense, allowing for direct comparisons across individual behaviors. Posey et al. [289] suggested that employees can be guardians of organizational information security. This is aligned with a recent interview study where cybersecurity professionals posited that empowering employees, rather than inhibiting their behaviors, can enable them as the last line of defense in organizations [364]. An increasing number of security behavior studies have examined employees' extra-role security behaviors [131, 55], beyond mere compliance. These studies investigated a wide range of security-related behaviors and tasks in the workplace, such as employees' self-driven security literacy learning [338], crowdsourced approaches to defending against phishing attacks [55], and security knowledge sharing in the workplace [318].

Understanding what drives employees to perform these *extra-role security* tasks can enhance organizational information security [93]. Extra-role information security tasks describe "security-related citizenship behaviors that go beyond prescription but nonetheless contribute to the organisational, social, and psychological InfoSec environment" [92, p.198]. In a focus group study, Chen et al. [70] found that various intrinsic factors influenced employees' intentions to report phishing emails. These motivators, deeply embedded in employees' psychological needs, include enjoyment, satisfaction, empowerment, and sense of belonging [70]. Recently, Frank and Kohn [131] proposed a taxonomy of motivation for extra-role security behaviors (*SDT-ER taxonomy*), based on SDT, in which motivators of extra-role behaviors are arranged along a continuum from extrinsic to intrinsic motivation. They [131] linked six out of nine dimensions of extra-role security behaviors with

autonomous motivators (including usefulness-driven, value-driven, and interest-driven). However, further research is needed to investigate other types of security behaviors related to autonomous motivation:

> **RQ2**. Which employee security behaviors (or behavioral intentions) related to autonomous motivation have been examined in the surveyed literature?

2.2.3 Fragmentation and heterogeneity of applying diverse theoretical frameworks

Sutton and Staw [356] discussed how references, data, variables, diagrams, and hypotheses can be mistaken for theory and shared their definition of *theory*:

> *Theory is about the connections among phenomena, a story about why acts, events, structure, and thoughts occur. Theory emphasizes the nature of causal relationships, identifying what comes first as well as the timing of such events.* [356, p.378]

In this review, we adopted Sutton and Staw's definition and emphasize the three characteristics of theory: (a) It consists of interrelated propositions and constructs; (b) it establishes clear relationships between constructs; and (c) it explains or predicts the occurrence of events [192]. We refer to a *theoretical framework* as the application of a theory or set of interrelated constructs derived from an established theory to guide the research [192]. By contrast, a *research model* involves synthesizing concepts, ideas, and constructs from multiple sources, including empirical findings and different theories, to create a unique model to address research problems [192]. We adopted Liehr and Smith's definition of *concept*: "an image or symbolic representation of an abstract idea" [192, p.188]. We conceptualize a *construct* as "a label for a cluster or domain of covarying behaviours" [41, para.2]. Constructs are key components of theories [371].

In the security behavior community, researchers have argued that theory is essential for the field [100]. Theory provides a structure for understanding complex behaviors and their underlying motivations [257] and for identifying intangible psychological factors that influence users' security-related choices. Moreover, theory guides the development of measurements [117] and interventions [338, 416], thus enhancing the rigor and validity of research outputs. For example, Faklaris et al. [117] created and empirically validated the Security Attitudes (SA-6) measurement with the guidance of the Theory of Reasoned Action. Zou et al. [416] developed and evaluated the effectiveness of two interventions grounded in Protection Motivation Theory, and Silic and Lowry [338] integrated intrinsic motivation into their security training to create an immersive learning experience for employees. Furthermore, a strong theoretical foundation helps to avoid the pitfalls of an "atheoretical black box" [100], ensuring that research contributions are meaningful. Given the

interdisciplinary nature of security behavior research [209], theories from *psychology, criminology*, and *organizational behavior* have been introduced to study security behaviors [234]. This integration of various theories has facilitated the exploration of diverse factors that influence security behaviors [264, 84], ranging from fear, desire, and self-efficacy to organization culture and societal influences, which offer numerous insights for understanding and promoting security behaviors [234]. However, this blossoming of adopted theories has also presented new challenges and raised new questions for researchers. For example, "How can we synthesize findings from varied and even competing theoretical frameworks?" This disparity and fragmentation of knowledge requires ongoing effort from researchers to synthesize findings from studies that have utilized different theoretical frameworks.

There are two common approaches for integrating existing findings from different theories, namely, theory-driven and empirical-data-driven approaches [264]. The theory-driven approach [234] can highlight important factors that are not visible in the empirical data of a specific case, whereas the empirical-data approach can find relevant constructs from a phenomenon and compare constructs from different theories [264]. Lebek et al. [234] conducted a literature review (n = 113) of theories related to security awareness in behavioral research. They proposed a meta-model by assembling the core constructs from the four most commonly applied theories in reviewed papers [234]. Moody et al. [264] empirically compared 11 theories with employees (n = 274), then they proposed and tested a unified model of ISP compliance (including constructs such as response efficacy, role values, and reactance) with 393 employees. To address the fragmentation and heterogeneity of research on cybersecurity self-efficacy, Borgert et al. [45] conducted a systematic literature review (n = 174). They made suggestions for standard and transparent self-efficacy measures and called for the pursuit of parsimony and falsifiability in self-efficacy theories, noting that inconsistencies often arise from deviations from the original theory and differing assumptions [45]. However, to the best of our knowledge, no systematic research has synthesized the role of autonomous motivation in organizational security behaviors.

To conclude, a plethora of theoretical frameworks have been utilized to study autonomous motivation and security behaviors. However, due to the distinctiveness of theoretical frameworks and a lack of synthesization of findings on the topic, it is still unclear how autonomous motivation influences employees' security behaviors. Consequently, we formulated the following research question:

> **RQ3**. Which theoretical frameworks have been employed to explore autonomous motivation in the domain of organizational information security, and how do these theoretical frameworks further our understanding of organizational information security?

Additionally, previous studies have indicated that study context, cultural background, and other various roles might influence the interpretation of study results [323, 161]. We propose that authors'

reflections on their study limitations and future opportunities from their work can provide information on how to further advance the field. Thus, our last two research questions are:

RQ4. What are the characteristics of the study contexts in terms of geographical location, industry sectors, and participants' job roles in the surveyed literature?

RQ5. What are promising avenues for studying autonomous motivation in the domain of organizational information security?

2.3 Systematic Literature Review

2.3.1 Preparation phase

In the preparation phase, we tested our search terms, verified whether the extracted papers were relevant, and refined the research questions and search terms. We began by identifying 77 papers that mentioned "intrinsic motivation," "security," and "employee" in their abstracts from the Scopus database. Of those, 22 were evaluated as relevant for defining the scope of our review and generating our research questions. These studies were conducted in the workplace or specifically mentioned that their study goal was to examine employees' motivation to engage in security behaviors.

2.3.2 Literature search

To cover the wide and interdisciplinary landscape of security behavior, we chose the ACM Digital Library (The ACM guide to computing literature) and the Scopus database as the initial data sources. The ACM Digital Library covers relevant computer science and IT-security-related publications. To complement the results with publications from related disciplines, Scopus indexes a wide range of peer-reviewed publications from different disciplines and is considered one of the most comprehensive databases [328]. Our review captures literature from fields such as usable security, information security management, and security behavior studies. These two databases cover most of the respective influential venues.

To address our research questions, we collected previous literature by conducting searches that combined the terms "autonomous motivators," "security behavior," and "workplace." We had to vary the terms slightly for each search engine. We provide all the searches as supplemental material. Here, we give an example search that we used in Scopus:

```
TITLE-ABS-KEY ( 'autonomous'' OR intrinsic'' OR endoge- nous'' ) AND
TITLE-ABS-KEY ( motivation ) AND TITLE-ABS- KEY ( security behavior''
OR security behavior'' OR cyber security'' OR information security'' OR
information tech- nology security'' OR it security'' OR information system
```

```
security'' ) AND TITLE-ABS-KEY ( employee'' OR workplace'' OR organization''
OR organisation'' OR company'' OR corpora- tion'' )
```

2.3.3 Preregistration

After conducting an initial screening of the resulting hits, we preregistered our review on the OSF to enhance the transparency of our process and facilitate the replication of the work by other researchers [74]. We follow the Generalized Systematic Review Registration Form proposed by Van den Akker et al. [375] for registration and documentation. In addition, we report our screening process, which adheres to the PRISMA guidelines [263, 160] for transparency and meta-analyses.

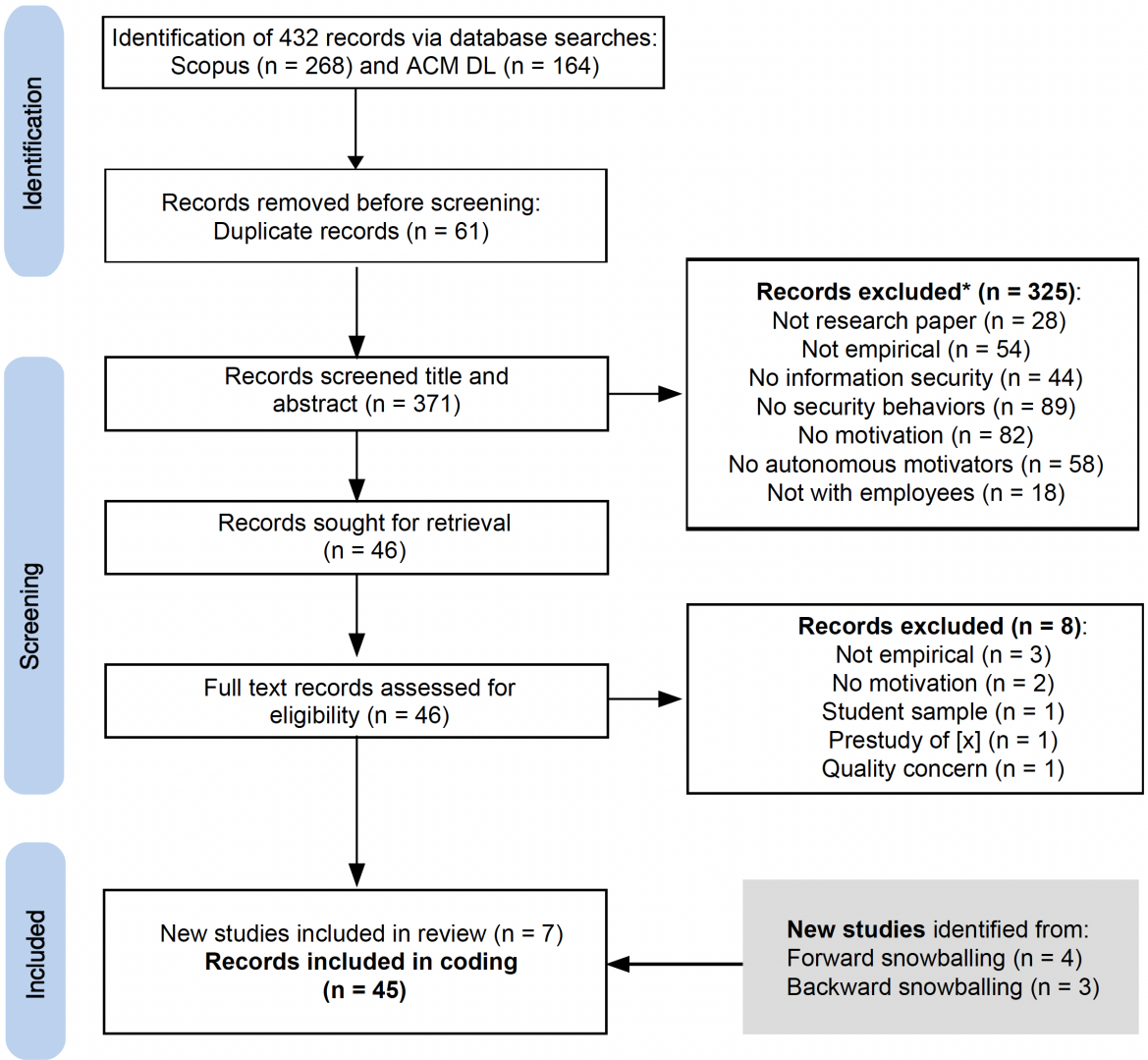

Figure 2.1: PRISMA flow diagram of study selection (Note: *Multiple reasons may apply; X: [136]).

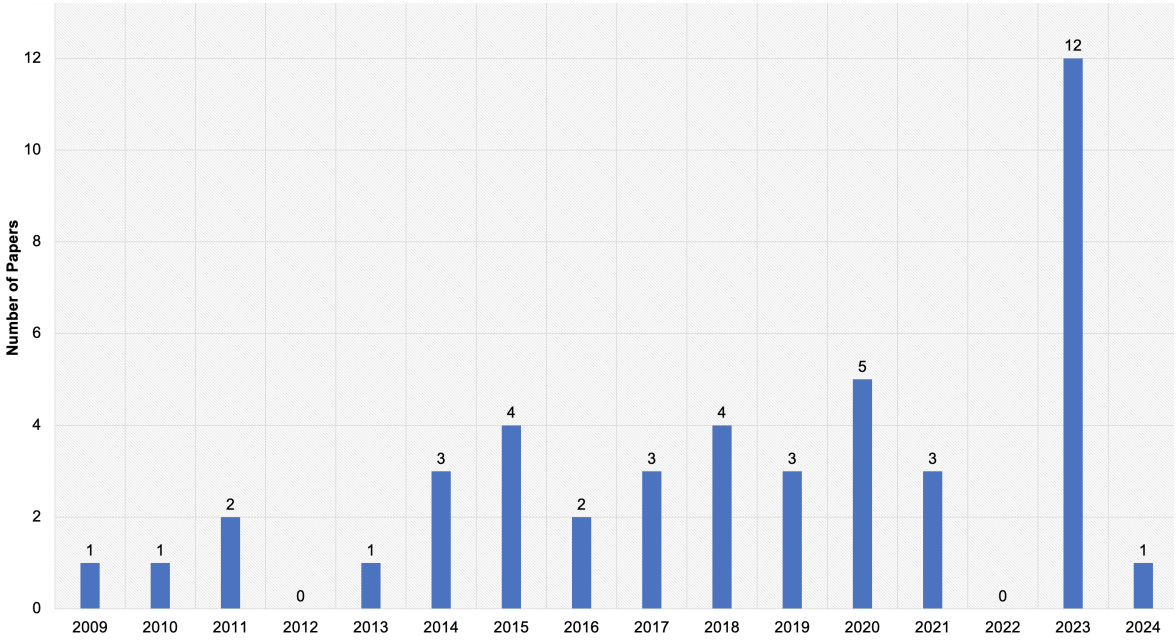

Figure 2.2: Number of papers by year.

2.3.4 Study selection

We manually searched and identified 432 publications of interest from the two databases on February 18, 2024. Subsequently, we downloaded their bibliographic information (including title, authors, abstract, and publication venues) and imported all of them into Rayyan, an online tool for systematic reviews[1]. Through Rayyan, automatic duplication identification, and manual confirmation, we removed 61 duplicates, and thus, 371 publications moved on to the screening process.

Two authors independently screened the same 26% of papers (95 out of 371) on the basis of our inclusion and exclusion criteria without being able to see the other coder's decisions. Of the 95 papers, the two authors agreed to include 13 publications and to exclude 78. The first author recommended the inclusion of one additional paper, whereas the second author recommended three other papers. Thus, to avoid potentially overlooking any relevant publications, 17 publications from this collection were included for further evaluation. The authors achieved 95.79% agreement (almost perfect agreement), with a Cohen's kappa of .84. The first author then screened the rest of the publications with the following inclusion and exclusion criteria and a total of 46 out of 371 publications moved on to further evaluation.

Inclusion: The study examined employees' autonomous motivation related to their cybersecurity behavior or intentions in the workplace as indicated by a screening of the title, abstract, and keywords.

Relevant *autonomous motivators* that were in line with the inclusion criteria comprised

[1] `https://www.rayyan.ai/`.

interest, curiosity, enjoyment, desire, satisfaction, empowerment, commitment, value, contribution, responsibility, fairness, moral belief, justice, ethics, legitimacy, endogenous motivation, autonomous motivation, intrinsic motivation, or fulfilling basic psychological needs (e.g., autonomy, competence, and relatedness).

Exclusion: The study was not peer-reviewed (e.g., doctoral dissertation), was not an empirical study focusing on employees (i.e., no participants were included or a student sample was used), or gave no indication of the data analytic methods or respective findings (e.g., it was a work in progress that lacked findings).

We used Zotero to screen the full text of retrieved publications. Eight publications were excluded for the following reasons: not an empirical study (n = 3), not related to employees' motivation (n = 2), student sample (n = 1), prestudy of another paper that was already included (n = 1), and concerns about quality[2] (n = 1). We applied forward and backward snowballing to the papers we retrieved in May 2024 and identified an additional seven papers that met our inclusion criteria. A total of 45 papers were included for our review (see Figure 2.2 number of papers by year). 32 were published in journals, addressing topics in information systems, security, and interdisciplinary fields. High-impact journals include *MIS Quarterly*, *Information & Management*, *European Journal of Information Systems*, *Computers & Security*, and *Computers in Human Behavior*. The remaining 13 papers appeared as conference proceedings at venues including the *Symposium on Usable Privacy and Security (SOUPS)*, the *Hawaii International Conference on System Sciences*, and the *AIS International Conference on Information Systems*. Appendix B contains a table of publication venues.

2.3.5 Paper extraction

To address our research questions, we used Microsoft Excel to extract and code (a) measurements of both motivators and behaviors, (b) theoretical framework related to autonomous motivation, (c) autonomous motivators, (d) security behaviors (or behavioral intentions), (e) research questions of the study, (f) relevant findings and conclusions, (g) study context, and (h) future work/study limitations. To standardize the coding process, the first author developed a detailed extraction manual to facilitate a systematic and consistent extraction that was jointly discussed and tested by all authors. Figure 2.3 illustrates the connection between the extracted content and our research questions.

First, all authors coded the same three papers to ensure a joint understanding. We resolved some ambiguities encountered during the process via discussion (e.g., when a paper mentioned multiple theoretical frameworks, we coded only the theoretical framework that was related to the autonomous motivators). Then, we extracted data from all papers. The data that each author extracted was

[2]The journal was de-listed from Scopus due to quality concerns in 2020. The paper was published in the same year of de-listing and does not match the journal's scope.

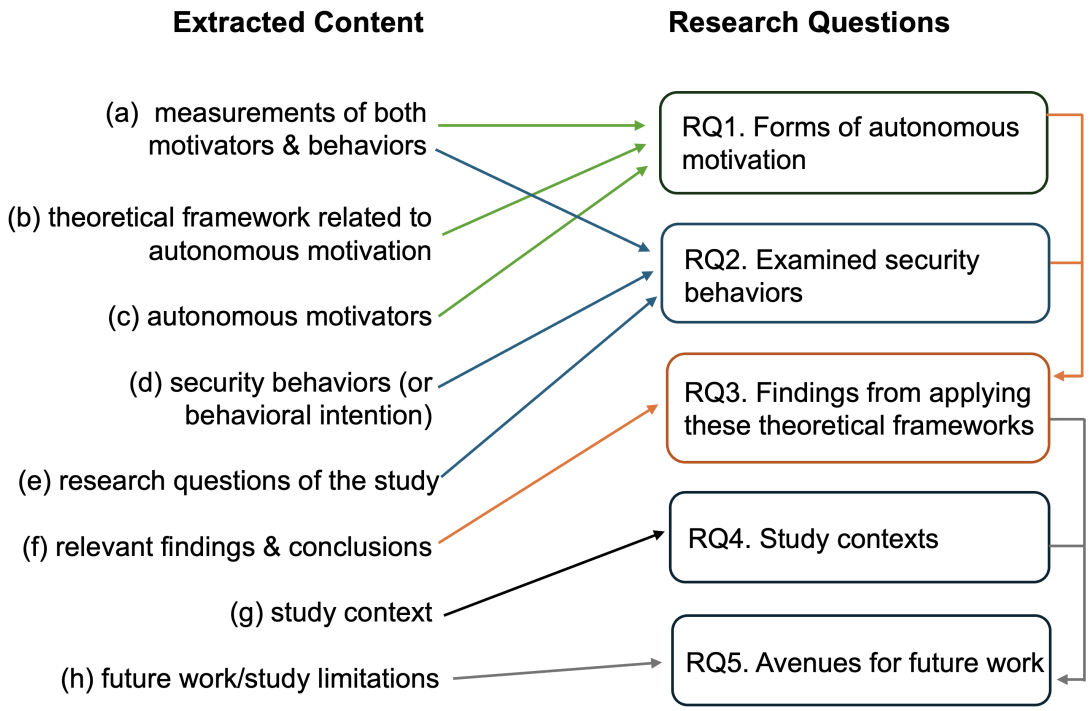

Figure 2.3: Matching extracted content with the research questions.

independently reviewed by another author to ensure accuracy and consistency. For transparency, the extraction manual is provided in the Supplementary Material along with all the extracted data.

2.3.6 Measurement extraction

During the extraction process, we observed that some references to the original sources that the authors borrowed or adapted their measurements from were quite vague. To ensure scientific rigor and to provide a useful toolbox of existing measurements for other researchers, we thus conducted a second iteration of the extraction phase with a specific focus on measurement extraction of autonomous motivators and security behaviors to identify all original references and also to describe the ways in which the original measures were adapted.

We also contacted the authors of 10 papers that did not include the complete measurements of either autonomous motivators or security behaviors to request the missing measurements. Four authors provided the requested measurements; however, for the remaining six papers, we were unable to obtain access to their full measurements. Consequently, our descriptions of their measurement and adaptation process are based on the information provided in their publications.

2.4 Results

2.4.1 Autonomous motivators examined in security behavior studies

Thirty-eight of the reviewed papers employed scales for measuring autonomous motivators in their study design. Among them, six papers' complete measurements were inaccessible. Thus, we extracted 97 measurements comprising 392 items from the remaining 32 papers. Additionally, there were four qualitative studies [131, 164, 43, 336] and three studies that used alternative measurements (i.e., Analytic Hierarchy Process questionnaires, a popular method for decision-making) [22], a single-item measure [136], and autonomous motivators as design principles [7]).

By analyzing the definitions, measurements (when provided), and references of all autonomous motivators, we identified 17 unique autonomous motivators that have been investigated in security behavior studies. We found that these motivators were driven by an individual's *interest*, the *intrinsic values* of engaging in the activity, or *personal values* and *expectations* associated with the activity. Additionally, a group of motivators was driven by *fulfilling psychological needs*, such as autonomy or relatedness. Table 2.1 provides an overview of all motivators, their definitions, and how often they have been studied. The category was developed from the SDT-ER taxonomy suggested by Frank and Kohn [131] but extended with insights from the empirical studies included in our review, thus resulting in five groups of motivators (see Table 2.1). We marked our proposed new categories with asterisks in Table 2.1.

To provide a useful toolbox for researchers or practitioners interested in measuring autonomous motivation in the security domain, we also extracted and categorized all provided measurements from the reviewed articles, their original sources where applicable, and whether they were adapted. Table 2.2 provides an exemplary excerpt of the complete overview table provided in the supplementary material.

Interest-driven motivators (n = 4).

Joy, curiosity, and satisfaction are intrinsic and interest-driven motivators [131]. In the SDT framework, these motivators are termed intrinsic motivation [313]. In the studies we reviewed, interest-driven motivators were either examined with specified terms such as joy (e.g., technophilia [131], interest [164], or satisfaction [318]) or applied as principles (e.g., joy and curiosity [338]) for training design. Compared with the other motivators associated with a specific task or values/expectations associated with engaging in an activity, intrinsic motivators emerge when an individual is internally driven.

Table 2.1: The taxonomy of autonomous motivation related to organizational security behaviors (N = number of reviewed studies that examined the motivator. *new categories proposed in the study).

Driver	**Motivator and definition**	**N**
Interest-driven	• *Joy, Satisfaction, Curiosity (Intrinsic Motivation)*: employees engaging with an activity because they like the activity, enjoy doing the activity, or derive satisfaction from performing the task [338].	4
Task-driven*	• *Intrinsic Task Value*: the anticipated or actual enjoyment derived from engaging in a specific task (e.g., feeling rejuvenated or content) [111].	11
	• *Job Satisfaction*: satisfaction or positive emotions derived from one's job [377].	4
Psychological Needs Fulfillment*	• *Competence-related*: includes (a) self-efficacy, which refers to the belief in one's ability to achieve specific outcomes [29], (b) perceived competence, and (c) perceived behavioral control.	24
	• *Autonomy-related*: an individual's perception of the extent to which they are engaging in an activity of their own choice [262, 257].	14
	• *Perceived relatedness*: "the degree of connectedness an individual feels toward others when interacting in a specific context" [257, p.1212].	10
	• *Protection motivation*: employees' intention/likelihood of protecting themselves and their organization from security threats [131, 287].	4
Value-driven	• *Altruism*: people perceiving "the act of helping as enjoyable and interesting" [131, p.4].	1
	• *Commitment*: "an affective attachment to the organization" [261, p.539] and "a willingness to exert effort on behalf of the organization" [191, p.71].	8
	• *Perceived Value Congruence*: the extent to which an employee and their employer share the same values [344].	3
	• *Organizational Justice*: employees' perceptions of fairness in the processes and outcomes of organizational decisions [236].	4
	• *Personal Responsibility*: employees' willingness to be accountable for their work-related choices, behaviors, and outcomes [186].	12
	• *IS Identity*: a person's self-concept of their roles, responsibilities, and importance of complying with information security policies [277].	2
	• *Normative Beliefs*: an individual's beliefs about whether important others or groups approve or disapprove of a specific behavior [4].	3
	• *Employee Involvement*: personal norms [191], attachment [191], involvement [191], user-IS exchange [92], and management support [364] in the workplace.	5
Expectation*	• *Perceived Benefits*: an individual's expectations of the benefits of performing a task, including saving time, convenience, increasing productivity [201], demonstrating impact [99], and making a difference [170].	4
	• *Response Efficacy*: an individual's belief in the effectiveness of a prescribed solution in mitigating the threat [257, 338].	12

Task-driven motivators (n = 11).

Intrinsic task value describes the positive experiences derived directly from a task that motivate individuals to engage with it [368]. Researchers sometimes intermix the terms intrinsic value and

Table 2.2: Excerpt of the overview of the applied measurements for autonomous motivation for all reviewed articles. (Note: DT = Deterrence Theory, EVT = Expectancy Value Theory, SDT = Self-Determination Theory, na = not available.)

Authors	**Theoretical framework**	**Motivator**	**Measurement**	**Example**	**Source**	**Adaption**
Alahmari et al. [7]	SDT	Psychological Needs Fulfillment	autonomy, competence, relatedness	na	na	na
Alhelaly et al. [12]	EVT	Interest-driven	Intrinsic Interest Value	In general, I find protecting my mobile identity is (extremely boring -extremely interesting).	[109]	yes
Alzahrani & Johnson [22]	SDT	Psychological Needs Fulfillment	autonomy, competence, relatedness (relation of needs to each other)	scale from 9-1 and 1-9 with competence and autonomy as end points	na	na
Alzahrani et al. [21]	SDT & DT	Psychological Needs Fulfillment, Value-driven	perceived autonomy/ relatedness/ competence; perceived legitimacy; perceived value congruence;	na	na	na
...	...	...	...	...	...	...

intrinsic motivation [185]. However, Eccles and Wigfield argued that it is necessary to differentiate the "internal origin of the desire to engage in an activity" (intrinsic motivation) from "the enjoyment of the task itself" (intrinsic value) [111]. We thus propose that intrinsic value[111] should be renamed intrinsic task value to make this term more distinguishable from intrinsic motivation. For example, Silic and Lowery surveyed employees' perceptions of feeling rejuvenated, lowering their stress, and passing the time more enjoyably when using a web system [338]. Other tasks have also been examined for the intrinsic value related to performing the task of protecting one's mobile identity [12] and complying with ISPs [53, 25, 12, 235]. These studies explored how employees' feelings of accomplishment, fulfillment, contentment [53, 25, 67], importance [99], and pleasantness [235] were associated with compliance. In addition, four of the studies we reviewed [377, 287, 288, 186] examined the relationship between *job satisfaction* and ISP compliance.

Motivators fulfilling psychological needs.

These motivators specifically address and satisfy individuals' basic psychological needs (e.g., competence, autonomy, relatedness [314], and safety).

Competence-related (n = 23). Competence-related motivators were the most frequently examined motivators in the studies we reviewed. They took the form of self-efficacy (n = 13), *perceived competence* (n = 6), and *perceived behavioral control* (n = 4). Despite different theoretical origins [29], papers in our sample often used self-efficacy and perceived competence interchangeably [218, 99]. SDT posits that the need for competence drives individuals to master significant tasks [300], whereas Social Cognitive Theory defines self-efficacy as the belief in one's ability to achieve specific outcomes [29]. Bulgurcu et al. [53] proposed self-efficacy, along with behavioral beliefs and normative beliefs, as an antecedent of attitudes when introducing the Theory of Planned Behavior. However, Ajzen, who postulated the Theory of Planned Behavior, stated that perceived behavioral control, which is conceptually similar to self-efficacy, captures the extent to which an individual has the ability to perform the behavior and how much the behavior is under their control [5]. Taking this viewpoint into account, we grouped perceived behavior control with self-efficacy and perceived competence.

Autonomy-related (n = 15). Spreitzer defined one dimension of *psychological empowerment* as self-determination to "reflect autonomy in the initiation and continuation of work behaviors and processes" [346]. One of our reviewed papers used the term "choice" to refer to self-determination [99]. Scrutinizing their cited source [346], we included "choice" [99] as one adapted measurement of *autonomy*. Additionally, Gerdenitsch et al. [138] used the measurement of *decision-making autonomy* to capture whether workers are given the autonomy in their work to make their own decisions. Thus, we consider decision-making autonomy as one application of autonomy in the work context.

Perceived relatedness (n = 10). In the studies we reviewed, some scholars used the term perceived relatedness to capture an individual's connection to their digital information and online accounts [257, 408, 278]; whereas others believe that the *camaraderie* and *group attachment* employees form in the workplace prompt their security behaviors [131].

Protection motivation (n = 4). Protection motivation has been investigated in organizational settings, regarding employees' intention/likelihood of protecting themselves [131] and their organization [287, 277, 288] from security threats.

Value-driven motivators.

These motivators are congruent with employees' endorsed goals, values, and ethics that are integrated into their identity [314].

Altruism (n = 1). Altruism was observed as motivating employees to help their colleagues with

security-related problems, share their security knowledge, and reduce the workload of IT personnel [131]. Frank and Cohn suggested that altruistic values enhance extra-role security behavior, especially by promoting helpful behaviors, stewardship, civic virtue, and organizational loyalty [131].

Commitment (n = 8). Three studies tested the relationship between *affective commitment* and compliance behavior [377, 287]. Others examined *organizational commitment* more generally and defined it as employees' acceptance of an organization's information security goal and policy, along with a willingness to invest effort in information security [191, 92, 67]. Posey et al. [288] suggested that organizational commitment serves as the mechanism that makes workplace security threats personally relevant to employees.

Perceived value congruence (n = 3). Son [344] introduced the concept of *perceived value congruence* to measure the extent to which an employee and their employer share the same values. Chen and Li [67] used the same measurement items as Son [344], but they renamed the construct perceived value fit.

Organizational justice (n = 4). Li et al. examined organizational justice's role in motivating employees' internet policy compliance intentions [236]. They tested four dimensions of organizational justice (i.e., procedural, distributive, interpersonal, and informational justice [236]). Li et al. [236] hypothesized that the four dimensions of justice beliefs, when enforced within an organization, influence employees' workplace ethics and subsequently enhance their intentions to comply. Another two studies measured whether perceived fairness positively affected employees' attitudes toward ISP compliance [21, 235]. Son [344] introduced a similar concept, *perceived legitimacy*, which measures the extent to which employees view the ISP as appropriate, desirable, and just.

Personal responsibility (n = 12). When employees consider security actions to be more their responsibility than their employers', they are more likely to perform security actions [186, 43, 131], such as installing software updates [42]. We identified some constructs that are closely related to responsibility, including locus of control [191] and ownership [201]. Ifinedo [191] measured *locus of control* (which refers to the extent to which people believe they have control over the course of events [345]); however, the items involved the concept of responsibility, such as "the primary responsibility for protecting my organization's information belongs to others and not me" [191]. In a security context, psychological ownership can be defined as a feeling of possessiveness an individual develops for a security task [201]. Psychological ownership has been empirically tested to confirm its relationship with ISP compliance [201] and updating software [42]. Additionally, three studies investigated how *perceived excessive responsibility* influences employees' compliance [201, 204, 203], that is, employees' sense of going beyond their regular work duties [204].

Information security identity (n = 2). Employees' information security identity (IS identity) may

be a driver of their compliance-based and voluntary security behaviors [277]. Two related constructs are *internalization of information security policies* (e.g., "I contribute to the organization by complying with its information security policy." [282]) and *internal perceived locus of causality* (e.g., "I comply with the requirements of the ISP because I want to find out how to ensure information system security." [218]).

Normative beliefs (n = 3) and *employee involvement (n = 5).* Normative beliefs and employee involvement have been examined as motivators of ISP compliance in the studies we reviewed. [218, 53, 191] examined the relationship between normative beliefs and ISP compliance intentions. In addition to being influenced by others, employees might be motivated to comply with ISPs for their own reasons. We put these motivators together with employee involvement, which includes *personal norms* [191] (personal belief in the relevance of complying), *attachment* [191] (communication with colleagues and respect for their ISP views), *involvement* [191] (actively involving oneself in information security), *user-IS exchange* [92] (employees' perceptions of and interactions with the information security department), and *management support* [364, 287, 67] (peer/higher management/technical support for information security).

Expectation motivators.

Expectation motivators are related to the expected outcome and the perceived benefits of performing an activity.

Perceived benefits (n = 4). Even though the following concepts were labeled "intrinsic" and "satisfaction," when we scrutinized the measurement items, they focused on expectations of a certain reputation and better cooperation (*intrinsic outcome expectations*, [129]) and whether the solution would be efficient (cost/benefits) or effective in protecting the organization (*self-worth satisfaction*, [318]).

Response efficacy (n = 12). Response efficacy is one of the most frequently studied Protection Motivation Theory components [158]. Employees might evaluate the efficacy of following an organization's ISPs or performing preventive measures. Perceived high response efficacy motivates individuals to comply with ISPs [389] and engage in security behaviors [257].

2.4.2 Security behaviors related to autonomous motivation

All 45 of the papers we reviewed examined security behaviors or behavioral intentions as outcomes of autonomous motivation, along with other factors of influence. Except for four qualitative studies [336, 43, 131, 164], the remaining 41 studies used single questions, items, or log data to evaluate participants' security behaviors. We found that general ISP compliance and specific security tasks

were often examined separately [80, 42, 131]. When switching from one security task to another, employees' intentions to perform a task can vary significantly [42]. Therefore, we present our findings of various security behaviors related to autonomous motivation in three categories: *information security compliance*, *extra-role security*, and *ISP violation behaviors*. See the overview of security behaviors related to autonomous motivation in Table 2.3. Similar to the excerpt provided in Table 2.2, we also provide an overview of the security behaviors and their measurements in the Supplementary Material to provide a useful toolbox for future research.

Table 2.3: Security behaviors related to autonomous motivation. (Note: [138, 277, 92, 202] examined two types of behaviors in their study.)

Type of behavior	**Behavior/Intention examined**	**Count**
ISP compliance (n = 24)	Intentions to comply with ISPs	16
	Attitude toward ISP compliance	2
	Performance of specific compliance tasks	6
Extra-role (ER) security behaviors (n = 22)	ER behavioral intentions	1
	ER volunteering intentions	1
	Participation in ER behaviors	5
	Protection-motivated behaviors	2
	Security knowledge sharing	4
	Actions not prescribed in the ISP	7
	Cybersecurity advocates	1
	Attack-focused tasks	1
ISP violation behaviors (n = 3)	Insider computer abuse	1
	Instrumental policy abuse	1
	Infringing ERM rules	1

ISP compliance behaviors

ISPs prescribe employees' responses for securing corporate information [344]. Employees' compliance with the organization's ISP is the key to enhancing information security [53]. All our reviewed ISP compliance studies (n = 24) examined either respondents' self-reported behavioral intentions/attitudes or likelihood of performing specific tasks.

Intention to comply with the ISP (n = 16). Nine studies utilized the measurement proposed by Bulgurcu et al. [53] that includes three statements referring to complying with the ISP requirements, protecting information and technology resources, and carrying out the responsibilities prescribed by the ISP. Four studies used Herath and Rao's measurements [170] to indicate their likelihood and certainty of following the organization's ISP [170]. Hong and Xu [186] used a scenario-based scale to survey respondents' intentions to comply with the ISP. They adapted the scenarios from the four scenarios (user authentication and access control, hardware, software, and the network) created by Guo et al. [155]. Two studies [21, 22] did not specify their ISP measurement items.

Attitude toward ISP compliance (n = 2). Awudu and Terzis [25] examined respondents' evaluative judgments of the importance, necessity, benefit, and usefulness of complying with the ISP. Tejay and Mohammed [364] surveyed employees' perceived value and the effectiveness of the information security program in protecting critical information. Additionally, employees were asked to indicate their views on whether the security program balances risks with security controls [364].

Performance of specific compliance tasks (n = 6). Son [344] surveyed employees' compliance with tasks such as (a) accessing information assets, (b) communicating via email, (c) handling internet and network resources, (d) performing antivirus actions, and (e) preventing unauthorized access. Li et al. [236] investigated whether employees followed their organization's internet use policy, whereas Jeon et al. [202] examined employees' use of enterprise rights management (*ERM*) systems in their organizations. Two studies [277, 136] assessed employees' adherence to the organization's ISP, regarding protecting sensitive information, changing passwords as per policy, and securing workstations when unattended. Interestingly, Vedadi et al. [377] expanded the role of management in their data collection and instructed supervisors to rate their employees' security practices with respect to the discussion of sensitive information, compliance with security procedures, and adherence to information security rules.

Extra-role security behaviors.

These behaviors comprise "spontaneous security actions that are not defined by organizational rules or policies" [131, p.2]. Examples include voluntarily helping others, actively intervening, accepting obstacles without complaint, and actively participating in improving security measures [131]. Twenty-two of the studies we reviewed examined such security behaviors that employees may view as extra-role security behaviors.

Extra-role behavioral intentions and participation (n = 7). Chen and Li [67] introduced *extra-role behavioral intentions* to assess employees' intentions to perform extra-role security behaviors in the workplace. They examined the extent to which employees promoted the information security program, put forth extra effort to enhance security, and voluntarily engaged in activities such as reporting risks or proposing new strategies [67]. Similarly, Davis et al. [92] used the concept *extra-role volunteering intentions* to survey employees' general intentions to engage in voluntary and proactive efforts to enhance information security. Furthermore, five studies asked employees to self-report their participation in extra-role security behaviors [138]. These self-reported questions were related to different aspects of organizational information security, such as helping colleagues/new employees learn about the ISP [277], evaluating the effectiveness of the system [332, 336], and reporting when suspicious emails had been received [282].

Protection-motivated behaviors (n = 2). Posey et al. [287] introduced protection-motivated behaviors (PMBs) to emphasize the critical role of employees' safe computing practices in organizational security. PMBs are defined as voluntary actions by insiders aimed at safeguarding both organizational information and the information systems that manage security threats [287]. In a multidimensional scaling study [286], Posey et al. categorized 67 PMBs into 14 clusters on the basis of levels of improvement needed, standardization and application, and reasonableness. These clusters include employees' behaviors, such as email handling, data protection, security training, software use, and account protection [286].

Information security knowledge sharing (n = 4). Information security knowledge sharing refers to sharing knowledge about information security to increase security awareness and mitigate security risks [318, 131]. Alahmari et al. [7] elaborated on the idea that security knowledge sharing implies a collaborative approach to cybersecurity, which is a powerful and efficient solution for mitigating cyber attacks. Frank and Ament [129] investigated the motivational factors influencing employees' intentions to share their information security incident experience. They argued that communicating incident experiences in the workplace can act as a social learning strategy that allows employees to learn from their colleagues' security incidents [129].

Security actions not prescribed in the ISP (n = 7). Researchers investigated a range of security actions that might enhance organizational information security, even though these actions were not outlined in the ISP [43, 136]. Blythe and Coventry [42] examined employees' intentions to *scan USB sticks* with anti-malware software and *install software updates promptly*. Alhelaly et al. looked into the motivational aspects of *mobile identity protection* due to the significant amount of important data stored on these devices [12]. Ogbanufe et al. [278] examined factors that motivate employees to voluntarily *adopt multifactor authentication*. Finally, Menard et al. [257] and Yang et al. [408] explored the application of security messages that appeal to individuals' psychological needs as a method for encouraging people to *adopt password managers*.

Roles of cybersecurity professionals (n = 2). Haney and Lutters [164] examined the work motivation of *cybersecurity advocates*. These security professionals promote, educate, and motivate workers to adopt the best practices for security in the workplace [164]. Hodges and Buckley [185] examined differences in motivation and self-efficacy between two cybersecurity behaviors: *attack-focused tasks* (e.g., red-teaming and exploit development) and defense-focused tasks (e.g., network design and policy writing). They asked security professionals to estimate the ratio between the amount of defense-focused work and attack-focused work they performed [185].

ISP violation behaviors.

Three studies examined employees' organizational ISP violation behaviors. These behaviors include insider computer abuse [58], instrumental policy abuse [386], and infringing ERM rules [202]. *Insider computer abuse* refers to unauthorized and deliberate employee behaviors that harm organizational information assets [58]. Welck et al. [386] argued that enterprises rely on information technology to facilitate work tasks, and merely prohibiting harmful use through security policies often leads to employees' *policy abuses*. Similarly, Jeon et al. [202] examined employees' behaviors with respect to *infringing ERM rules*, for example, accessing information through a borrowed account.

2.4.3 Applied theoretical frameworks and key findings

Among the 45 reviewed papers, 24 different theoretical frameworks related to autonomous motivators were mentioned. We include a glossary of 24 theoretical frameworks from the studies we reviewed in Appendix A. SDT was the most frequently cited (n = 16), followed by Protection Motivation Theory (n = 7), Theory of Planned Behavior (n = 6), and Deterrence Theory (n = 3). In the following subsections, we summarize the key findings from the reviewed papers on the basis of the approaches the studies used to engage the theoretical frameworks, that is, deductive, inductive, and design approaches [100]. Two papers [204, 332] did not indicate a specific theory in their research; thus, we exclude them from this subsection leading to 43 papers.

Deductive approach (n = 35)

The deductive approach refers to papers that test predefined hypotheses or research models. Online surveys were used most frequently in the reviewed papers. Most authors chose this methodology to test the assumed relationships between the constructs from their conceptualized research models. This process included (a) verifying whether a theoretical framework from other disciplines was useful for interpreting IS topics [92]; (b) comparing two theoretical frameworks to examine which one provided more explanations about IS phenomena [257]; and (c) extending an established theory with constructs from another theory [191] or factors of influence identified in prior studies [42]. We apply our proposed taxonomy to summarize the findings from the reviewed studies (deductive approach) and provide an overview in Table 2.4.

Intentions to comply with/attitude toward ISPs. From studies examining the antecedents of employees' ISP compliance, we found that some autonomous motivators were consistently related to employees' intentions to comply/attitude toward compliance, whereas others demonstrated non-significant or mixed results:

Table 2.4: Security behavior/intentions and autonomous motivation matrix from the studies we reviewed utilizing the deductive approach. (Note: n = the number of security behaviors that have been examined more than once; non-sig = the motivator did not demonstrate statistical significance, otherwise, the motivator was found to be significant; mixed = the motivator had mixed results regarding significance from different studies on the behavior type; inversely = the motivator is inversely related to employees' intentions to perform the behavior; otherwise, the motivator was found to be positively related to employees' intentions to perform the behavior. Motivators related to security behavior or intentions via a moderator are not included in this table.)

	Expectation	**Value**	**Needs fulfillment**	**Task**	**Interest**
Intention/attitude to comply with ISPs	Perceived benefit; Response efficacy	Organizational justice (non-sig); Commitment; Personal responsibility (4); Perceived value congruence; Normative beliefs (3)	Autonomy-related; Competence-related (mixed, 5); Perceived relatedness	Intrinsic task value; Job satisfaction (2)	
Compliance with specific security tasks	Perceived benefits (non-sig)	Organizational justice (2); Personal responsibility (mixed, 2); Perceived value congruence; IS identity	Protection motivation	Intrinsic task value (non-sig)	
Extra-role and Protection-motivated behaviors/intentions	Response efficacy; Perceived benefits	Perceived value congruence; IS identity (2); Normative belief (2); Commitment (mixed, 2)	Competence-related; Protection motivation (mixed, 2)	Intrinsic task value (2); Job satisfaction (non-sig)	
Security knowledge sharing	Perceived benefits	Normative beliefs	Competence-related		Intrinsic motivation
Actions not prescribed in the ISP	Response efficacy (4)	Personal responsibility (mixed, 4)	Competence-related (mixed, 5); Autonomy-related (3); Perceived relatedness (mixed, 2)	Intrinsic task value	
Attack-focused tasks			Competence-related	Intrinsic task value	
Insider computer abuse		Personal responsibility (inversely)			
Instrumental policy abuse	Response efficacy (inversely)		Autonomy-related (inversely)		
Infringing ERM rules	Perceived benefits (inversely)	Personal responsibility (inversely)			

- **Significant antecedents**: Intrinsic task value [235], Job satisfaction [186, 377], Autonomy-related [21], Perceived relatedness [21], Personal responsibility [201, 203, 191, 186], Perceived value congruence [21], Normative beliefs [218, 53, 377], Commitment [377], Perceived benefit [201], and Response efficacy [389].

- **Mixed results**: Competence-related [21, 53, 191, 389, 203].

- **No significant effects**: Organizational justice [235].

Compliance with specific security tasks. Li et al. [236] found that personal responsibility (work-related ethical beliefs) had a positive impact on employees' internet use policy compliance than the sanction-based approach (sanction severity and certainty). Organizational justice (procedural and distributive justice) had a positive impact on compliance intentions directly and indirectly by fostering work ethics [236]. Similarly, the autonomous motivators (perceived legitimacy and value

congruence) contributed significantly more to the explained variance in employees' compliance than extrinsic motivators (deterrent certainty and severity) [344]. Furthermore, Ogbanufe and Ge [277] revealed that whereas protection motivation and IS role identity were positively related to compliance behaviors, intrinsic task value did not demonstrate a significant relationship with in-role compliance. Empowerment-based ERM can enhance employees' perceived responsibility and benefits; however, this empowerment does not lead to increased compliance with ERM regulations [202].

Extra-role and protection-motivated behaviors/intentions. Employees who internalized information security policies self-reported more security practices compared with those who merely complied with the policies [282]. Perceived control, IT competence, and user-IS exchange are positively associated with information security commitment [92]. Information security commitment [92, 67] promoted employee participation in extra-role behaviors. Organizational commitment made information security threats personally relevant to employees [288]. Whereas intrinsic task value and IS identity were positively related to extra-role security behaviors, protection motivation was negatively related to extra-role security behaviors [277]. Response efficacy showed a strong positive correlation with both protection motivation and self-reported engagement in protection-motivated behaviors [288]. While management support had a significant effect on insiders' protection motivation [288], increased job satisfaction did not significantly impact their protection motivation [287]. Intrinsic task value and perceived benefits significantly influenced an individual's intention to protect their mobile identity [12].

Security knowledge sharing. Perceived benefits (e.g., reputation and sense of accomplishment) were more effective at encouraging the sharing of incident experiences than external rewards such as incentives [129]. Perceived behavioral control and normative beliefs also significantly influenced employees' sharing intentions [318]. Additionally, intentions to share and trust had significant effects on security knowledge-sharing behavior within organizations [318]. Satisfaction of curiosity positively influenced employees' intentions to share knowledge within information systems, mediated through their attitudes [318].

Actions not prescribed in the ISP. Response efficacy significantly facilitated anti-malware behaviors [42]. Employees with a stronger sense of personal responsibility for security had greater intention to engage in anti-malware software and software updates [42]. Self-efficacy emerged as the strongest predictor of both anti-malware software use and email security behavior but had no impact on software update behavior [42]. In another study, Blythe et al. [43] found that whereas employees accepted some responsibilities, they diffused others onto their organization [43]; additionally, low response efficacy, driven by a lack of feedback on the effectiveness of employees' responses, was identified as a potential barrier to certain security practices [43]. Furthermore, Ogbanufe et al. [278] revealed

that autonomy and relatedness exhibited significant correlations with intrinsic task value, whereas competence did not. Subsequently, intrinsic task value was significantly associated with the voluntary use of multifactor authentication [278]. Perceived autonomy, competence, and relatedness were significantly related to home users' intention to install a password manager [257]. However, in a replication study with organizational users, Yang et al. [408] found that only autonomy demonstrated significant correlations. Lastly, in a survey with 137 cybersecurity professionals, Hodges and Buckley [185] found that individuals who chose to focus more on *attack tasks* were more internally motivated, with a higher intrinsic task value and higher self-efficacy than those focused on defensive tasks.

ISP violation behaviors. The reviewed studies suggested that empowering employees and fostering their sense of responsibility can serve as remedies for curbing behaviors that violate rules in organizations [202, 58]. Jeon et al. [202] indicated that granting employees the autonomy to access information within a defined set of rules leads to greater perceived benefits and an added sense of responsibility compared with those using control-based systems, which reduce users' intentions to circumvent access rules [202]. In an online vignette experiment, Welck et al. [386] found that two dimensions of psychological empowerment, self-determination (autonomy) and impact (response efficacy), had a significant negative effect on employees' intentions to abuse the rules [386]. Burns et al. found that employees' perceptions of maladaptive financial benefits and psychological contract violations were positively related to insider computer abuse [58]. Employees' personal responsibility (self-control) was found to negatively moderate the relationship between their abuse motives and insider computer abuse.

Inductive approach (n = 3)

We categorize papers as inductive if their approach was to derive theoretically cohesive abstractions from observations. Frank and Kohn [131] examined various types of extra-role security behaviors and their motivators by conducting in-depth interviews (n = 29). They found that interest, competence, autonomy, and a sense of connection influence these behaviors. Employees exhibited different extra-role security behaviors based on distinct motivational factors, suggesting the need for targeted interventions. Organizations should also identify highly motivated employees and clarify the boundaries of acceptable extra-role security behaviors. Through interviews with cybersecurity professionals (n = 28), Haney and Lutters [164] identified several intrinsic drivers of cybersecurity advocacy, including interest, a sense of duty, self-efficacy, evidence of impact, camaraderie, and, to a lesser extent, awards and monetary compensation [164].

In a case study of Ghanaian government employees, Awudu and Terzis [25] explored attitudes toward ISP compliance and perceptions of intrinsic and extrinsic rewards. Their findings indicate

that, despite the absence of a formal ISP and related training, a positive information security culture existed within the organization [25]. Employees recognized the necessity, benefits, and importance of the ISP, and they felt content, satisfied, accomplished, and fulfilled when they adhered to it. However, perceptions of extrinsic rewards were less clear. Whereas experienced staff reported that they generally believed that extrinsic rewards do not motivate compliance, the viewpoints of inexperienced staff were uncertain [25].

Design approach (n = 5)

Here, we refer to papers that adopt theories to inform the design of a tool, intervention, or product. Silic and Lowry [338] tested a *gamified security training* versus an email-based training in the field. Their longitudinal findings suggested that the gamified training inherently motivated employees to learn and adhere to security policies, and perceived intrinsic usefulness and curiosity increased employees' behavioral intentions to follow security policies [338]. Similarly, Alahmari et al. [7] designed a *mobile intervention* by using elements such as badges and a leaderboard to encourage sharing of security knowledge in the workplace. The intervention, designed to address employees' basic psychological needs, improved their knowledge of and their responses to security incidents, in comparison with the control group [7].

Shojaifar et al. [336] introduced CYSEC, an automated cybersecurity *communication tool* designed to promote cybersecurity practices in the workplace. The tool leverages SDT constructs to guide and motivate companies toward adopting effective cybersecurity measures. Their observations indicated that enhancing self-efficacy positively influenced users' self-motivation, whereas providing choices supported both autonomy and self-motivation [336]. Lastly, two studies combined constructs of SDT with established *assessment methods*. Alzahrani and Johnson [22] developed a questionnaire, using the Analytic Hierarchy Process method, to survey the weights for autonomy, competence, relatedness, and behavioral intentions to comply with ISP. Gangire et al. [136] created an information-security-compliant behavior questionnaire with questions related to competence, relatedness, and autonomy based on the Human Aspects of Information Security Questionnaire (*HAIS-Q*).

2.4.4 Study contexts and control variables

We analyzed the context-related and control variables of all 45 papers. Below, we highlight the more widely reported and impactful contextual factors. A complete overview of the demographic and contextual factors can be found in the Supplementary Material.

Study context:

Demographic information. The number of participants in the papers varied, ranging from 15 to 993 participants. Naturally, quantitative papers tended toward a larger sample size with a median of 289. For qualitative papers, the median number of participants was 25.

Ten papers did not report participants' ages, and two more provided only very vague information. However, of those 12 that reported, the median mean age was 38.43. The remaining papers provided age ranges, where 16 reported the largest proportions for groups between 25 and 49 years of age. Overall, only few papers investigated very young or older participants as a primary target population, with two papers reporting a large proportion of participants below 25 [12, 25] and three papers reporting a large proportion of participants above 50 years old [389, 236, 235]. None of the papers reported any age-based recruitment criteria.

Nine papers did not report participants' gender. The majority of the remaining papers reported a balanced distribution of genders, whereas 12 papers had a skewed proportion where any gender took up more than 60% of the entire sample. Of these, seven were skewed toward a male population, whereas five were skewed toward a female population.

Geographical location. The majority of studies took place in Western countries, with a particular focus on the United States. Seven online studies did not specify the location or geographic composition of its sample. Geographical locations are summarized in Table 2.5.

Table 2.5: Geographical areas of studies.

Region	**Count**
North America	17
Europe	7
Other Western	1
Middle East	4
East Asia	7
Africa	2
Not specified	7

Industry. The studies collected data from people who were actively employed, although industry was not always explicitly stated. There were a broad variety of sectors, and education, finance, government, and healthcare were prominent in the reviewed studies.

Job roles. Most papers investigated employees in general, irrespective of their exact job role. Whereas some papers investigated specialized groups, such as cybersecurity specialists [164, 185] or security managers [22], the studies did not explore differences between these groups and other employees or roles.

Control variables in the studies and their findings

Of the 40 papers employing quantitative analyses, 20 reported the use of control variables. We provide an overview of the frequencies of these control variables and their impacts in Table 2.6. In all cases, age was used as a control variable. Gender and education were used in 18 and 12 papers, respectively. We grouped together control variables that determined the degree of experience an employee might have, such as tenure with an organization, general job market experience, or amount of experience with specific systems. 12 papers controlled for these factors. Finally, 12 papers also controlled for factors indicating either job role or job status, such as whether an employee was a specialist or was working in a managerial position. A few papers also controlled for various other factors such as organization size [21, 92, 170, 218], self-efficacy [389], or security-related awareness [92].

Table 2.6: Control variables and their impact.

Attribute	Count	Impact
Age	20	Higher age can increase compliance intentions and behavior.
Gender	18	Women can demonstrate higher compliance intentions.
Education	12	Mixed effects.
Experience	12	No statistically significant effect reported.
Job role/Status	12	Generally positive influences of more specialized and hierarchical positions.

A total of 18 papers reported on the effects of control variables. Of these, three found no significant influence of any variable. In the following, we detail any control variables for which significant effects were reported in more than one case. Age appeared to have a positive influence on compliance intentions and behavior in some cases, with ten papers reporting no significant effect and five papers finding that older individuals showed more security intentions and behavior [21, 203, 204, 236, 389]. Gender seemed to have minor influence overall, with no significant effect reported in 11 papers but women demonstrating higher compliance intentions in two cases [170, 191]. The effects of other variables were less clear. Education was reported as not significant nine times, whereas significant effects were reported three times, with one paper stating a positive relationship between education and extra-role behavioral intentions [67], and two showing that it decreased information security engagement intentions [21, 92].

Although several studies investigated job role or managerial status, the variables were not consistently applied or measured. However, the three papers that reported on their effects noted positive influences of higher specialization or managerial status, such as increasing protection motivation [288] or perceptions of success [364]. Computer self-efficacy showed mixed effects, in two cases decreasing security compliance intentions [203, 204] and in one case increasing compliance behavior [344]. Finally, a variety of general security awareness or knowledge of specialized systems were positively

associated with security engagement and behavioral intentions (e.g., [92, 203]).

2.4.5 Future study opportunities suggested in the reviewed papers

We systematically analyzed the future work and study limitations discussed by the authors of our reviewed papers. We coded and categorized the extracted 130 suggestions into the following four themes:

Theoretical framework refinement, integration, and testing (n = 33).

Testing and refining theoretical frameworks (both established theories and conceptual models) is essential for advancing cybersecurity research [170, 257]. Existing models, such as SDT and Protection Motivation Theory [257, 287], require ongoing refinement to maintain their relevance in security behavioral research. Researchers should integrate meaningful constructs [170, 257] into their research models to increase the explanatory power. Furthermore, researchers have called for new paradigms in security behavioral research [344, 257]. For instance, Burns et al. [58] advocated for a paradigm shift from Deterrence Theory to theories that emphasize self-control and motivation. Similarly, authors have recommended investigating how organizational commitment [288], autonomy [203], emotion [99], and psychological empowerment [361] are related to security behaviors, as well as the antecedents of these psychological constructs [99, 389]. Additionally, the authors proposed interdisciplinary perspectives for enriching human-centered security research. Tejay and Mohammed [364] argued for the incorporation of theories from anthropology, which could introduce new perspectives to cybersecurity culture that are currently underexplored. Finally, it is crucial to test and apply research models to different types of behaviors to validate their robustness and explainability [287], ensuring that these theories can be utilized to address different cybersecurity challenges.

Methodology improvement (n = 54).

A total of 54 recommendations focused on improving research design, measurement, and data collection. First, many authors advocated for the use of longitudinal research designs to capture changes in behavior over time [235, 277, 92, 170] and to unveil causal relationships between motivators and security behavior. Some authors proposed that qualitative methods should be utilized, such as case studies [53], observation [164], and focus groups [191], to investigate psychological constructs and behaviors in more depth. Several authors emphasized the need to incorporate additional control variables, such as geography [12], education [12], and social desirability [170], to improve the accuracy of research findings. Second, a significant number of authors suggested future work to improve the measurements of motivators [218] and security behaviors [236], including developing more valid

scales [218] and triangulating self-reported data with objective logs [257, 278], as self-reports can induce biases [235]. Third, randomized sampling [136] and the inclusion of diverse participants from different organizations [21] and various job roles (e.g., management, external stakeholders [25]) are recommended to ensure the robustness of the findings. Fourth, to enhance the generalizability of research findings, several authors proposed that studies should be replicated across different organizations [202] and that the research models should be tested with specific ISPs [53]. These suggestions call for rigorous methods, data collection, and the use of both qualitative and quantitative approaches to improve research on information security behaviors.

Examining personal, organizational, industrial, cultural, and contextual differences (n = 21).

To mitigate potential biases in cybersecurity, it is essential to study underinvestigated sectors and demographics to extend beyond heavily regulated industries and traditional geographic regions [377, 22]. More diverse research samples should be used across industries, departments, and occupations to improve the validity of findings [204]. Future research should also explore cultural influences on security behaviors, as factors such as national culture [364] and individualism [287] may significantly impact employees' security decisions. Conducting cross-cultural studies can further illuminate how security behaviors vary across different social environments [364, 129]. Moreover, examining individual differences [58], such as personality traits [236, 99], is crucial for developing a better understanding of employee security behaviors; for example, personality differences might vary their acceptance of intrinsic and extrinsic appeals [236]. Finally, organizational contexts, including policies [138], security task characteristics [99], and leadership dynamics [138], also play a vital role in shaping employee behaviors.

Intervention design and practical application (n = 22)

Many authors suggested that organizations should promote organizational culture and foster an ethical climate that is aligned with personal values [67], create interventions based on specific industry needs [202], and embed security responsibilities within the organizational culture [43]. Encouraging creativity [185] and collaboration [318] in cybersecurity practices might enhance their effectiveness. Several authors proposed that employees' needs for autonomy, competence, and relatedness should be fostered in the workplace to improve their security behaviors [257, 186]. Employees should be provided with the freedom to control their tasks, which can reduce negative emotions and increase compliance intentions [204]. Additionally, some authors suggested that gamification and innovative media should be incorporated into security training to engage employees [338]. The gap between static training and evolving threats was noted, with recommendations for more iterative and dynamic

training approaches [7]. Furthermore, the authors highlighted the role of empowerment in promoting cybersecurity compliance. Empowering employees to participate in decision-making and feel more autonomous was also seen as crucial for enhancing job satisfaction and organizational commitment [186, 99]. These suggestions call for a shift toward more culturally aligned, autonomous, and empowerment-focused approaches in cybersecurity.

2.5 Discussion

We provide an overview of the key findings in Table 2.7. Next, we discuss the development of our taxonomy, practical implications of our review, and suggestions for future studies.

2.5.1 A taxonomy of autonomous motivation related to organizational security behaviors

Reflection on the role of theories in our work: We began with the definition of autonomous motivation proposed in the SDT framework [97]. After analyzing the reviewed papers, we found that neither the SDT motivation continuum [313] nor the SDT-ER taxonomy [131] could accommodate the five groups of autonomous motivators identified in the reviewed empirical studies. Therefore, we introduced two core constructs — intrinsic (task) value and expectation, from Expectancy-Value Theory [111] — into the SDT-ER taxonomy and integrated psychological needs fulfillment as an additional reason for behavior. See the taxonomy of autonomous motivation related to organizational security behaviors in Figure 2.4. The following rationales support our adaptation:

- The reviewed studies [377, 186] suggested that job satisfaction has a positive association with ISP compliance behavior. [377, 186] defined job satisfaction as the pleasurable or positive emotional state resulting from one's job or job experience. Job satisfaction cannot readily be categorized into the three categories in the SDT-ER taxonomy, i.e., usefulness-driven, value-driven, and interest-driven motivators. We propose a new category "task-driven" to accommodate job satisfaction because it emphasizes the enjoyment individuals derive from an activity [111]. Thus, job satisfaction and *intrinsic task value* were incorporated into the taxonomy under the task-driven category.

- In the SDT framework, "satisfying human needs for competence, relatedness, and autonomy creates sustainable (i.e., enduring) motivation" [352, p.77], namely, autonomous motivation. However, multiple reviewed studies have suggested that these motivators and protection motivation are *directly* related to employees' security behavioral intentions, not necessarily

Table 2.7: Summary of key results.

Research Questions (RQs)	**Key results**
RQ1. Forms of autonomous motivation (Section 2.4.1)	We developed a refined *taxonomy of autonomous motivation* related to security behaviors and categorized 17 unique motivators into five groups: interest-driven, task-driven, psychological needs fulfillment, value-driven, and expectation.
RQ2. Related security behaviors (Section 2.4.2)	The most examined security behavior in the reviewed studies were ISP compliance behavior/intention (n = 24), followed by extra-role security behaviors (n = 22). Additionally, three studies investigated employees' ISP violation behaviors.
RQ3. Applied theoretical frameworks and findings (Section 2.4.3)	*24 theoretical frameworks* have been applied to study autonomous motivators. The most frequently applied theories were Self-Determination Theory (n = 16), Protection Motivation Theory (n = 7), and Theory of Planned Behavior (n = 6). The most frequently examined autonomous motivators were competence and personal responsibility, with mixed results (some studies found statistically significant effects, while others did not). *Commitment, perceived value congruence, information security identity, perceived benefit,* and *intrinsic task value* were less studied but seem to be positively related to several security behaviors. Autonomous motivators have *positive relationship* with ISP compliance behavior/intention and extra-role security behaviors, and some motivators (i.e., response efficacy, personal responsibility, and autonomy) *are inversely related to* ISP violation behaviors. The majority of studies in our review adopted deductive approaches (n = 37), with only *three* employing inductive methods and *five* using design-based approaches. Among the 45 papers reviewed, we found only *one* longitudinal study (six months) and *one* replication study.
RQ4. Study contexts (Section 2.4.4)	*Most examined regions* were: North America (n = 17), Europe (n = 7), and East Asia (n = 7). Education, finance, government, and healthcare were the most studied *sectors*. Of the 40 papers employing quantitative analyses, *only 20* reported the use of control variables. Age, gender, education, and job roles might influence employees' security behaviors.
RQ5. Promising avenues (Section 2.4.5)	Theoretical framework refinement, integration, and testing. Methodology improvement: longitudinal designs, more qualitative studies, better measurements and sampling, inclusion of control variables, and replication studies. Examining personal, organizational, industrial, cultural, and contextual differences. Intervention design and practical application.

through moderators [21, 408, 288]. Thus, we created the new category in our taxonomy:

"psychological needs fulfillment."

- We proposed to use "expectation," instead of usefulness-driven, to categorize response efficacy and perceived benefits to express their future-oriented nuance. According to Expectation-Value Theory, *expectation of success* is one of the core constructs that directly influence individuals' choice of performing an activity [111, 70]. The measures of response efficacy and perceived benefits assess employees' anticipated outcomes (e.g., securing the workplace network [42]) of the recommended security actions using future-oriented language. These outcomes are neither static nor aligned with the concept of "usefulness" in relation to employees' job roles. For this reason, we integrated expectation into our taxonomy.

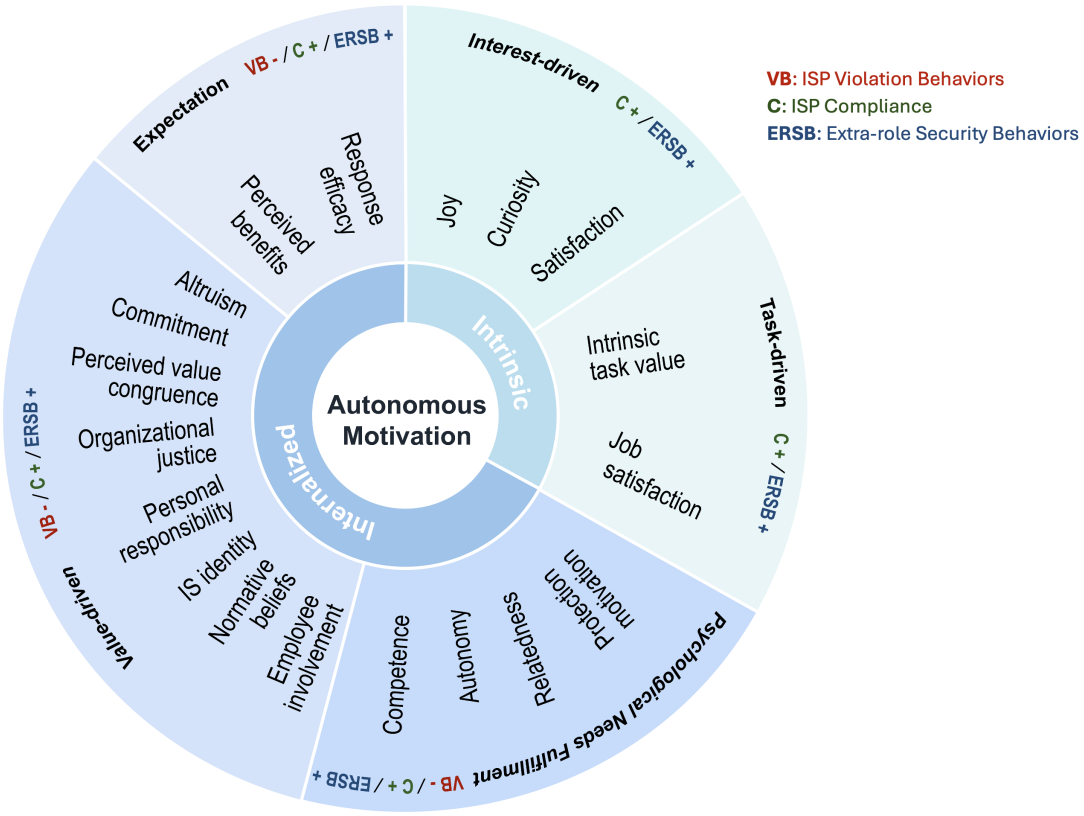

Figure 2.4: The taxonomy of autonomous motivation related to organizational security behaviors (based on the work of [131, 97]; -/+ indicates that at least one motivator from the located category is negatively (-)/positively (+) related to the behavior).

This review makes a theoretical contribution by clarifying a construct emphasized by multiple theories and carefully aligning the findings from reviewed empirical studies with established theories. This approach resulted in a theoretically robust and practically relevant taxonomy, demonstrating the potential for fostering autonomous motivation and promoting security behaviors in organizations. Behavioral security scholars can draw inspiration from studies that employ relevant theoretical

frameworks. For example, through the lens of SDT, the *social condition and process* that provide rationale for the activity, acknowledge individuals' perspective and feelings, and support their experience of choice (while minimize the use of pressure) foster autonomous motivation [96]. In practice, interventions targeting management stakeholders have been carried out to change organizational climate, with the goal of improving employees' satisfaction and trust in the organization [339]. Further, task challenges *at moderate levels* strengthens employees' autonomous motivation, hence, simulating their work-related well-being [359]. These findings could be leveraged to improve cybersecurity management and security task design at the workplace.

How autonomous motivators influence employees' organizational security behaviors: Applying our taxonomy to analyze the findings from reviewed studies, we found that all five categories of autonomous motivation demonstrated positive statistically significant relationships with employees' intentions to comply with ISPs (see Table 2.4). [338] achieved a significant result with a design-based approach (not included in Table 2.4). This suggests that all categories of autonomous motivators might positively influence employees' intentions to comply with ISP in organizations.

Among all the motivators, personal responsibility and competence-related motivators were examined the most, with mixed results (some studies demonstrated significance, whereas others did not; see Table 2.4). Given the diverse demographics and contexts of the papers we reviewed, this result requires further scrutiny. Other motivators, such as commitment, perceived value congruence, IS identity, perceived benefit, and intrinsic task value were less studied but were positively associated with several security behaviors.

Autonomous motivators were not only positively associated with compliance behaviors and extra-role security behaviors but also inversely related to violation behaviors. Specifically, personal responsibility, response efficacy, autonomy, and perceived benefits were found to be negatively associated with employees' violation behaviors (see Table 2.4). Only three studies investigated employees' violation behaviors, partly due to a scarcity of research on employees' risky cybersecurity behavior [17]. These previous findings suggest that autonomous motivation may help reduce employees' violation behaviors.

2.5.2 Practical implications

Human-centered security and autonomous motivation

Autonomous motivation is instrumental to design user-centric security policy and interventions. One prominent theme of human-centered security is the user-centered design of security mechanisms. *User needs* have long been emphasized as a primary design goal when developing usable and secure systems

[418]. Security mechanisms and policies that fail to consider employees' job contexts, the feasibility of organizational strategies, and usability issues reduce employees' *motivation to engage* with security measures [1]. Nevertheless, studies show that security measures still cause frictions [34] and security officers "regularly shift responsibility either to the management (by demanding more support) or to the employees (by blaming them)" [177, p.2311]. Other streams of research within the community examined the behavioral aspects of organizational security, such as the learning curve of security behavior [321, 176], designing tools to aid security tasks [39], and creating engaging experiences for security learning [133]. The autonomous motivation reviewed in this study can influence an employee's decisions on whether to perform a security behavior when they are capable and to engage with non-mandatory security tasks. When security mechanisms align with employees' interests, tasks, psychological needs, values, and expectations, the associated behaviors are more likely to be accepted and maintained.

Empowering employees

Multiple authors [386, 99, 415] have suggested empowerment as a complementary measure to technical measures and sanctions in promoting compliance and preventing ISP violations. *The empowerment approach* emphasizes that people's strengths and abilities should be identified and built upon, rather than blaming them for their difficulties [285]. Empowerment can influence employees' security behaviors through different approaches [99, 70, 202]. First, psychological empowerment positively influences employees' ISP compliance intentions [361]. Second, empowerment also informs the design of management tools [202]. For example, an empowerment-based management system has demonstrated the potential to minimize the circumvention of access rules [202]. Finally, when an employee feels empowered, they are more likely to engage in proactive security behaviors, such as becoming security champions, i.e., proactive security advocates who often have a good knowledge of security practices and can promote security culture among employees [135, 156, 360, 258]. While research showed that this promising approach faced certain challenges in the past, such as the selection of appropriate people [35, 135] and difficulties arising based on a lack of management support [156], our findings might inform criteria for the successful implementation of a security champion program. For example, people with high autonomous motivation might be good candidates for becoming security champions. To foster or maintain that motivation, security champions should not only be appointed without further rights, but also be enabled to act and be included in the development and discussion of security measures.

Security training programs, access to security strategies, and inclusion in decision-making have been shown to enhance employees' psychological empowerment, as highlighted by [361]. Moreover,

employees' perceptions of managerial practices, workplace support, leadership, and work design characteristics significantly impact their sense of empowerment [330]. Building on these findings, our taxonomy of autonomous motivation could be a useful tool for organizations aiming to enhance the perceived fairness of their information security policies, align work environments with employees' internal values, and foster a stronger information security identity. By integrating autonomous motivation into these efforts, organizations can not only empower their employees but can also promote employees' engagement in security practices. In the following subsection, we use the designing of security training programs as an example to illustrate how the taxonomy of autonomous motivation can be applied in practice.

Evaluating and improving security training programs

Security managers deploy "a combination of tangible activities, material delivery, and ongoing engagement" with the goal of raising employees' security awareness [178]. Employees often perceive these activities as a burden and disengage with security training [70]. Some researchers [325, 33] suggest using tools such as scenario-based surveys, HAIS-Q, or Security Attitude Inventory (SA-13) to differentiate employees and deliver targeted interventions for specific employee groups, to avoid burdening employees with unnecessary interventions. Others propose to improve the training design to engage employees with intrinsic motivation being frequently applied as a guiding principle in designing awareness campaigns [69], particularly in gamified training [338, 369]. This approach often integrates intrinsic motivators such as joy, curiosity, and satisfaction, into the learning experience and has demonstrated its effectiveness [338, 397]. As noted by Bennett and Mekler, in the Human-Computer Interaction (*HCI*) and user experience (*UX*) communities, "very little attention has been paid to motivational factors related to the outcomes of the activity, and how these relate to the values and goals users bring to the interaction" [37, p.26]. Which opportunities exist for organizations to apply autonomous motivators to evaluate and improve security training?

Security managers can apply the constructs from the taxonomy of autonomous motivation to audit the training (e.g., perceived benefits: "How well does the training benefit employees and enhance their knowledge in protecting against attacks?"). They can utilize motivators to collect employees' feedback on the training (e.g., competence: "To what extent do you feel the training has improved your ability to identify incoming threats?"). Involving employees in the training design process and ensuring organizational transparency and fairness can improve their engagement. Additionally, management can support employees by addressing technology-related frustrations (leading to higher job satisfaction) and encouraging self-development in technology use (e.g., enhancing personal competence and fulfilling curiosity). Moving forward, organizations should consider conducting regular evaluations

of their security awareness campaigns to ensure ongoing relevance and effectiveness [178].

Table 2.8: Observed challenges and our recommendations in conducting theory-informed studies.

Observed challenges	Recommendations
Theories	
• Naming theories with terms different from those used in cited sources.	• Maintain consistency in terminology by using the same names for theories as those found in cited sources.
• Introducing one theory in the related work section while using another theory for measurement.	• Ensure consistency by clearly explaining the choice of theory in both the related work and measurement sections.
• Categorizing autonomous motivators under terms that differ from the theoretical framework, such as classifying them as intrinsic motivation when based on SDT.	• Use terminology that accurately reflects the theoretical framework and definitions of the constructs.
• Stating theoretical propositions without correct or sufficient citations.	• Provide accurate citations to support all theoretical propositions, ensuring the relevance and quality of the sources.
• Lack of linkage between the introduced theory and the proposed research model.	• Clearly present the research model and state how the theory informs the model.
Measurement	
• Using the same measurement as the cited source, but giving it a different name.	• Standardize measurement terms across studies and provide clear explanations to avoid confusion.
• Names of the concepts/constructs do not intuitively match items used for their measurement.	• Define concepts and constructs clearly to match the items used for their measurement.
• Removing or adding items to measurements without providing a reason.	• Justify any changes to measurement items with clear reasoning and documentation.
• Cited sources cannot be retrieved, and items were not included.	• Include all items used in the Appendix of the study to maintain transparency and reproducibility.

2.5.3 Looking into the future of autonomous motivation in human-centered security

Recommendations for conducting theory-informed studies

Throughout our review, we encountered the following challenges at least once (see Table 2.8). However, we deliberately avoid pointing out individual papers to maintain a focus on providing constructive and future-oriented recommendations for avoiding pitfalls and enhancing the transparency and replicability of research.

Future avenues

Our analysis of suggestions in the reviewed papers identified four future directions for security behavior studies (see section 2.4.5). Regarding publication venues, information system journals seem to be in favor of deductive methodology and studies with clear theoretical contributions. Much of the reviewed design-based and inductive studies were published in interdisciplinary journals, as well as security and privacy venues that commonly accept HCI studies. Only three of the reviewed studies examined beyond general roles, that is, cybersecurity specialists [164, 185] and security managers [22]. Stakeholders that design cybersecurity policy and manage cybersecurity tasks (e.g., CISOs and system administrators) are under-represented in our review. Ensuring organizational security is their primary task, unlike most employees for whom security is usually a secondary task [398]. Security professionals usually have higher cybersecurity expertise as compared to general employees. Based on these differences, it would be highly relevant to compare security professionals' and general employees' motivation toward security tasks. However, professionals are much more limited in numbers as compared to general employees and hence harder to reach. For example, previous research on security professionals sometimes relied on computer science students as an alternative (e.g., [269]), involved high payment for professional work included in the studies (e.g., [270]), or had limited sample sizes (e.g., [86, 179]). Despite these challenges, insights from this comparison could guide the design of security measures that reduce the conflicts between primary and secondary tasks.

Our systematic review also allowed us to make methodological recommendations for future research. Most of our reviewed studies employed surveys and only five studies used design approaches (including one longitudinal intervention study [338]), three exploratory studies, and one replication study among 45 papers. We encourage future research to include more qualitative studies, intervention studies, longitudinal designs, and the development of tools (e.g., [336]) to support organizational security practices. Whereas subjective measurements from self-assessment questionnaires remain the most commonly used method [210], incorporating objective assessment methods could offer valuable insights. For instance, eye gaze data might provide an objective measure to complement others [195, 18, 275]. However, as eye gaze data are often challenging to interpret on their own, researchers commonly supplement eye-tracking data with methods such as Retrospective Think Aloud [152]. These methods often require manual and labor-intensive segmentation and labeling of the data, which can be especially daunting for large or complex datasets [388]. While automating such processes shows promise, they still lack in precision and contextual understanding [378, 275]. Scenario-based assessments [155] and the triangulation of data from multiple sources—such as management evaluations [377], self-reports, and system log data [338]—show promise in providing a more complete understanding of security behaviors. However, there are noteworthy challenges regarding both data collection and analysis of

behavioral data in the field. Accessing organizational log data often requires researchers to collaborate closely with the organization's security officers [72, 407]. Additionally, the sensitivity of behavioral data demands rigorous processing protocols that comply with local data protection laws. Coordination among the legal team (e.g., for non-disclosure agreements), the data protection office, and the Ethical Review Panel at the research institute can take months.

Previous reviews [234, 17] have indicated that the Theory of Planned Behavior, Protection Motivation Theory, and Deterrence Theory are the ones that have been examined most extensively in security behavior research, primarily focusing on competence, threat appraisal, and deterrence (see section 2.2.1). However, our review on autonomous motivation highlights the increasing relevance of SDT. This shift in theoretical frameworks reflects how different theories are driven by specific perspectives, demonstrating the adage, "What you look for is what you find." When research is limited to exploring how threat appraisal leads to protection motivation, it inevitably reinforces those findings, leaving little room to explore the influence of psychological needs and other motivational factors. Future research should further evaluate other frameworks, such as SDT and Expectancy-Value Theory, in the security context, to continue moving beyond deterrence. Additionally, there is the potential to integrate autonomous motivation with 20 other less-explored theories (Appendix A) in the security domain. The application of these theories could further capture the complexities of organizational security behaviors.

2.5.4 Limitations

Our review converges with recent security behavior studies within the HCI community, such as "self-efficacy and security behavior" [45], "cognition in social engineering empirical research" [56], and "emotions in cybersecurity" [385]. However, due to the scope of this review, we did not investigate the relationship between autonomous motivation and other influencing factors examined in [45, 56, 385]. Future studies can synthesize findings from these recent studies and this review and comprehensively examine these factors with specific security behaviors. While we proposed a refined taxonomy, the SDT framework's definition still captures the essence of autonomous motivation. Our taxonomy is not exhaustive, as it builds on previous taxonomies and the findings of reviewed papers. We encourage future research to validate our taxonomy and examine the interactions among motivators through empirical studies.

2.6 Conclusion

Scholars have suggested that autonomous motivators hold untapped potential in promoting security behaviors without relying on controlled motivation alone [257, 344]. Prior work has used a variety

of theoretical frameworks from various disciplines to study autonomous motivators, leading to fragmented and heterogeneous literature. It is unclear how autonomous motivators connect with security behaviors.

To reconcile scattered findings, we systematically reviewed and analyzed 45 empirical studies examining autonomous motivation in organizational security contexts. We propose a refined taxonomy of autonomous motivation related to organizational security behaviors. We identify three types of security behaviors that have been examined in relation to autonomous motivation, synthesize findings and suggestions from the reviewed studies, and chart a path for conducting theory-informed studies on autonomous motivation in human-centered security.

2.7 Data Availability Statement

We have included the anonymized preregistration for peer review [74]. The data supporting the findings of this paper, including the preparation phase extraction, search query results, extraction manual, and the extraction table, are available as supplemental material.

2.8 Acknowledgments

Author 1 acknowledges the financial support of the Institute for Advanced Studies at the University of Luxembourg through a Young Academic Grant (2021). The Doctoral School in Humanities and Social Sciences at the University of Luxembourg supported the project with the Research Support Grants for 2024 and 2025. We thank Sophie Doublet and Muriel Frank for their support in discussing and visualizing autonomous motivation. We thank the ACs and reviewers for their constructive feedback.

Chapter 3

What Motivates and Discourages Employees in Phishing Interventions: An Exploration of Expectancy-Value Theory

Published as: Chen, X., Doublet, S., Sergeeva, A., Lenzini, G., Koenig, V., & Distler, V. (2024, August). What Motivates and Discourages Employees in Phishing Interventions: An Exploration of Expectancy-Value Theory. *Proceedings of the Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*. USENIX, Berkeley, United States.

Abstract Organizations adopt a combination of measures to defend against phishing attacks that pass through technical filters. However, employees' engagement with these countermeasures often does not meet security experts' expectations. To explore what motivates and discourages employees from engaging with user-oriented phishing interventions, we conducted seven focus groups with 34 employees at a European university, applying the Expectancy-Value Theory. Our study revealed a spectrum of factors influencing employees' engagement. The perceived value of phishing interventions influences employees' participation. Although the expectation of mitigation and fear of consequences can motivate employees, lack of feedback and communication, worries, and privacy concerns discourage them from reporting phishing emails. We found that the expectancy-value framework provides a unique lens for explaining how organizational culture, social roles, and the influence of colleagues and supervisors foster proactive responses to phishing attacks. We documented a range of improvements proposed by employees to phishing interventions. Our findings underscore the importance of enhancing utility value, prioritizing positive user experiences, and nurturing employees' motivations to engage them with phishing interventions.

3.1 Introduction

Phishing was the most reported cybercrime in the U.S. between 2019 and 2022 [121]. Phishing emails deceive people into clicking on malicious links, disclosing sensitive information, or installing malware on their devices [9]. Phishing attacks endanger organizational intellectual property and institutional reputation, causing billions of losses [184, 15, 121]. Organizations employ a range of measures to defend against phishing attacks. Despite the implementation of technical filters, even if deep learning models achieve an accuracy rate of more than 96% [140, 27], a substantial number of phishing emails still end up in employees' inboxes. While technical solutions play a critical role in mitigating phishing attacks, employees are the last line of defense in organizations [245].

To raise employees' security awareness and educate them about phishing attacks, some organizations deploy online security courses as a cost-effective way to educate their employees [88]. Some organizations utilize simulated phishing tests in an attempt to track whether employees can identify phishing emails [57, 106]. Further, organizations broadly advocate for employees to report phishing emails, which enables IT teams to promptly detect incoming phishing attacks [229]. Research suggests that phishing interventions promote safe responses to attacks [409, 222], and reporting can serve as an effective crowd-sourced approach to counteract phishing [229, 72]. However, these user-oriented phishing interventions are not always embraced by employees [308, 227], as participation in the interventions requires time and effort and can interrupt the working routine [216, 133].

Motivation theories from educational psychology can be useful in explaining employee's (dis-)engagement. Recently, Expectancy-Value Theory (EVT) has received attention from scholars working in information management [367]. EVT seeks to explain individual behaviors with two central constructs: "expectation of success" and "subjective task value" [78]. We find these constructs particularly relevant and under-investigated in security behavior studies [80].

In this paper, we examine employees' engagement with *phishing awareness campaigns*, which include online security courses and simulated phishing tests, as well as *reporting phishing emails* through the lens of EVT. By deepening understanding of the influencing factors associated with phishing interventions, organizations can improve their implementation of these interventions. We pursue the following objectives: 1) examining factors that motivate and discourage employees from engaging with phishing interventions, and 2) exploring what could be improved to increase employee engagement with these interventions. Focus groups are a qualitative method frequently applied to elicit ideas [402] and confront different viewpoints [401]. Educational institutions are frequently targeted by cybercriminals in recent years [128, 370]. Examining factors that influence university employees' engagement with phishing interventions is highly relevant to the current threat landscape. In light of this, we conducted seven focus groups with 34 employees (including research and non-research roles) in a European university.

Contributions. This paper makes empirical contributions, providing an enriched understanding of how various factors influence employee (dis-)engagement with phishing interventions. Our findings and adaptation of EVT suggest that it is a valuable theoretical framework for explaining how motivational factors influence employees' engagement with phishing interventions, highlighting its potential as a framework for future security behavior studies. This paper makes a theoretical contribution and highlights the possible adaptations to EVT for future use in organizational cybersecurity. Additionally, we offer practical suggestions for improving phishing awareness campaigns and reporting procedures in organizations, advocating user-centric approaches.

3.2 Related work

3.2.1 Phishing awareness campaigns

Simulated phishing tests are a tool for both assessment and educational purposes at organizations [177, 106]. Prior studies primarily utilized employees' click-through and reporting rates in phishing tests as indicators of employees' security behavior and their resilience to phishing attacks [106, 409, 222]. A recent case study highlighted that conducting simulated phishing tests at an organization requires significant time and effort from different stakeholders [50]. Moreover, some organizations have experienced side effects from phishing tests that have burdened CISO's relationship with employees [177]. When organizations neglect privacy concerns, fail to receive approval of simulated materials, don't specify the purpose of tests, or withhold appropriate feedback, it can lead to negative reactions from employees [308]. Phishing tests also increase employees' workload, potentially making them more susceptible to phishing attacks [50, 308]. Brunken et al. suggest involving employees in future research to better understand how simulated phishing tests impact them and their overall productivity at the workplace [50].

A variety of formats have been introduced to engage individuals with online security training [188]. Comic and game-based online trainings have reported notably high levels of satisfaction in user evaluations [223, 397]. A meta-analysis revealed that trainings combining text and comics demonstrated large effects in reducing victimization compared to comics or game-based trainings [54]. Online phishing quizzes, such as jigsaw puzzles, effectively improved participants' skills in detecting phishing emails [395]. Volkamer et al. created and evaluated a five-minute phishing awareness video, which significantly enhanced participants' ability to recognize phishing attempts both immediately and after an eight-week interval [382]. User feedback praised their video's clarity and simplicity, with suggestions for more phishing examples and a concluding summary [382]. Anti-phishing training utilizing storytelling led to higher levels of curiosity, self-efficacy and phishing detection ability than training employing comics in an online experiment [190]. To improve the effectiveness of security trainings, both the content and format of trainings were re-designed to engage learners.

While some studies suggest that offering educational materials after simulated tests improved employees' safe responses to phishing [222, 409], there are concerns about the effectiveness of this embedded training approach [50, 229]. Kumaraguru et al. found that employees who trained with anti-phishing materials after clicking links in simulated phishing emails exhibited a decreased likelihood of clicking on links in subsequent phishing tests compared to their untrained colleagues [222]. Yeoh et al. reported that the immediate provision of anti-phishing materials following phishing tests led to more safe responses than merely administering phishing tests [409]. Despite these

findings, researchers suggested that only a small percentage of employees who clicked the phishing tests subsequently engaged with training materials [50, 102]. Thus, further investigation is required to better integrate simulated phishing tests and online security courses.

3.2.2 Phishing email reporting

Recent studies have begun to investigate factors that influence individuals' intention to report phishing emails. A survey with American college students [227] revealed that perceived self-efficacy, expected negative outcomes (*concern for mishandling of reports of spear phishing emails*), and cybersecurity self-monitoring increase the likelihood of reporting spear phishing emails. In alignment with [227], Kersten et al. suggested that user's intention to report phishing emails was negatively associated with the perceived "believability of the email" (the extent a user considers the email to be credible) in an online controlled experiment [214]. In an in-situ deception study [102], Distler found that employees' motivations for reporting phishing included improving email filters and receiving positive feedback. Obstacles to reporting entailed uncertainties regarding the reporting process and rationale, coupled with concerns about "getting colleagues into trouble" for sending legitimate emails that were misinterpreted as phishing attempts. Additionally, participants believed that reporting became redundant once they had clicked on the link in a simulated phishing email [102]. In a survey with US workers, factors such as self-efficacy, subjective norms, and altruism tendencies increased reporting intention. Conversely, "sportsmanship" hinders individuals from reporting phishing emails [246]. Other than utilitarian motives, Franz proposed that the design features and risk indication influence participants' acceptance of reporting tools and suggested further research into the role of hedonic motives in the reporting process [132]. Additional factors may influence an individual's intention and behavior regarding the reporting of phishing emails, warranting further investigation.

3.2.3 Theoretical models applied to study user security behavior

Prior research on user security behaviors has frequently focused on fear appeals, as seen in studies that examine the constructs of Protection Motivation Theory (PMT) [158]. PMT explains protection behavior through two processes: threat appraisal and coping appraisal. In threat appraisal, people evaluate their perceived vulnerability and the perceived severity of a situation, while coping appraisal entails assessing response efficacy, self-efficacy, and response cost [309, 376, 287]. However, there are limitations and constraints in applying PMT to study user security behaviors. Originally constructed to explain health protection behaviors, PMT is based on the premise that the threat is relevant to the individual; however, this might not be the case in the information security context [257]. In a Relative Weight Analysis, attitude, personal norms & ethics, and normative beliefs demonstrated the

highest effect sizes and relative importance in explaining security compliance behaviors, emphasizing employee psychological and ethical traits [80]. These constructs are not included in the theoretical model of PMT. To overcome the limitations of PMT, recent studies have begun to integrate constructs from other motivational theories to examine user security behaviors [158].

Expectancy-Value Theory (EVT) [108] is an influential motivation theory in educational psychology [168]. According to EVT, individuals' beliefs about how well they will do on an upcoming task and the subjective values they attributed to it influence their engagement with the task [111] (see Appendix C for the core constructs of EVT). EVT shares the same theoretical root as PMT, as both theories developed from Atkinson's expectancy-value model [309, 399]. EVT examines individuals' anticipation and subjective task values in educational contexts [399], whereas PMT employs fear appeals to motivate protective actions in health management [309]. However, EVT has rarely been applied to security behavior studies [80]. In an experiment incorporating EVT constructs, Jenkins et al. found that the highest levels of security behavior were associated with minimal technical controls (*number of passwords a participant was forced to use and remember*) combined with security education [199]. A recent structural modeling study that applied EVT revealed that achievement, along with intrinsic and extrinsic motivations, are determinants in explaining the motivational values associated with users' intention to protect mobile identity [12]. Applying EVT to investigate the factors that influence employee engagement with phishing interventions appears promising.

3.2.4 Research objectives

Low employee engagement with phishing interventions continues to be an obstacle to achieving information security in organizations [227, 409]. EVT has been utilized to examine learners' motivations in various contexts, including organizational [187, 64]. Applying EVT can elicit employees' motivational factors associated with phishing interventions. Further, beliefs and values form attitudes in the cognitive process, which in turn guide behavioral responses [199]. Expectation and subjective task values directly influence people's choices and performance in the EVT framework (see figure 3.1). Consequently, we propose to utilize EVT constructs to address the following research questions (**RQ**):

RQ1: Which factors motivate employees to engage with phishing interventions?

RQ2: Which factors discourage employees from engaging with phishing interventions?

RQ3: From the employees' perspective, which aspects of phishing interventions could be improved?

3.3 Study design

We conducted focus groups with 34 employees at a European university to address these research questions. Focus groups are a form of interviewing where multiple participants come together to express and deliberate on their views regarding a predetermined topic in a collective discussion [104]. Focus groups are especially useful for gathering diverse and in-depth perspectives from interactions among participants [134, 401, 402], allowing us to gain an exhaustive understanding of the factors influencing employees' engagement with phishing interventions.

3.3.1 Study context

The study was conducted at a research-oriented European university that employs approximately 3,900 individuals. 38% of them are employed in research roles, whereas the remaining employees fulfill administrative functions. The organization uses a phishing awareness campaign sourced from a security service company. The IT team sends a simulated phishing test to all employees via the management software on a random date each month. Employees who click the link or download the attachment within the phishing test land on a page displaying "you clicked on a simulated phishing test" and "rules to stay safe online". Afterwards, the IT team sends a web link to online security courses to those who responded unsafely. Employees who reported the simulated email to the IT team receive an automatic reply within a couple of minutes with the subject line "congratulations, you've spotted a phish".

To raise phishing awareness, the IT team sends every new employee an email during their first week that includes links to online security courses and suggested responses to suspicious emails. To defend the organization against phishing attacks, the IT team encourages employees to report any suspicious emails to "report-a-phish@anonymized". When the reported email is a simulated test, a program automatically sends out a reply; otherwise, a security expert manually reviews the reported email. Normally, it takes one or two working days for the expert to reply with the verification result of the reported email. When a reported email is a phishing attempt, the expert sends a phish alert to individuals who also received the phishing attempt. When the email is legitimate (not a phish), the expert replies with "It is a legitimate email".

At the time of our investigation, all employees automatically received simulated phishing emails as part of their cybersecurity training without prior informed consent. Employees could either actively engage by reporting the simulated test in accordance with the organization's suggestions for handling suspicious emails or ignore these simulated tests.

3.3.2 Participants

We used multiple approaches to recruit study participants, including posters across three administrative buildings, LinkedIn posts, email invitations, and direct outreach. Forty-five employees registered their interest in participating in our study. We assigned them to different groups based on the similarity of their job roles and the diversity of faculty. We did not exclude any specializations (e.g., computer scientists) when scheduling our focus groups. Due to personal reasons, 34 of the 45 interested employees participated in seven focus group sessions (20 female, 13 male, and one non-binary) between November 2022 and January 2023. Each session consisted of three to seven participants. Participants included 19 researchers, 12 administrative staff, and 3 software developers. On average, the research staff had worked at the organization for 1.3 years (SD=0.9), and the non-research staff 7.3 years (SD=6.7). The participants' age ranged from 25 to 56 years (mean=37.6, SD=10.8). In the demographic questionnaire, 32 (94%) participants indicated that they had encountered phishing attacks previously; 29 (85%) had received simulated tests from the IT team[1]; 25 (74%) had reported phishing emails to the IT department, and 14 (41%) had previously participated in online security courses. We include the participant demographic information in Appendix F.

3.3.3 Procedure

Prior to data collection, we conducted two pre-test sessions (N=11) to refine our protocol. During the first pre-test, we led the discussion using a synthesized framework of motivation theories [168]. Introducing concepts from multiple theories led to cognitive overload for participants during the focus group. In the second pre-test, we narrowed our focus to EVT. According to the preliminary analysis, observations, and participants' feedback on the pretests, we improved our discussion questions and added templates and brainstorming activities. The revised focus groups included four parts: a warm-up activity, a group discussion, a brainstorming activity, and the debriefing. Each focus group took approximately 90 minutes.

First, we conducted a warm-up activity to familiarize the participants with the lab and to elicit what motivates and discourages them from engaging with a self-selected leisure activity through *Template 1*. This stage lasted for 10 minutes.

In the second part, the participants were involved in a group discussion on phishing awareness campaigns for 25 minutes. Then, we instructed them to complete *Template 2* to record their motivating and discouraging factors for reporting suspicious emails. Following this, participants continued discussing the factors influencing their reporting. This stage planned a total of 60 minutes and

[1]Every employee is scheduled to receive a phishing test monthly. These five employees, who reported not receiving any phishing tests, may have simply not clicked on or noticed the tests.

included 12 questions to examine *general opinions, self-concept of their ability, goal setting*, and *role identification*, as well as their subjective task value (*costs, benefits*) related to participating in phishing interventions. These questions were adapted from the core concepts within EVT framework that affect individual's choices and performance (see Figure 3.1).

In the third part, participants were asked to brainstorm as if they were the new chief information security officer in response to an increase in phishing emails targeting the university. Participants were tasked with designing strategies to engage employees with phishing interventions in groups. This round lasted 15 minutes.

Lastly, the participants were debriefed by introducing the standard practices suggested by the IT department to avoid any misunderstandings caused by opinions mentioned during the discussion. We provide the *two templates* and full focus group *protocol* in Appendix D.

3.3.4 Data collection and analysis methods

We recorded audio and video of the focus group sessions. We used the audio recordings (11 hours in total) for the analysis[2]. The audio was transcribed automatically using Microsoft Word and reviewed to ensure accuracy. We pseudonymized the transcripts to protect the identity of participants prior to analysis.

The answers to "Template 2. What motivates/discourages you from reporting" were transcribed into an Excel spreadsheet. The first and second author then independently coded the template, following a thematic analysis procedure [76]. Then the two authors categorized the generated codes into preliminary groups in a discussion, which yielded an initial set of codes. Concurrently, a coding workshop was conducted with five researchers experienced in qualitative research and coding. This workshop, which employed an inductive approach [142], analyzed the transcripts from two focus group sessions. Consequently, a second set of codes was created. By integrating the template codes with those from the workshop, the first author established a code system in MAXQDA [379]. The code system was reviewed and revised by three authors. All transcripts were subsequently coded by the first author using MAXQDA. Theme saturation [297] was reached after completing the coding of data from the sixth group. The second author thoroughly reviewed all coded transcripts for consistency and accuracy. A few disagreements were resolved before the final summary of findings via discussion between authors and reviewing the context of the coded segments. We include our coding scheme in Appendix E.

[2]Videos were recorded with the lab's default system as a backup resource in case of audio disruption and were deleted after transcription.

3.3.5 Ethical considerations

The study received approval from the university's ethics review board prior to the pretest. We emphasized that "the session is strictly confidential" to assert peer confidentiality in the email confirmation prior to each session. All participants were informed of their right to withdraw both during and after the study and provided informed consent. The raw data collected in this study were kept confidential to the researchers and stored in line with the General Data Protection Regulation (GDPR) and the ethical guidance of the research institution. Each participant received a €40 gift voucher as compensation for their 90-minute participation. We only used pseudonymized data for analysis.

3.4 Results

We present the factors thematically according to the core concepts of EVT framework and highlight those that could not be located within the framework (see Table 3.1). Unlike qualitative data from individual interviews and open-ended questionnaires, the factors emerging from focus group conversations represent a co-creation among participants. There were occasions when participants filled in specific factors in the template (e.g., P28: "being a good citizen") but did not mention them during discussions, or situations where a factor was articulated in depth by one participant, leading others to choose not to repeat it. Providing the frequency of each theme mentioned by participants would thus not be meaningful.

3.4.1 Phishing awareness campaigns

Factors that motivate employees

Gaining phishing knowledge and *enhancing phishing awareness* are the two utility values mentioned by many participants. They noted that the awareness campaign demonstrated that phishing attacks are constantly changing and evolving. They learned that it is critical to remain informed of evolving phishing techniques, which can support their decision-making in responding to suspicious emails. Additionally, phishing campaigns keep them vigilant of phishing attempts in their daily work. Not only beginners who were not tech savvy could benefit from the campaigns but also experienced employees could be reminded that they need to be cautious of contextual factors. As P2 stated, "even if you're aware of the problem and know how to check . . . you can still fall for it (phishing test) if you don't pay attention, if there's a lot of stress and you're going faster." Additionally, a few participants considered participating in phishing campaign to be a game (P8), and some parts of the online training were "awesome" and "*fun*" (P26).

Table 3.1: Motivational factors associated with phishing interventions.

	Phishing Awareness Campaigns		**Report Phishing Emails**	
	Motivating	**Discouraging**	**Motivating**	**Discouraging**
Expectation	Cyber safety	Optimism bias	Expectation of mitigating, Fear of consequences	Lack of feedback, Lack of communication
Utility value	Phishing knowledge, Phishing awareness	Perceived low value, Lack of incentive	Protecting oneself, Safeguarding the workplace	Low utility value
Intrinsic value	Fun	Lack of interest	Enjoyment, Satisfaction, Pride	
Attainment value		Other priorities	Core values	
Cost		Time constraint, Interrupting workflow, Opportunity cost, Negative inference	Easy to report	Usability issues, Worries and privacy concerns
Competence	Acquiring skills	Overconfidence	Empowerment	Low self-efficacy
Social identity			Recognition, peer influence, sense of belonging	
Goal	Personal development			
Self-schemata		Procrastination		Habitual behavior
Previous experience		Fear of failing the training	Phishing experience	
Outside of EVT				Contextual factors

Acquiring skills in identifying whether emails are legitimate or not from awareness campaigns was mentioned by some participants as a motivating factor. Through the campaigns, they increase their competence (self-concept of one's ability). They perceived the phishing campaign as beneficial in "training people to recognize what is phishing and prevent them from actually falling into one when it happens" (P22). Consequently, they held this expectation of maintaining *cyber safety*. As P9 shared, the campaign not only benefited them in terms of protecting their own data and e-mail accounts, it also "helped the university as an institution to be better protected."

A few participants believed that receiving training on security-related knowledge could benefit

their life and improve their computer literacy, contributing to *personal development* or long-term goals. P29 stressed that cybersecurity knowledge would become a fundamental skill for them to perform daily tasks with digital tools, and "it's not only about fear of being attacked, you need to understand what's inside these technology tools ... everything related to cybersecurity is very fundamental now and, in the future, would become even more fundamental, like reading."

Factors that discourage employees

Perceived low value discourages participants from taking online security courses, as indicated by P9, "not sure this kind of course will help me to be more precise in making judgments." On the one hand, the course was perceived as low value for some participants who had received security training before working in the current organization. On the other hand, some participants had concerns that the course might be in technical language, which can be difficult for people who are not tech-savvy to understand, "I'm going to attend it, but I'm not going to understand it" (P13). Furthermore, participants shared that the *lack of incentives* discouraged them from participating in security course. If the organization offered incentives, such as course credits (for doctoral researchers), compensation, and praise from the team leader, they would be more likely to participate in the security courses. As P24 asked, "what is my incentive to do an optional course here?"

Some participants expressed that even though they had intended to learn from the security course, the cover image and name of the course gave them the impression that it would *not be interesting*, resulting in them disengaging with the courses (P16). Participants thought that the course exercises were too simple; "the exercises were so obvious that you would truly have to make an effort to answer wrongly" (P2).

Participants frequently mentioned *time as a constraint* that discourages them from engaging with awareness campaigns. Participants found it difficult to allocate time to the awareness campaign due to their packed schedules. Time spent on the campaign was seen as an *opportunity cost*, as P23 stated, "instead of achieving something for your project, for example, a good experimental result, you spend time on the phishing campaigns, and you lose that opportunity." Multiple participants shared that a downside to engaging with awareness campaigns was heightened worry about potential threats - "*Negative inference*" (P30). An awareness campaign might lead them to experience more stress, compelling them to exercise increased caution in their daily lives (P5 and P25).

Participants expressed less interest in the campaign if the course content was not relevant to their area of expertise or interests. "*Other higher priorities*, such as course work and the experiment, would discourage me from participating in the awareness course; for me, the security courses were super boring" (P23). Participating in awareness campaigns requires people to switch from their tasks at

hand to phishing-related content. The switching *interrupted their workflow* (P25). Switching between tasks meant that it took additional hours for them to perform their duties (P27).

Participants' belief that they were less likely to experience phishing compared to others led to less involvement with the awareness campaign (*optimism bias*). As illustrated in P14's case, "I always had this thinking, it won't happen to me because this (phishing email) is so stupid." Participants also indicated that *overconfidence* in their knowledge of the topic made them less likely to engage with the awareness campaign (P28).

Previous negative experiences with security courses might evoke a *fear of failing the training*, which discouraged employees from participating. As P8 shared, "the fear or the worry that if I failed the course, it would be tracked. Because I experienced that in the previous job. If you didn't get a certain grade, then you would be forced to retake it and retake it." Additionally, participants shared that *procrastination* resulted in delaying or forgetting to take the courses (P32 and P33).

3.4.2 Report phishing emails

Factors that motivate employees

Participants had specific *expectations* when they reported phishing emails. Reporting was a practical way of notifying colleagues and alerting them of phishing attempts. Participants expected that the organization would improve its spam filters with their reported emails, which would benefit them in terms of receiving fewer spam and phishing emails in future. "The main benefit of reporting is that the IT team could create more filters for phishing emails if they have more data (from reporting), making us safer" (P27). They expected that the organization could contain the damage, retrieve stolen data from attackers and mitigate risks. *Worries and fears* related to the consequences of phishing attacks prompt participants to report. Specifically, participants worried that they would get into trouble, lose information, suffer from financial risks, and involvement in cyber crimes if they did not report promptly. Several participants emphasized reporting to avoid potential reputational damage and financial losses for their workplace (P13).

Participants indicated that reporting *protected their personal data*, financial assets, and other valuable possessions, including personal accounts. When suspicious of an email, they received support from the IT department in assessing the reported email. Beyond work-related protection, one participant felt safer in their personal life after reporting a phishing attempt to law enforcement, specifically an email accusing them of financial misconduct. Their concerns were alleviated once the email was confirmed as a phishing attempt. Participants also regarded reporting as a measure to *safeguard the workplace*. Firstly, reporting phishing attempts protected the organization's confidential data, documentation, work tools, internal network and servers from external access (P23). Secondly,

reporting was viewed as a way of raising awareness of phishing attempts in the organization. Not only the IT team needed to be notified of phishing attempts, but also their colleagues (P11 and P12). Thirdly, participants regarded reporting as a collaborative approach to countering phishing. The IT team assisted the employees in verifying the legitimacy of emails, and employees assisted the IT team in detecting the phishing attempts in real-time (P19).

Participants shared their *experiences receiving phishing emails*. Some received suspicious emails from professors, colleagues or family members asking for money or directing them to fraudulent websites. Others fell for phishing attempts while using online hotel booking platforms. P19 is a doctoral researcher in computer science who got phished a week before the focus group, "I lost two days of my life trying to correct just one click. During the backup, I lost a bunch of documents (erased a password for storing work documents), so there were other consequences after that." Even though the incident happened in their private life, it impacted their work. After the phishing incident, P19 wanted to warn others about phishing attacks and was motivated to report phishing attempts.

The ease of reporting phishing emails was mentioned as a reason why some participants reported phishing frequently. They referred to the one-click reporting button as straightforward, which made the reporting process simple and not time-consuming. They emphasize the one-click option for quick responses. The positive user experience of the reporting button facilitated participants to report, as exemplified by P31: "It's easy so it doesn't take even two seconds. If you suspect, click, click, and then you're done."

Participants regard the "congratulations" email that they received from the IT team when they reported a (simulated) phish as a kind of "*recognition*" and extrinsic reward for their reporting (P9). While P21 used to ignore phishing emails, one colleague told them it's better to report (*peer influence*). After that, P21 started to report suspicious emails. The *sense of being part of the community* prompts participants to report, as exemplified by the following conversation:

> P32: "We need to participate. We're all active users and it's not just IT who has to deal with it."
>
> P34: "We are actors within the community. So, we are together."

Participants described that they experienced feelings of *enjoyment, satisfaction*, and *pride* when reporting phishing attempts, likening the process to a game, feeling proud of their vigilance, and deriving a sense of satisfaction from reporting. As P28, P11, and P8 indicated:

> "When you click to report phishing attempts, then you receive 'congratulations'. I'm happy and it's like a game." (P28)
>
> "I can relate to the sense of satisfaction. Once you've reported it, you feel like you played

> your role. You did a good job." (P11)
>
> "I don't want to break my streak of always reporting the phishing attacks ... I'm quite proud of that." (P8)

Several participants mentioned a number of *core values* (guiding principles that shape people's attitudes, actions, and decisions) that drive them to report phishing attempts, including "help others" and "vulnerable" groups (P2 and P15), "duty" (P11), "being a good citizen" (P28 and P33), and "contributing to the fight against phishing" (P33). Additionally, a few participants considered reporting as an approach to take control and make a difference (P6). In P16's case, "I had the initiative to defend against the phishing attack. And knowing that I can stop spreading this attack for other people and for my future self really helps me, like *empowering*."

Factors that discourage employees

Multiple participants felt discouraged from reporting suspicious emails because they received *no feedback* on the outcome of their actions. They expected to receive more information about the outcomes of their reporting (P12). As P31 emphasized, "we don't know what the effectiveness of reporting phishing emails is. We don't know the numbers, so it would be really good to have a kind of feedback status. What has been done last year? What was the success rate?" Further, even for participants who reported diligently, they sometimes felt discouraged from reporting due to not knowing whether their colleagues were reporting or not (*lack of communication*).

> "I report phishing emails regularly and religiously, but I'm thinking is everyone else doing the same as me, putting in the same effort as I am on reporting? It takes maybe 30 seconds of your time, but I'm still very careful about it." (P25)

The perceived *low utility value* discouraged participants from reporting phishing emails. Firstly, the belief that the "phishing" email is merely a test from the IT department reduces the perceived need to report it, as stressed by P27, "for me, every phishing email that I received was a simulated one. So, I didn't see the point of reporting that because I knew that it was from IT." Secondly, if the participants believed most people would be able to recognize the email as a phish and posited a low threat to others, they chose not to report (P16). Thirdly, worries of additional burden due to reporting discouraged participants from following the reporting procedure. These assumed negative outcomes included "bog me down with questions" (P13), getting "more emails" (P17), and "fear of annoying IT staff" (P28). Lastly, the belief that reporting doesn't lead to effective outcomes, such as prevention or resolution of the attack, discouraged participants from reporting. As exemplified by P19, "the lack of results discourages me. It seems like we try to do something nice and nobody really cares."

Participants highlighted several issues related to ease of use, functionality, and efficiency in the reporting process as discouraging factors (*usability issues*). Some participants found the reporting procedures ambiguous. For instance, P8 only learned about the "report-a-phish" email address from a colleague after observing the absence of a reporting button following an update of the email client. P26 wondered about the preferred method of reporting, stating, "I forwarded it to report-a-phish, and they said, 'Oh no, can you please send it as an attachment instead of forwarding it'." For participants who frequently reported suspicious emails on their laptops mentioned that they often delete or disregard such emails when viewed on their smartphones. P9 shared, "I wanted to report it and I had trouble doing that with my phone. So I always try to be extremely careful, almost like you have something burning in your hand." Despite their caution, they still accidentally clicked on the email when trying to report it, leading them to ignore phishing emails on their phones. Moreover, Linux and Mac OS users felt the reporting process demanded too much effort. It's easier to just delete the suspicious email than to forward the email as an attachment to the IT department. As emphasized by P24, "if it's anything more than a one-button click would be a little bit more discouraging."

Participants expressed they would not report when they were concerned that the suspicious emails "*disclose their private information*" or cause false impressions about their personal life (P4, P28). Additionally, *worries about being judged* by the IT team were shared as a discouraging factor by participants. As the conversation between P33 and P34 revealed:

> P34: "I have this feeling that IT guys, they're always like a bit, 'they don't know they're doing really.' And I feel I'm so stupid. If I report Netflix or something as phishing, then they would think 'stupid woman'."
>
> P33: "They could judge us."
>
> P34: "So this feeling unnerved me and discouraged me from reporting. Because they give you this feeling sometimes. I experience it, I call the help desk and get this 'again'."

Participants shared that they frequently postponed or forgot to report because they reverted to their *old habits* of simply deleting emails. They mentioned that the reporting process is unique to their current workplace, contrasting it with their usual habit of deleting or marking suspicious emails as spam. As P11 stated, "in my personal life, when I encounter a suspicious email, I just delete or mark it as spam. However, this report-a-phish button is quite specific and new." Participants noted that if they *lacked confidence* in identifying whether an email is phishing, they would typically ignore it. Furthermore, some participants cited "laziness" as a reason for not reporting.

Contextual factors, such as task overload, stress, and time pressure, could deter participants from reporting phishing emails. When focused on one's tasks and in the status of flow, they perceived incoming emails as a distraction, resulting in less intention to report (P27).

3.4.3 Improvements proposed by participants

Participants proposed various ideas to make phishing interventions more engaging during the brainstorming sessions. We categorized them into the following themes:

Gamification elements: Participants suggested adding achievement, competition, virtual reputation, and fun elements to the reporting process. There should be rewards or acknowledgments for the department that actively participates in awareness campaigns and reports the most phishing emails. Participants recommended providing incentives for participation in phishing campaigns, such as gifts, praise, and course credits. Participants suggested that role-playing and leaderboards would engage employees with the security training.

New employees & Mandatory training: During the onboarding week for new employees, the university should provide a mandatory training session to equip them with knowledge about phishing and the reporting procedure. The IT team should walk in the shoes of new employees and find out the potential attack points within their work activities. Participants also suggested making a security course mandatory for frequent clickers of phishing tests and for departments that receive a high number of phishing attacks.

User experience: Participants suggested to improve the user experience of phishing interventions. Real-time verification of reported emails and shorter, more relevant and interactive trainings would attract employees. Course content should be personalized according to different levels of phishing knowledge. Participants suggested using pop-up quizzes instead of online videos to raise phishing awareness because the latter took too much time.

Communication: Participants suggested that the IT team provide regular updates or host information sessions with employees. The positive impacts that phishing interventions have on the university should be communicated quarterly or annually. Seminars drawing from diverse expertise areas like IT, HR, and research were recommended to bolster organizational defense and collaborations between departments.

Feedback: The IT team should gather feedback on phishing interventions from employees, provide statistics on phishing interventions, and be transparent about the state of the art and the efficacy of current solutions. Participants also suggested the IT team provide individual feedback on what happens after an employee reports phishing.

Present real incidents: Participants suggested the IT team present real phishing attacks and their consequences as examples to raise awareness. Providing concrete examples of how data breaches happened through phishing would raise employees' phishing awareness.

Authentication of internal emails: Participants suggested implementing digital signatures to authenticate internal communication, which would enable fast detection of phishing emails that

masquerade as internal communication. Additionally, participants suggested recruiting *more IT employees* to host training sessions regularly, noting that the IT team seems occupied with an overload of tasks. Lastly, one group proposed a *punishment* approach, that is, increasing the number of simulated phishing emails for employees who repeatedly clicked simulated phishing emails.

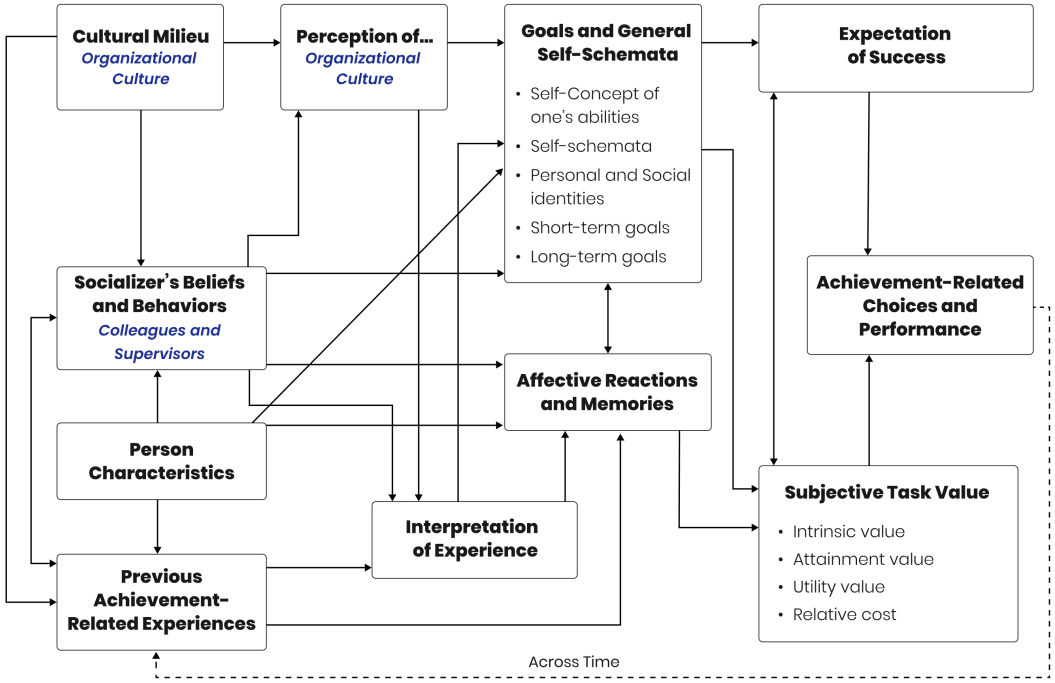

Figure 3.1: The expectancy-value model (adapted from Eccles and Wigfield [111]).

3.5 Discussion

3.5.1 Applying EVT to the context of organizational cybersecurity behaviors

In this study, we investigate how Expectancy-Value Theory (EVT) can illuminate the factors influencing employees' engagement with phishing interventions. Building on our findings and considering that EVT was created to interpret achievement-related choice and performance in educational settings, we propose incorporating an organizational dimension into EVT model (refer to Figure 3.1, our adaptations to EVT are in blue italics). Hence, we suggest integrating the organizational dimension in the form of "organizational culture" [404] into a "cultural milieu" construct, which can be described as a system of social roles, each with its associated responsibilities and obligations [400]. Perception of the organizational dimension can be interpreted through the lens of the "psychological contract", which refers to an unwritten set of expectations and beliefs about the obligations that exist between an employee and their employer [153], also including employees' beliefs about their responsibilities in organizational security [163]. During group discussions, employees consistently highlighted that, through their security behaviors, they aim to collaborate with the IT department in fighting against

phishing attacks and safeguarding the organization. Despite the absence of explicit organizational policies dictating such obligations, this inclination can be attributed to the implicit norms acquired through the organization's unspoken rules and in general - organizational culture as a proxy for information security culture in the organization [341]. Our results suggest that the perception of the organizational culture, communicated through socializers' beliefs and behaviors, can contribute to a constructive "us vs. them" (organization vs. attackers) mentality, where employees have a self-concept of a contributor to organizational security.

In accord with past studies [390, 89], we observe that "peer influence" and "knowledge sharing" among colleagues influence employees' intention to report phishing emails and participate in online security courses. Pursuing this line of thought, we can extend the EVT model's "socializer" construct to include "colleagues and supervisors." These people convey their knowledge of the organization's unwritten norms to other employees, aiding in shaping security protective identities. Furthermore, we propose that employees' security consciousness stems from their social identity in EVT. Being a "responsible" employee dedicated to the organization, in harmony with other foundational roles, makes up one's social identity. This is connected with the "extra-role security behaviors" phenomenon [239, 131], in which some employees are self-motivated to take additional responsibilities to secure the organization, even if these responsibilities go beyond their contractual role. In our study, we found strong evidence that this type of motivation is one of the core drivers of reporting phishing emails.

In summary, our empirical findings demonstrate that the EVT framework can be specialized for use in organizational security settings. We specify certain concepts of EVT in the organizational setting, proposing to focus on the organizational culture, colleagues and supervisors, and "previous experience" on the left side of the framework (see Figure 3.1). Our findings also support the original EVT framework with findings that subjective task values, expectations, goals, and general self-schemata influence employees' engagement with phishing interventions. The discovery paves the foundation for future studies to apply EVT in studying organizational security behaviors.

3.5.2 Subjective task value of phishing interventions

The majority of educational interventions based on EVT focus on altering individuals' "Subjective task value" [112]. Subjective task value is the core construct within EVT, in which the value of engaging with an activity can be considered as the ratio between perceived benefits and associated costs [111]. People tend to opt for activities that have a higher benefit-to-cost ratio. Our findings showed that many of the discouraging factors of phishing awareness campaigns are associated with different types of costs, such as psychological cost, time cost, and opportunity cost. The findings align with previous studies on imposing security measures within organizations, which found that

employees perceived the security measures as extra burdens that encumber their work [216, 50, 308]. Previous literature proposes remedies such as reducing the friction associated with security measures and automating security protocols [133, 417, 81].

In the EVT framework, another promising avenue for exploration emerges: the potential for security managers to tip the scale in favor of security measures by reducing their associated costs. This shift could engage employees more with security measures. This idea aligns with studies showing positive outcomes from security trainings in short video format, with participants regarding the training "informative"/"useful" [413] and expressing interest in extended sessions [382]. The increased benefit-to-cost ratio in such cases can be attributed, in part, to the brevity and density of the training content. Our study echoes employee preferences for succinct training, as exemplified by "don't give me a half hour course for two minutes' value" (P13) in the group discussion. Similarly, participants in different groups proposed providing employees with shorter but more frequent security trainings.

Our study identifies a cluster of motivators associated with the intrinsic values of reporting. These motivators, deeply embedded in employees' psychological needs and desires, include satisfaction, empowerment, and core values (citizenship and altruism). Our findings are congruent with previous studies, which suggest that autonomy, personal values and principles influence users' security behavior [218, 257]. These elements, often sidestepped in security behavior research, weave a complex network of factors influencing phishing reporting intentions. Considering that security messages that appeal to individuals' desires are more likely to elicit secure responses than those based on fear [257, 344], organizations should establish reporting procedures that resonate with employees' psychological needs. Integrating "fun" [338] and "experiential learning" [72] elements into training programs can enhance their intrinsic value, thereby engaging employees with phishing awareness campaigns. Furthermore, Eccles and Wigfield suggested developing attainment value-based interventions [112]. These interventions could take the approach of informing employees about the connection between anti-phishing practices and their personal values.

3.5.3 Previous experience, expectation of success, and personal development

Our study reveals that even motivated employees can become disheartened if they lack clear feedback and perceive their actions as ineffective. Several discouraging factors for phishing interventions can be categorized under "previous achievement-related experience." According to EVT, the "interpretation of experience" can influence "expectation of success" by altering goals and subjective task value. This attenuation is often due to negative experiences from prior engagement with the task. Employees are more inclined to adhere to security protocols if they deem the processes effective in mitigating

phishing attacks [338]. Various employees in our study identified the lack of feedback and clarity about subsequent steps after reporting an email as discouraging factors, often provoking uncertainty and negative emotions. Such a phenomenon was also observed in employees' attitudes towards phishing awareness campaigns where previous unfavorable experiences shaped their perceptions. Over the last 20 years, research has persistently emphasized the critical role of feedback in fostering secure behavior within organizations [1, 321, 26]. Our study further explores the mechanisms through which an absence of feedback can alter motivation, even for motivated employees.

Intriguingly, we noted that prior experiences with being phished emerged as a strong motivator for some employees to report phishing, propelling their goal to prevent others from undergoing similar negative consequences. We hypothesize that the negative experience altered the subjective value they placed on reporting, which necessitates further study of this transformation from victim to defender in the context of phishing. Recognizing this transformative process can inform the development of support structures within the workplace. Employees who encounter cybersecurity incidents often experience guilt and shame. Workplaces should provide support, instead of blaming, to contain damage caused by the incidents and empower their employees [305, 102].

Employees demonstrated interest in acquiring security-related knowledge, linking it with their personal and professional growth. This interest suggests a pathway for organizations to refashion their security training to better align with employees' long-term goals. Given that all employees manage valuable accounts and passwords, and are often influenced by media reports or personal experiences of cybersecurity incidents, the imperative to adeptly navigate digital protection is clear. Similarly, Reeves et al. suggest shifting from a compliance-driven to a user-driven approach in security training to enhance the efficacy of training programs [301]. Incorporating employees' personal learning needs into organizational training paradigms could motivate employees to engage with security trainings.

3.5.4 Practical implications

We found that many of the discouraging factors related to the phishing awareness campaign are associated with its perceived value. Several usability-related factors discourage employees from reporting phishing emails. Fear, worries, and concerns about phishing interventions discourage employees from engaging (see Table 3.1). Leveraging insights from both the employee-generated suggestions and the EVT framework, we have proposed several improvements:

For phishing awareness campaigns: Clear communication of the campaign guidelines, expectations, goals, and consequences can alleviate the discouraging factor of "fear of failing training." Specific time slots should be allocated for employees to participate in the training sessions, addressing the discouraging factors of time constraints and interruption to their workflow [402]. This might

not be possible in the case of knowledge workers who autonomously allocate time and tasks, for whom training will inevitably cut into their "productive" time. Making the training content relevant to individual job roles would enhance its relevance and applicability to daily tasks. Regular updates on evolving phishing attacks should be provided to increase awareness among employees. Gamification elements in the training program might enhance engagement [338].

For reporting phishing emails: Organizations should clearly communicate how reported incidents are managed by the IT team [127]. Timely feedback mechanisms should be established, reinforcing employees' sense of contributing to security. Regular updates (e.g., intranet, messages, displays) are beneficial for keeping employees informed about security efforts and emerging threats. Providing statistics on reporting and organizational benefits can underscore the personal value of reporting incidents. The reporting process should be frictionless to alleviate usability concerns. Ongoing awareness initiatives can foster engagement [87]. Training new employees is crucial to acquaint them with countering phishing practices and maintain a consistent level of awareness throughout the organization.

3.6 Limitations and future work

Despite their advantages, focus groups have a few limitations which we were careful to mitigate through purposeful moderation. The discussion might veer into narratives outside the scope of research. Also, dominant speakers might hijack the discussion while some participants might remain silent and not willing to confront others. This requires researchers' facilitation to steer back to the planned agenda and engage participants with contributing. Furthermore, much of the collected data is expressed informally, necessitating careful interpretation by researchers. Thus, we involved multiple researchers in the data analysis process. Participants' viewpoints might be influenced by the others' arguments during group interaction. Thus, we recorded individual opinions prior to the group discussion on reporting to obtain individual viewpoints.

Although we utilized diverse strategies to recruit employees from the organization, we might have attracted people who are particularly interested in the topic. We hypothesized that an important power imbalance exists between the IT security team and other staff regarding the topic of the study. We did not have IT security officers as participants. We acknowledge that focus groups were composed of participants with multiple roles, potentially creating a perceived power imbalance that inhibited participation. The investigated university has no strict rules regarding phishing awareness campaigns, reporting, and the use of personal devices for work. Thus, while our findings offer valuable insights, critical interpretation is warranted when extrapolating results to different organizational contexts. Future studies should use quantitative methodologies to test the hypotheses drawn from our results.

We found that contextual ("situated") factors, such as task overload, time pressure and stress, influence employees' response to phishing emails (in line with [102]). Contextual factors are not represented in the original EVT framework, although the authors later highlighted that the processes underlying the EVT model are influenced by the immediate situation in which a decision occurs [111]. Recent early-stage work suggests using knowledge about momentary user states to better tailor security interventions [18], for example proposing security interventions or training in opportune moments. We suggest future studies investigate how to integrate contextual factors into EVT when applying it to study information security behaviors.

3.7 Conclusion

Employees are the last line of organizational defense against phishing attacks [414]. It is important to train and engage employees and encourage reporting of phishing attacks to enable organizations to respond promptly. This engagement can be achieved by enhancing the perceived value of the task, reducing its relative costs, and making *phishing awareness campaigns* more user-centric and relevant to employees.

We find that Expectancy-Value Theory is a valuable theoretical framework for studying user security behavior in an organizational context. EVT helps explain how organizational culture, social roles, and the influence of colleagues and supervisors foster proactive responses to phishing attacks.

Our study reveals a spectrum of factors that influence employees' intentions to *report phishing emails*. Some factors not previously discussed in phishing studies include those associated with social roles (safeguarding the workplace, sense of belonging, and collaboration with IT) and intrinsic factors (satisfaction, enjoyment, and empowerment). Among the factors discouraging employees, the absence of feedback and perceived low utility value are particularly detrimental. This lack not only affects the perceived value of reporting but also undermines employees' confidence in the effectiveness of countermeasures. Given that users devote considerable time and effort in addition to their role to engage in security tasks, it seems justifiable to provide them with more feedback about how their actions fortify the organization's defenses against phishing attacks. A month after our focus group session, we received an email from P18—a highly motivated employee who indicated that they always report suspicious emails. They allowed us to cite:

> *I have now finally stopped reporting phishing emails. Yesterday, I received two that were exactly like the ones I've been getting dozens of times over the past years. It feels a bit like an insult to be asked to report phishing emails when this information is so evidently not utilized. I expressed this sentiment in my final report, but of course, it was ignored.*

We see this loss of engagement with phishing reporting as an understandable but regrettable behavioral response. Envisioning such sentiments and the resulting behavior at scale, with possibly large numbers of employees ending up disappointed and disengaging from phishing interventions, we can only speculate regarding the negative effects on the organizational security of an organization. We hope that this paper can help avoid such frustrating experiences for employees in the future by providing a better understanding of the motivating and discouraging factors for phishing interventions through the lens of EVT.

3.8 Acknowledgments

Author 1 acknowledges the financial support of the Institute for Advanced Studies at the University of Luxembourg through a Young Academic Grant (2021). The study was supported by the User Lab of the University of Luxembourg. Thank you to our reviewers for their constructive feedback. We thank all our participants. A shout-out to Eric J. Francois for his suggestion of role-playing, which inspired the development of a subsequent "role-playing as hackers" training [72].

Chapter 4

The Effects of Group Discussion and Role-playing Anti-Phishing Training: Evidence from a Mixed-design Experiment

Published as: Chen, X., Sacré, M., Lenzini, G., Greiff, S ., D istler, V ., & S ergeeva, A. (2024, May). The Effects o f G roup D iscussion a nd R ole-playing T raining o n Self-efficacy, Support-seeking, and Reporting Phishing Emails: Evidence from a Mixed-design Experiment. *In Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1-21).

Abstract Organizations rely on phishing interventions to enhance employees' vigilance and safe responses to phishing emails that bypass technical solutions. While various resources are available to counteract phishing, studies emphasize the need for interactive and practical training approaches. To investigate the effectiveness of such an approach, we developed and delivered two anti-phishing trainings, group discussion and role-playing, at a European university. We conducted a preregistered[1] experiment (N = 105), incorporating repeated measures at three time points, a control group, and three in-situ phishing tests. Both trainings enhanced employees' *anti-phishing self-efficacy* and *support-seeking intention* in within-group analyses. Only the role-playing training significantly improved support-seeking intention when compared to the control group. Participants in both trainings reported more phishing tests and demonstrated heightened vigilance to phishing attacks compared to the control group. We discuss practical implications for evaluating and improving phishing interventions and promoting safe responses to phishing threats within organizations.

4.1 Introduction

Phishing was the most reported cybercrime between 2019 and 2022 in the US [121]. Globally, 4.7 million attacks were recorded in 2022 [24]. Phishing attacks exploit human factors by using social engineering techniques to deceive individuals into divulging confidential i nformation o r installing malware on their devices [124, 167]. Phishing is typically used as the initial entry point for more advanced attacks, including ransomware attacks, intellectual property theft, and business scams [144], which can result in billions of dollars annual losses for organizations [120]. As communication channels like email, instant messenger, and team collaboration tools become popular, attackers are exploiting new vulnerabilities to target technology users [167, 9, 194].

[1]Preregistration link: `https://aspredicted.org/qi8ka.pdf`

Organizations employ multiple technical measures to reduce the number of phishing attacks, but these measures may fail [60]. In such cases, employees' vigilance against phishing can serve as the last line of defense for organizations [245, 414]. Employees' awareness and proactive responses to phishing attacks improve organizational information security [181, 32]. Employees who are aware of their organization's information security policies and procedures demonstrate greater competence in managing cybersecurity tasks than those who are not [237]. In addition, employees' self-efficacy in information security positively influences their intentions to comply with security policies [376]. Consequently, organizations implement a range of phishing interventions to heighten employees' awareness and develop their competence in countering phishing [133].

In addition to on-site/online education, organizations send simulated phishing emails to monitor employees' responses and resistance to phishing attacks (referred to as *simulating phishing tests*) [198, 106]. While simulated phishing tests have been commonly adopted by organizations to raise employees' phishing awareness, they alone may not adequately train users to respond safely [383]. Accordingly, the *embedded phishing campaign* was developed, in which employees who respond unsafely to the simulated phishing email are directed to a webpage containing educational resources. Studies on the effectiveness of embedded phishing campaigns in training employees for safe responses have yielded mixed results [222, 409, 229, 94]. Given the evolving nature of phishing attacks and the critical role of employee vigilance [148], studies suggest developing more practical and interactive anti-phishing training for employees [197, 402, 362].

Role-playing training helps participants gain experience with various situations, equipping them with the skills and knowledge to anticipate, adapt to, and recover from undesirable situations [207, 406]. A recent role-playing anti-phishing training, "What.Hack", was found to be effective in improving users' performance of identifying phishing emails within a controlled laboratory setting [397]. This effectiveness was evident when compared to the results of two alternative training programs [397], but engaging with What.Hack does not require social interactions between users and the contextual factors embedded within the game narrative might potentially lack relevance in the work context [397]. Given that previous research has shown that contextual factors affect employees' susceptibility to phishing attacks [130] and workplace social interactions can help employees protect themselves against such attacks [102], it is imperative that we incorporate these aspects when designing anti-phishing training programs.

To further explore practical and interactive training approaches tailored for the work context, we have specifically designed two anti-phishing trainings: group discussion and role-playing. Both trainings are designed with the goal of enhancing employees' *anti-phishing self-efficacy* and *support-seeking intention*. They aim to train employees to respond safely to phishing emails. During

the group discussions, participants are encouraged to *share their experiences and anti-phishing practices* with their colleagues [319]. In the role-playing training, we guide participants to *think like a hacker* and design phishing emails to infiltrate the organization [115]. By comparing these two training approaches, we intend to address the following research questions (RQ):

- **RQ1**: What is the effect of role-playing training on employees' anti-phishing self-efficacy compared to group discussion and the control group?
- **RQ2**: What is the effect of role-playing training on employees' support-seeking intention when receiving phishing emails compared to group discussion and the control group?
- **RQ3**: How do group discussion and role-playing training influence employees' response to phishing attacks?

To address these questions, we employed a *mixed-design experiment* [65], assessing training effects both within and between subjects. We incorporated repeated measures at three time points, included a control group, and conducted three simulated phishing tests in our study. This method is novel in that it allows us to assess the effectiveness of anti-phishing trainings with respect to participants' self-efficacy, support-seeking intention, and responses to phishing emails. This paper makes four primary contributions:

- We contribute to the understanding that group discussion and role-playing are effective anti-phishing training approaches that enhance employees' perceived self-efficacy, support-seeking intention, and vigilance towards phishing attacks.
- Our study highlights the significance of discussing phishing incidents and anti-phishing practices in the workplace, demonstrating its potential to promote safe responses to phishing attacks.
- We introduce a new and useful measurement for evaluating anti-phishing trainings: support-seeking intention when receiving suspicious emails.
- Our study is one of the first to employ a mixed-design experiment to assess the effects of anti-phishing trainings in the field, measuring both self-reported and behavioral changes. We demonstrate that it is possible to obtain informed consent from research participants for simulated phishing tests and still receive useful insights from the results.

Reflecting upon the results of our study, we advocate for role-playing as an enjoyable and effective approach to engage employees with anti-phishing training. Our study is among the first to demonstrate the effectiveness of two trainings aimed at increasing employees' propensity to report phishing emails through in-situ phishing tests over a time period of three weeks.

4.2 Related Work

In Subsection 4.2.1, we review previous studies on anti-phishing education and training[2]. Then, we review the methods that have been employed to evaluate educational and training interventions in Section 4.2.2. Lastly, we examine the role of self-efficacy and social interaction in countering phishing attacks in Section 4.2.3.

4.2.1 Anti-phishing education and training

Simulated phishing campaigns and on-site/online education are among the popular interventions adopted by organizations to bolster their phishing resilience [197, 409, 224, 145]. For organizations with a large number of employees, simulated phishing campaigns seem to be a convenient solution to raise phishing awareness [88]. Longitudinal observations from an Australian educational institute revealed that employees exhibited safer responses after six cycles of embedded phishing campaigns, as opposed to the period when only simulated phishing tests were administered [409]. However, another large-scale and long-term study in Switzerland found contradicting results and argued that embedded training during simulated phishing tests does not make employees more resilient to phishing [229]. Besides these mixed results, deploying phishing campaigns is expensive and requires dedicated human resources in organizations [50, 308]. Instead of off-the-shelf phishing campaigns, some organizations have designed their own training programs for employees; for example, 409 employees of a German organization improved their skills in distinguishing phishing email screenshots significantly after attending on-site tutorials [302]. When comparing instructor-, computer-, and text-based anti-phishing training based on the same content, Stockhardt et al. found that instructor-based training was more effective in transferring knowledge than the other two formats and had the highest scores in user satisfaction and confidence [351]. In an experiment conducted within an organization, participants who underwent an adversarial training, adopting the mindset of a cybercriminal, revealed a nearly threefold decrease in susceptibility to phishing attacks compared to those who received a video training [413].

Role-playing has been a central approach employed by several anti-phishing digital and card games to engage users in learning [412]. Digital games, including Anti-phishing Phil [335], Bird's Life [394], and What.Hack [397], have been created to teach users how to identify phishing elements. These games have primarily been evaluated with university students [397, 394]; therefore, the effectiveness of these trainings for organizational employees require further empirical investigation [165, 36]. On

[2]Anti-phishing education and training: Following the phishing intervention taxonomy [133], we examine studies related to *phishing education* (which involves developing knowledge and understanding of phishing) and *training* (aimed at cultivating skills that users can apply when encountering phishing). Refer to the Supplementary Material for an overview of the role-playing and other types of anti-phishing trainings reviewed in our study.

the other hand, card games, which often incorporate red team (attackers) and blue team (defenders) designs, have also been introduced to assess the vulnerabilities posed by social engineering attacks [36, 165] and raise awareness of excessive online information disclosure, enhancing phishing awareness [118]. Card games appear to be more accessible than digital games, but they tend to exhibit a level of complexity that can challenge users' ability to quickly grasp the game rules [30], necessitating additional learning efforts and expert guidance [118].

4.2.2 Evaluating educational and training interventions

The majority of our reviewed studies conducted either user evaluation [165, 394, 118], which focuses on studying the usability of the intervention, or between-subjects experiment [397, 190, 413], which compares the training effects with alternative interventions; only one study chose the approach of measuring the training effects over time in the field [302]. Post-training questionnaires were frequently used to assess participants' learning experience [397], self-efficacy [190], perceived effectiveness [165], and feedback [335]. A few studies compared pre- and post-training questionnaires to evaluate training efficacy. The measurements include phishing knowledge [394], confidence level [75], behavioral tendencies, self-reported computer skills and perceived risks [354]. Further, participants' demographic information was commonly collected to examine their relationship with phishing intervention outcomes [118, 61, 229].

Log data of simulated phishing tests and participants' performance in distinguishing phishing emails/websites from legitimate ones have been used as indicators of training efficacy in many previous intervention studies [66]. The reporting rate and click-through rate of phishing campaigns have been used to evaluate participants' responses to phishing emails [106, 409] and to evaluate the effectiveness of different education approaches [61, 229, 392]. Through online surveys, Reinheimer et al. measured employees' performance of distinguishing phishing email screenshots from legitimate ones [302]. In the laboratory, participants were instructed to classify emails/websites as phishing or legitimate ones before and after the intervention to assess their effectiveness [225, 335, 397]. In a literature review [66], Chaudhary et al. compared existing evaluation methods and considered simulated phishing tests provide more realistic view of participant's responses than question-based tests, assuming they comply with data protection laws and are conducted in ethical ways.

Furthermore, observation of time spent on the task [30], group discussions [36], and participants' designs [118] have been analyzed to evaluate the interventions. In an in-situ deception study, Distler combined observation and interview data to study employees' responses in the context of their typical work tasks to spear phishing attacks, including social interactions, and their reporting behavior as well as rationalizations [102]. In-situ studies evaluating educational and training interventions deliver

important insights as they maximize ecological validity [302, 66], enable researchers to observe how context influences reactions to a social engineering attack, and allow for the capture of natural reactions at the critical moment when an employee is exposed to a social engineering attack. However, many ethical challenges are associated with conducting ecologically valid phishing studies; for a discussion refer to [306].

4.2.3 Self-efficacy and social interaction in anti-phishing

Self-efficacy is the most frequently studied construct from Protection Motivation Theory when applied to users' information security behaviors [158]. Self-efficacy in information security is defined as *a belief in one's capability to protect information and information systems from unauthorized disclosure, modification, loss, destruction, and lack of availability* [307]. Studies show self-efficacy positively influences employee's information security compliance intention [257, 171, 376]. Specifically in anti-phishing studies, self-efficacy positively influenced an individual's likelihood of reporting phishing emails [227, 246] and mobile users' motivation to avoid phishing [380]. Employees with stronger self-efficacy were more inclined to share their negative experiences with colleagues, alerting them about phishing attacks in a financial company [77]. In this study, we define *anti-phishing self-efficacy* as a belief in one's ability to recognize suspicious emails and keep up to date with phishing techniques [403, 272]. Several self-efficacy scales have been developed to measure users' self-efficacy in domains such as information security [307] and smart home security [46]; however, there is currently no comparable scale available for measuring anti-phishing self-efficacy.

Promoting social interaction within organizations can benefit the organization in defending itself against phishing threats. Firstly, workplace social interactions can support employees in assessing and responding to suspicious emails [102]. By motivating employees to report suspicious emails, organizations can detect phishing attempts within minutes after a new attack is launched [229]. Secondly, Das et al. found that "social cues", where individuals engaged with or observed others' actions, were the predominant category of triggers that led to recent security and privacy behaviors in an online survey (N = 852) [91]. Thirdly, stories shared by peers, second only to expert advice [392], led to lower click rates in phishing tests compared to those who received other forms of training materials in two experimental studies [248]. Thus, research suggests that promoting social interaction and experience-sharing at the workplace holds promise as an effective approach for defending against phishing attacks.

4.3 Methods

4.3.1 Group discussion and role-playing training design

For the purpose of the study, we developed two training programs: a group discussion and a role-playing training. The underlying mechanism behind using group discussion to train employees stems from a study showing that small group discussions can bring about powerful and lasting changes in information security awareness and behavior [8]. The role-playing training design was inspired by two previous studies showing that role-playing is an engaging approach to train anti-phishing skills [397] and that by assuming the role of hackers, students improve their awareness of spear phishing risks [118]. We chose a face-to-face training approach, instead of a digital format, to foster more social interaction between participants [157].

To facilitate the comparison between group discussion and role-playing training, we used the same set of materials to design them: recent real phishing emails targeting the organization, a phishing definition [124], content cues [331, 123], attack channels, and phishing techniques [9]. In the process of designing both training programs, we consulted two experts from the organization's Information Security Office with expertise in phishing and three professors working on cybersecurity and Human-computer Interaction. Prior to data collection, we conducted a pilot study of the role-playing training with 7 employees to gather feedback and refine the training procedure. The group discussion and role-playing training shared the same structure and length; they both started with a brief introduction of the study and training schedule, a tutorial on phishing fundamentals, group discussion or group work, and a conclusion, as outlined in Table 4.1.

Table 4.1: Group discussion and role-playing training procedure.

	Group discussion	**Role-playing**
5 min	Introduction	
20 min	Phishing fundamentals: definition, content cues, channels, and techniques	
50 min	Analyze real phishing emails with a template (10 min). Discuss phishing emails, perceived vulnerability, and coping strategies (40 min).	Group work on real phishing emails and design one phishing email (40 min). Phish each other and identify the phishing email (10 min).
10 min	Conclusion phase	

In the group discussion condition, we first asked each participant to scrutinize two real phishing emails with a template with questions on suspicious elements of the emails and the difficulty of identifying them as phishing. Then, we moved to discuss the following questions in groups of 4 to 6 people facilitated by a researcher:

- *What surprised you most about these real phishing emails?*
- *Have you received phishing emails on your work accounts?*

- *How do you respond to these suspicious emails related to work?*
- *Which contextual factors related to your job position might be exploited by attackers?*
- *What would happen if you were to click on a suspicious email or download malware to your work laptop?*

In the role-playing condition, we asked participants to play the role of hackers aiming to infiltrate the organization. We randomly divided participants into two groups, each comprising 2-4 individuals. Each group was equipped with a computer, an email account of a fictitious persona, and two legitimate work-related emails in their draft box. The group work started with a discussion on suspicious elements and attack techniques of real phishing emails and, subsequently, creating one phishing email together to phish the other group. 40 minutes later, the participants sent the created phishing email and two legitimate emails to the other group. After both groups sent their "phishing" email and legitimate emails, they were asked to identify the phishing email created by the other group.

In the conclusion phase, participants from both conditions shared strategies and practices they intend to use to protect their workplace from future phishing attacks, as well as the lessons they learned during the training. Additionally, we provided participants with additional tips for identifying phishing emails [273] and phishing awareness resources available at the organization.

4.3.2 Participants

We employed multiple methods to recruit participants for our study. We sent study invitations via email to all employees across four university faculties (Humanities, Engineering, Computer Science, and Medicine), posted recruitment materials on campus in the form of printed posters and digital displays, distributed flyers in two employee cafeterias, and conducted door-to-door recruitment in two office buildings (we include the recruitment poster in the Supplementary Material). 118 employees registered interest to participate in the study by completing an online questionnaire indicating their faculty, email address, and availability. We did not exclude any participants, and any current employee who had a work email account was allowed to participate in our study. We invited all employees who expressed interest to participate in our study.

Among the 105 employees who participated in our study, 60 reported being female, 40 male, one non-binary, and four chose not to disclose their gender. 60% (N = 63) of participants were aged between 25 and 34, 21% (N = 22) were between 35 and 44, and 13% (N = 14) were between 45 and 54. In terms of their professional background, 47 were from the Humanities faculty, 22 from Engineering, 9 from central administration, 9 from Medicine, 8 from Computer Science, and 10 from other departments. 39 participants worked as doctoral researchers, 15 as postdoctoral researchers or research scientists,

and the remaining participants held roles as research facilitators, Research & Development specialists, administrators, and professors. Their work experience at the current organization varied between 1 and 187 months (mean = 42.6, SD = 46.6).

4.3.3 Study procedure

Conditions. Participants were randomly assigned to one of three conditions to compare the effects of both trainings to a control condition[3]. The *control group* (N = 35) received no intervention from us and was conducted remotely. The *group discussion* (N = 35) and *role-playing* (N = 35) training sessions took place in the same user lab. We invited participants to attend a "Phishing Resilience Workshop" in groups, but they were unaware that there were two different training programs.

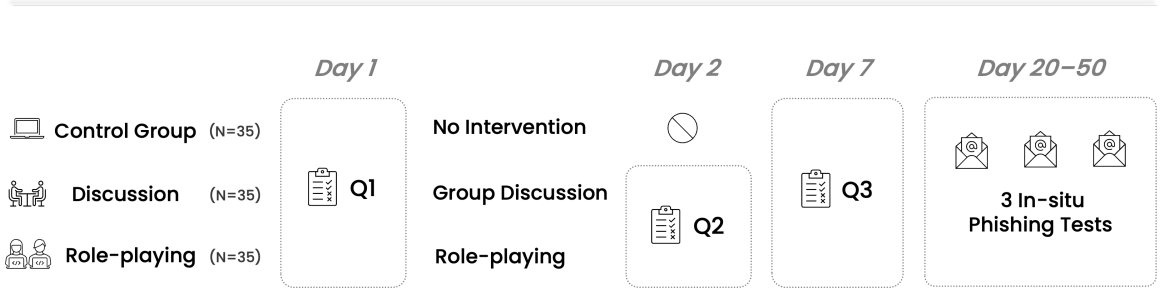

Figure 4.1: The experimental design: three conditions, repeated measures, and in-situ phishing tests.

Pre-training assessment. After we assigned registered participants to three conditions, we sent them study invitations with an attached information sheet and consent form for the study. We included a link to the first questionnaire (*Q1*) in the invitation. For the control Group participants, we instructed them to reply with a signed consent form to participate in the "Phishing Resilience Study". They could answer Q1 immediately after they sent us the consent form. For participants of two treatment conditions, we instructed them to fill out Q1 prior to their attendance of the training session.

Post-training assessment. Immediately after each group discussion and role-playing training session, we sent the second questionnaire (*Q2*) to the session attendees and asked them to complete Q2 within 24 hours. The control group participants did not receive Q2 because they had not taken part in an activity.

On the seventh day after each training session, we sent the attendees the third questionnaire (*Q3*) to measure knowledge retention and gather feedback on the training. Additionally, on the seventh day after the control group participants answered Q1, we sent them Q3 with adapted questions.

[3]We requested participants to indicate their availability among the 20 provided timeslots in the registration form. Timeslots that reached 8 or more participants were designated for group discussion or role-playing training randomly. For timeslots that had fewer than 8 participants, we assigned them to the control group.

In-situ phishing tests. To assess the impact of trainings on employees' real-life responses to phishing emails, three simulated phishing emails were sent to all study participants. We collaborated with two security experts who were in charge of phishing campaigns at the organization to design these tests with the themes of "Email client upgrade", "Data breach", and "Security alerts". Between 20 and 50 days after answering Q1, all participants received three phishing tests from the IT team. There were intervals of 6-7 days between each test, and participants received each test simultaneously. If a participant clicked the link within the phishing email, they would be directed to a webpage indicating that "you clicked on a simulated phishing test". We included one example of a phishing email and the webpage in the Supplementary Material. The tests were distributed in a sequence of easy, moderate, and difficult to identify as phishing[4]. Refer to figure 4.1 for an illustration of the study procedure.

4.3.4 Measures

We collected demographic information and participants' current responding "strategies or practices" when receiving suspicious emails in Q1. Demographic questions included participants' job positions, the organizational entities they worked for, starting date at the current organization, gender, and age group. Additionally, we employed scales and simulated phishing test records to evaluate the training effects and gathered feedback from participants.

Assessment scales. We included a self-efficacy (*SE*) scale and support-seeking (*SS*) scale (adapted version) in our pre- and post-training assessments (Q1, Q2, and Q3). Given the absence of a validated anti-phishing self-efficacy scale, we utilized two dimensions of self-efficacy sourced from studies with good construct validity and reliability. These dimensions evaluate distinct aspects of self-efficacy. Self-efficacy 1 (SE1) measured participants' confidence in learning and updating their knowledge of phishing attack techniques with three items [403], while Self-efficacy 2 (SE2) evaluated participants' confidence in recognizing suspicious emails with four items [272]. Meanwhile, we adapted the *Instrumental Support Seeking scale*[5] to evaluate participants' intention to seek support when receiving suspicious emails [147]. Three researchers reviewed and adapted the SS collaboratively. Afterward, one external expert assessed and confirmed that the adapted SS was more accurate in measuring support-seeking in the phishing context compared with the original scale. We used SE and SS with pilot study participants and received positive feedback. We include SE and SS in Appendix G.

Performance metrics. We used the number of reported and non-clicking on links within simulated phishing tests as indicators of participants' resilience to phishing emails. There are mixed findings

[4]Two researchers ranked the emails with the phishing scale [350], while one IT security expert relied on their experience with simulated phishing tests. They independently assessed and reached mutual agreement on the difficulty of identifying the three phishing emails.

[5]Instrumental Support Seeking scale: assesses the inclination to seek advice, information, and feedback from one's social network during stressful situations [147].

regarding the effectiveness of phishing campaigns as a form of intervention [229, 409, 177]. Nonetheless, in agreement with [66, 302], simulated phishing tests may be an ecologically valid evaluation method, reflecting participants' natural responses to phishing attempts, if the measures of success are carefully designed, especially in combination with other evaluation methods. We also recorded when participants reported the phishing tests to investigate how quickly they reported a simulated phishing email. We provided all study participants with the same instructions, "the IT department will send you three simulated phishing tests in the coming month. If you spot any suspicious emails, please report them: forward them as an attachment to report-a-phish@anonymized" (the standard reporting procedure at the organization).

Training feedback. In Q2, we asked the participants the following open-ended questions:

- What strategies or practices would you apply when receiving suspicious emails?
- Which aspects of the training, if any, do you consider useful for learning?
- How do you anticipate that this knowledge will help you in your work?

In Q3, for the control group, we collected their feedback on the phishing campaigns at the organization with the above three questions. For two treatment conditions, we prepared the following two questions:

- Please rate the effectiveness of the training in helping you defend against future phishing attempts. (Select from: Poor, Fair, Good, Very good, Excellent) Optional question: "*Please enter your comment*".
- How likely are you to recommend this training to a colleague? (Select from: Very unlikely, Unlikely, Neutral, Likely, Very likely)

4.3.5 Ethical considerations

The study design received approval (ERP 22-061) from the ethical review board before data collection. The first author signed a non-disclosure agreement with the organization to access the log data of simulated phishing tests and reporting records for the purpose of this study. We ensured that our study posed no potential harm to the participants. There was no dangerous link to follow, no malicious attachment to be downloaded, and nothing that could lead to a leak of personal information in our simulated phishing tests. The experimental design avoids any emotional distress due to doubts about having really fallen victim to a phishing attack. During the recruitment process, we informed the participants that there were two conditions (in-person and remote) and that they would be randomly assigned to one condition. We were transparent regarding the tasks they would perform in the

registration questionnaire. To compensate for participants' time commitment, we offered €25 gift vouchers to each participant of treatment conditions and a €10 gift voucher to each participant of the control group the day after we sent Q3 (even if they did not complete all the questionnaires). We informed participants about their right to opt out of in-situ phishing tests through the information sheet provided for all three conditions and reiterated this at the end of each in-person training session. We only used pseudonymous data in our analysis to protect participants' privacy [306]. We preregistered the study before launching it[6].

4.3.6 Data collection and analysis

Data collection. We conducted six sessions each for group discussion and role-playing training over 19 days in July 2023. We collected 105 complete answers from Q1 (all participants), 70 complete answers from Q2 (all training attendees), and 103 complete answers from Q3 (34 from the control group, 35 from group discussion, and 34 from role-playing). We have non-clicking and reporting records from the three in-situ phishing tests from 105 participants[7].

Preliminary analysis. Prior to our main analysis, we examined the scale validity and randomization of our group assignment. We converted the five-point scale SE1 into a seven-point scale to integrate the two dimensions of SE [349]. We examined the scales' factor structure and measurement validity following the recommendations of Kline and Schmitt et al. [324, 217]. First, we screened the data for missing entries or errors and evaluated item distribution. The data exhibited a non-normal distribution: 11 items showed skewness beyond the -1 to 1 range, and four items displayed kurtosis exceeding 5. Second, we specified measurement models for both scales and employed the maximum likelihood mean adjusted (MLM) method to estimate model parameters, addressing non-normality. Third, model fit was assessed using the Robust Comparative Fit Index (CFI), Robust Tucker-Lewis Index (TLI), Robust Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). A favorable fit is indicated by CFI and TLI values exceeding .9, and RMSEA and SRMR values below .08 [59, 208, 217]. Fourth, we examined modification indices for models with suboptimal fit to identify areas for enhancement. Only intra-latent covariances were introduced. Fifth, Cronbach's alpha was calculated to assess internal consistency reliability, with values exceeding .70 generally deemed acceptable. Lastly, to examine the randomization of our group assignment, we used the chi-square analysis and the Kruskal-Wallis test to check the distribution of demographic factors among the three groups in Q1.

[6]Pre-registration: `https://aspredicted.org/qi8ka.pdf`

[7]In addition to the aforementioned data, we captured video recordings of role-playing sessions and audio recordings of group discussion sessions. For role-playing, we collected 12 phishing emails designed by the participants. Furthermore, we gathered the completed templates for analyzing real phishing emails by group discussion participants. Due to length constraints, the analysis of these collected materials is deferred to future studies.

Main quantitative analysis. Given our relatively small dataset, which was not normally distributed and contained several outliers in each condition (refer to the box plot in Appendix K), we applied non-parametric analysis to the primary study variables (*SE* and *SS* scores). Therefore, we used Friedman's one-way repeated measures analysis (non-parametric analogy to repeated-measures ANOVA) for within-groups analysis to investigate whether the trainings have effects on participants [125]. We applied the Kruskal–Wallis one-way analysis of variance to compare the training effects between groups [79]. We applied the Bonferroni correction to all p-values obtained from post-hoc pairwise comparisons in our tests and present Bonferroni adjusted p-values in the findings [196].

We performed chi-square and Kruskal-Wallis tests to analyze the *phishing test* results. We conducted regression analyses, with the dependent variables being the sum of non-clicking and the sum of reported to examine whether the performance of non-clicking and reporting was influenced by the measured variables. We applied the non-parametric Mann–Whitney U test to analyze the differences between groups in perceived effectiveness and likelihood to recommend. Main and supporting quantitative data were analyzed and visualized with SPSS 28 and R (lavaan and ggplot2 packages). We provide an overview of main quantitative analysis methods and the corresponding research questions in Table 4.3.6. We include the anonymized dataset and SPSS syntax used for analysis in the Supplementary Material to allow verification and reproducibility.

Qualitative analysis. We conduct a qualitative analysis on the collected feedback (4.3.4) with MAXQDA [379]. The first author read through the feedback, took notes, generated codes from meaningful segments, and organized these codes into preliminary categories [219]. Subsequently, two other authors reviewed the formulated code system and held two discussion meetings to improve the clarity and precision of the code system. Following this refinement, the first author coded the feedback in MAXQDA. As part of quality assurance, another author examined the coded segments to ensure accuracy and consistency. Afterward, to compare differences in counter-phishing measures among three conditions, we employed MAXQDA to retrieve and visualize the frequency of "counter practices" coded in the participants' responses. Regarding the usefulness of the training, two authors thoroughly reviewed the coded segments and categorized them into four distinct themes [76]. We include the code system and example quotes in Appendix H.

4.4 Quantitative results

4.4.1 Preliminary analysis

Validity and reliability of the scales. The self-efficacy (*SE*) scale showed a good fit after adding covariance between two items ($\chi^2(12) = 17.757$, p = .123, CFI = 1.000, TLI = 0.999, RMSEA =

Table 4.2: Overview of research questions and corresponding quantitative analysis methods.

Research Question	**Analysis Method**	**To Examine**
RQ1: What is the effect of role-playing training on employees' anti-phishing *self-efficacy* compared to group discussion and the control group?	Related-samples Friedman's two-way analysis of variance by ranks	whether the trainings have effects on participants (4.4.2)
	Kruskal–Wallis one-way analysis of variance with Bonferroni-adjusted pairwise z-test for post-hoc analysis	the training effects between groups (4.4.2)
RQ2: What is the effect of role-playing training on employees' *support-seeking* intention when receiving phishing emails compared to group discussion and the control group?	Related-samples Friedman's two-way analysis of variance by ranks	whether the trainings have effects on participants (4.4.2)
	Kruskal–Wallis one-way analysis of variance with Bonferroni-adjusted pairwise z-test for post-hoc analysis	the training effects between groups (4.4.2)
RQ3: How do group discussion and role-playing training influence employees' *response* to phishing attacks?	Chi-square Analysis	whether there are difference in non-clicking and reporting between groups (4.4.3, Appendix I)
	Kruskal–Wallis one-way analysis of variance with Bonferroni-adjusted pairwise z-Test for post-hoc analysis	the differences in non-clicking and reporting in three simulated tests combined (4.4.3)
	Linear Regression analysis	the effect of SE, SS, and demographics on the performance of non-clicking and reporting (4.4.3)
Feedback analysis	Mann–Whitney U test	whether there are difference in "perceived effectiveness" & "likelihood to recommend" (4.4.4)

0.0144, SRMR = 0.023): "I am confident I can recognize a suspicious email" and "I am confident I can recognize suspicious email headers". The reliability of the SE scale was excellent in Q1, Q2, and Q3 (see Table 4.3). The support-seeking ((*SS*)) scale showed an acceptable model fit after adding covariance between two items ($\chi^2(19)=38.582$, p = .005, CFI = 0.939, TLI = 0.911, RMSEA = 0.091, SRMR = 0.066): "I try to talk and explain the suspicious elements of an email in order to get feedback from my colleagues" and "Before clicking anything within a suspicious email I'll talk with a colleague about it" (see Appendix L to see factor loading of both scales). The reliability of the SS scale was good in Q1, Q2, and Q3 (see Table 4.3).

Randomization check. The results from the chi-square (χ^2) analysis found no significant differences between the three groups in terms of gender proportion ($\chi^2(6, 105) = 2.850$, p = .827), faculty ($\chi^2(4, 105) = 6.648$, p = .156), and age group ($\chi^2(8, 105) = 11.855$, p = .158) in Q1. The Kruskal-Wallis test also revealed no significant differences in organizational tenure (H(2) = 2.05, p =

Table 4.3: Descriptive and Cronbach's alphas (α) for SE and SS scales.

	n missing	**Mean**	**SD**	**min**	**median**	**max**	α
Q1 SE	0	36.32	7.45	16	37	49	0.91
Q2 SE	35	39.89	7.32	10	41	49	0.93
Q3 SE	2	39.70	7.19	11	41	49	0.91
Q1 SS	0	22.23	5.05	8	23	32	0.83
Q2 SS	35	25.61	4.36	14	26	32	0.81
Q3 SS	2	24.06	5.45	12	24	32	0.89

Note: Q1 SE is the sum of the seven items in Q1, ranges from 7 to 49. Higher scores indicate higher self-efficacy. *Q1 SS* is the sum of the eight items in Q1, ranges from 8 to 32. Higher scores indicate higher support-seeking intention.

.359).

4.4.2 Training effects on SE and SS

Training effects compared within groups. We employed the Related-samples Friedman's two-way analysis of variance by ranks to assess if there are training effects on SE and SS in both trainings (refer to Table 4.4 to see the full results of the analysis).

Table 4.4: Related-samples Friedman's two-way analysis of variance by ranks.

Group Discussion (N = 35)

	Self-Efficacy		**Support-seeking**	
	$\chi^2(2) = 16.924$	Sig. < .001	$\chi^2(2) = 16.217$	Sig. < .001
	Pairwise comparison			
	Z-stat.	Adj.sig (Sig)*	Z-stat.	Adj.sig (Sig)
Q1-Q2	-2.749	.018 (.006)	-3.287	.003 (.001)
Q1-Q3	-3.884	< .001	-3.167	.005 (.002)
Q2-Q3	-1.135	.769 (.256)	.120	1 (.905)

Role-playing Training (N = 34)

	Self-Efficacy		**Support-seeking**	
	$\chi^2(2) = 8.835$	Sig. = .012	$\chi^2(2) = 25.878$	Sig. < .001
	Pairwise comparison			
	Z-stat.	Adj.sig (Sig)	Z-stat.	Adj.sig (Sig)
Q1-Q2	-2.304	.064 (.021)	-4.366	< .001
Q1-Q3	-2.425	.046 (.015)	-3.638	< .001
Q2-Q3	-.121	1 (.903)	.728	1 (.467)

* *Adj.sig (Sig)*: Bonferroni adjusted p-value (unadjusted p-value).

Immediate training effects (Q1-Q2): For group discussion condition, we found statistically

significant positive effects on both SE (adjusted $p < .001$) and SS (adjusted p = .003), comparing the measurements before and after the intervention. For role-playing training, we found statistically significant positive effects on the results of the SS (adjusted $p < .001$). However, we did not find significant immediate effects on the SE (adjusted p = .064).

Day 7 training effects (Q1-Q3): For group discussion, we found statistically significant positive effects on both SE (adjusted p = .018) and SS (adjusted p = .005), comparing the measurements before the training and on Day 7. For role-playing training, we also found statistically significant positive effects on the SE (adjusted p = .046) and SS (adjusted $p < .001$).

Training effects compared between groups. To compare the Day 7 training effects between the three groups, we analyzed the *deltas* (score difference between Q3 and Q1) in SS and SE with Kruskal–Wallis one-way analysis of variance.

Support-seeking: The analysis revealed a significant difference between three conditions regarding the deltas of SS (H(2) = 7.169, p = .028), as shown in Figure 4.2. The post-hoc analysis (Dunn Pairwise Z-Tests) identified a significant difference in deltas between role-playing and control Group (Z = 2.621, adjusted p = .026). We did not find significant differences between group discussion and role-playing (Z = -.845, adjusted p = 1)[8].

Self-Efficacy: The analysis did not reveal a significant difference between any groups regarding the deltas of SE (H(2) = 3.859, p = .145), as shown in Figure 4.3. A closer examination with Wilcoxon signed–rank test revealed that the control group increased significantly between day 1 and day 7. The p-value and effect size suggest that the increase is not only statistically significant but also of moderate magnitude (T = 338.5, Z = 2.18, p = .029, r = 0.37).

4.4.3 Phishing test results

Performance comparison. In Table 4.5, we present the amount of non-clicking and reported from the three simulated phishing tests. Chi-square analyses for each test revealed no statistically significant difference in non-clicking behavior among the three phishing tests. However, significant differences in reporting behavior were observed across the three tests (see Appendix I).

We examined the differences between three conditions in *non-clicking* (sum) and *reporting* (sum)[9] with the Kruskal-Wallis test. The analysis showed no statistically significant differences in non-clicking behavior (H(2) = .002, p = .999), but it did reveal statistically significant differences in reporting behavior (H(2) = 14.662, $p < .001$).

Post-hoc analysis with Dunn Pairwise Z-Tests revealed statistically significant differences in

[8]Similarly, we did not find any statistically significant difference in deltas (Q2-Q1) between two trainings SS (U = 532.5, p = .346), SE (U = 677, p = .467)

[9]The non-clicking (sum) represents the total non-clicks on the links of three phishing tests, while reporting (sum) represents the total number of reported phishing tests.

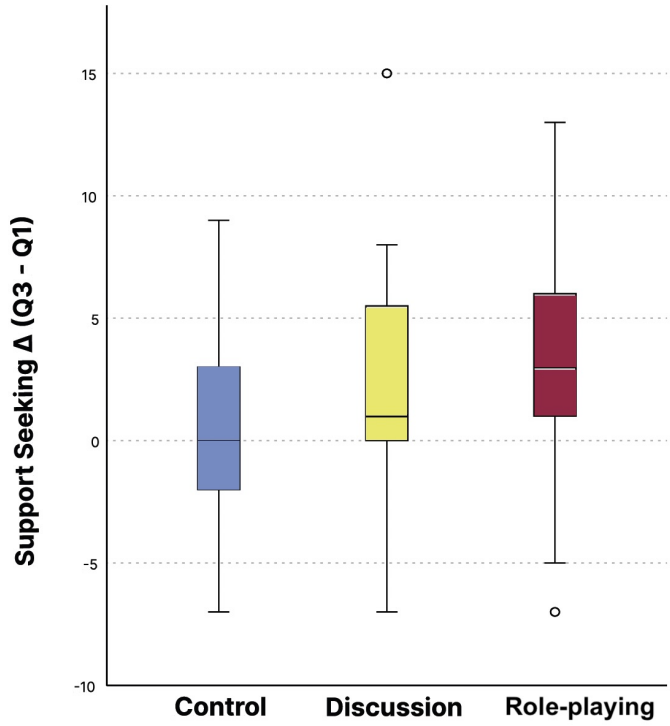

Figure 4.2: Kruskal–Wallis test of support-seeking deltas.

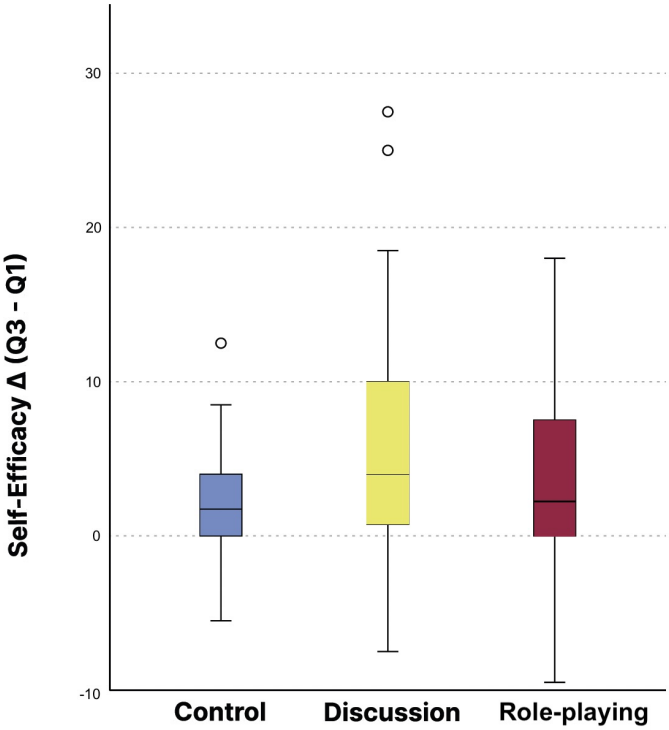

Figure 4.3: Kruskal–Wallis test of self-efficacy deltas.

reporting (sum) between group discussion and control group (Z = 3.235, adjusted p = .004) and between role-playing and control group (Z = 3.391, adjusted p = .002). However, no statistically significant differences were found between two treatment conditions (Z = -.156, adjusted p = 1).

Table 4.5: Number of participants (N) who did not click on the link within the simulated phishing test and reported it to the IT team. Each condition has 35 participants.

	Non-clicking			**Report-a-phish**		
	Client upgrade	Data breach	Security alerts	Client upgrade	Data breach	Security alerts
Control Group	35	33	34	3	7	8
Discussion	34	34	34	10	18	19
Role-playing	32	34	35	4	23	20

Several participants reported simulated phishing emails within minutes after receiving them. In the first test (Email client upgrade), one participant reported it within one minute, and three participants reported it between 6 and 10 minutes afterward. In the second test (Data breach), five participants reported within one minute, and eight participants reported within 5 minutes. In the third test (Security alerts), five participants reported within one minute, and six participants reported within 5 minutes.

Linear regression results. We estimated linear regression models to examine whether SE and SS predict non-clicking and reporting behaviors. We estimated two separate models for the dependent variables "non-clicking (sum)" and "reporting (sum)" [10], with the independent variables "SE (Q3)" and "SS (Q3)". As control variables, we included "working month", "gender", and "faculties". The regression results indicated that none of the independent and control variables in both full models had a statistically significant effect on the dependent variables. We provide the regression results in Appendix J.

4.4.4 Feedback analysis

The Mann–Whitney U test indicated that there were no significant differences between two trainings in perceived effectiveness (U = 725.5, p = .084), and an analysis of the means revealed high scores for both trainings. Specifically, group discussion had a mean of 4 (CI: 3.74-4.26), and role-playing had a mean of 4.29 (CI: 4.03-4.56). As for the likelihood to recommend (U = 550, p = .527), group discussion had a mean of 4.6 (CI: 4.39-4.81), while role-playing achieved a mean of 4.5 (CI: 4.27-4.73). This suggests that participants found both trainings highly effective and recommendable.

4.5 Qualitative findings

4.5.1 Changes in counter-phishing practices

In this section, we present a comparative summary of participants' counter-phishing practices as reflected in responses from the three questionnaires (4.3.4). We did not identify noticeable differences

[10] Contrary to our pre-registration, we use non-clicking and reporting as dependent variables, and SE and SS as independent variables. This adjustment is in line with findings from previous studies suggesting that *SE increases intention to report* [227].

among the three conditions when assessing the number of participants who specified specific categories of counter-phishing practices in Q1 (refer to Table 4.6). Furthermore, we noted only minor variations in the responses provided by the control group in Q1 and Q3. Considering these factors, we adopt the control group as the baseline for comparing coded segments between groups.

Table 4.6: Number of participants (N) who mentioned a specific category of counter-phishing practices; in each cell, a participant is only counted once.

Practices		**Control**	**Discussion**	**Role-playing**
Check email header	Q1	18	20	20
	Q2	N/A	18	14
	Q3	17	18	19
Evaluate email content	Q1	14	15	14
	Q2	N/A	16	20
	Q3	10	17	17
Do not respond	Q1	16	12	16
	Q2	N/A	10	8
	Q3	19	11	11
Block/report	Q1	6	13	11
	Q2	N/A	22	21
	Q3	7	18	21
Interact with colleagues	Q1	1	0	1
	Q2	N/A	7	7
	Q3	3	7	9

Identify phishing with header and content. More participants of both trainings mentioned checking email details to discern phishing compared to the control group (see Figure 4.4). The training attendees mentioned evaluating incoming emails based on "theme and content", "expectation and context", "read with caution", and "quality of the text" more than those in the control group. Additionally, more participants of role-playing training indicated reviewing specific email elements, such as "checking attachments", "verifying URL", and "analyzing requests" than group discussion.

Do not respond. After the training session, the group discussion generated the highest occurrence of "do not respond" subcodes among all groups in Figure 4.4. It is noteworthy that a few participants indicated "do not respond" and "report" at the same time. The response ("do not respond") does not necessarily indicate participants' exclusive way of responding to phishing emails.

Report phishing emails. The number of participants mentioning "report-a-phish" (report phishing emails to the IT team) increased noticeably after the training interventions (see Figure 4.4). However, only a few participants mentioned their rationale for this change in the questionnaire, including "informing the IT team" (P21), "detecting the phishing attempt" (P23), as well as "helping others" (P53).

Code System	Control	Discussion	Role-playing
Check email header			
Search online/official Web	5	11	5
Email subject	4	1	5
Verify sender	18	25	26
Evaluate email content			
Check attachment	3	4	7
Check URL	7	11	16
Analyze the request		5	11
Theme and content	4	7	7
Expectation and context	3	8	10
Read with caution	4	11	11
Quality of the text	6	12	11
Do not respond			
Do not click/respond	18	21	15
Delete	10	12	14
Block/Report			
Block the sender	1		3
Reporting	4	5	5
Report-a-phish	5	25	27
Interact with Colleagues			
Talk with colleagues	3	12	10
Inform my colleagues/friends	1	2	7

Figure 4.4: Number of Participants (N) mentioned specific counter-phishing practices across all questionnaires. To enable comparison between the control group and the treatment groups, in each cell, a participant is only counted once, even if they mentioned a topic in multiple questionnaires. Bright red indicates the largest number of the row.

Interact with colleagues. We noted an increased number of participants indicated talking with/informing their colleagues (friends) regarding phishing emails following the trainings (refer to Figure 4.4). This communication often involved *alerting colleagues* about phishing attacks (P11) or seeking support in determining whether the email they received was a phishing attempt (P43).

4.5.2 Usefulness of the training

In this section, we summarize our findings from analyzing the open-ended question answers regarding "the usefulness of the training" and "how this knowledge will help participants in their work" (4.3.4). We begin by presenting the utility values that participants attributed to both trainings. Following this, we describe the particular aspects emphasized within each training, the varied levels of enjoyment reported by the participants, and some negative effects of the trainings.

Utility values of the trainings. Participants gained knowledge of various phishing techniques,

communication channels, and attack tactics after both trainings. The training was "a good reminder of what to look for to assess phishing emails" (P42). Furthermore, participants emphasized their "vigilance" and heightened awareness against phishing threats (P7). Notably, a majority of participants expressed their intention to "respond with caution" when receiving emails (P21) and communicate with their colleagues when receiving suspicious emails. The practical examples and group discussion were perceived as useful, for "they allowed sharing knowledge on the latest phishing attacks" (P53).

Learning through group interaction. Participants from both trainings found discussion with colleagues useful in deepening their understanding and improving their skills to counter phishing attacks (seventeen participants from group discussion and twelve from role-playing). As P40 of group discussion commented: "Discussing topics in person helps to remember *anecdotes and stories from others* that will be helpful in similar future events." To clarify, participants in both trainings spent around 40 minutes in discussions. The role-play participants focused on discussing how to craft a phishing email, while the group discussion mainly involved exchanging phishing-related experiences.

Thinking like a hacker is useful. Fifteen role-playing training attendees referred to the task of designing a phishing email as useful, as it requires participants to *think from the perspectives of hackers* (P18), identify vulnerabilities of the organization (P11), and examine the elements of phishing emails (P14). Some participants became more aware of the complexity of phishing attacks and attack techniques after role-playing, as P30 indicated:

> Designing phishing email forces us to check how to proceed, and reminds us how the hackers are proceeding/thinking. Really interesting. Adding two normal emails among the phishing emails is really good, as we really have to check and we realized that this is really difficult to see the truth in an email. So this impacted my vision of phishing.

Varied levels of enjoyment. Seven participants from role-playing training mentioned that the training was *interesting/fun*, while four participants from the group discussion stressed that it was interesting for them to learn about phishing techniques or analyze phishing emails. In contrast, four participants from both trainings commented that they were not the target recipient of this training. This opinion is associated with their high self-efficacy scores and perceived medium effectiveness of the training (P13 and P68). As exemplified by P19: "The workshop was excellent in teaching employees about phishing. However, as a computer scientist, I do not know if I was the right target for it."

Negative effects of the trainings. A dozen participants from both trainings lowered their level of self-efficacy after training. Through the questionnaire responses, we get to know that in P8's case, their lowered levels of self-efficacy were due to the realization that it was easier than they thought to create a phishing email and quite difficult to distinguish the phishing email from legitimate ones in the

role-playing training, and "ignoring suspicious emails remains the easiest thing to do, and I cannot ask for support all the time" (P8). Additionally, in the cases of P27, P36, and P58, they self-reported quite high self-efficacy scores prior to the training and lowered their scores after the training.

4.5.3 Summary of results

We conducted a mixed-design experiment to assess the effectiveness of role-playing and group discussion training within and between groups. Combining the quantitative and qualitative analysis, we found that:

- **RQ1**: Group discussion yielded a significant improvement in perceived *anti-phishing self-efficacy* in both the immediate and Day 7 assessments. The role-playing training did not demonstrate a significant improvement in the immediate assessment but did so for the Day 7 assessment. Both trainings worked similarly well when comparing their effects on anti-phishing self-efficacy on Day 7 (no statistically significant difference in effects between the two trainings). However, these improvements of both trainings did not reach statistical significance when compared to the control group.

- **RQ2**: Group discussion and role-playing training significantly enhanced *support-seeking intention* in the immediate and Day 7 assessments. Again, both interventions worked similarly well when comparing their effects on Day 7 (no statistically significant difference in effects between the two trainings). However, only role-playing training showed statistical significance when compared to the control group.

- **RQ3**: Both trainings were effective in prompting employees to *report* phishing emails between 20 to 50 days after the training. Employees in the treatment conditions reported phishing tests statistically more often than employees in the control condition. We did not find a difference in *non-clicking* between groups, but clicking numbers might have been too low to detect differences. In terms of qualitative results, both trainings enhanced employees' vigilance towards phishing emails, increased reporting intention, and promoted their intention to interact with colleagues when receiving phishing emails. Both trainings were perceived as highly effective and were highly recommended by the employees.

4.6 Discussion

4.6.1 Training effects on anti-phishing self-efficacy

Our research reveals that after participating in a group discussion training, employees perceived higher levels of anti-phishing self-efficacy in the immediate and Day 7 assessment (see 4.4.2). When we compare the two trainings, group discussion participants spent more time *sharing knowledge* on the latest phishing attacks and discussing *anecdotes and stories* than in the role-playing training (see 4.5.2). A recent online experiment by Hull et al. might help explain why employees perceived higher self-efficacy after group discussions, finding that anti-phishing training with stories resulted in higher self-efficacy and more accurate phishing detection than training using mindfulness techniques [190]. In addition, sharing knowledge is anticipated to reduce the likelihood of information security risks [363]. Security stories from others can serve as informal lessons, and these stories impact people's thought processes and corresponding behaviors when making security-relevant decisions [296]. Our findings suggest group discussion can be an effective method for delivering rich narrative/story-based anti-phishing training.

In the immediate assessment, the role-playing training did not yield a significant improvement in anti-phishing self-efficacy. However, participants in the role-playing training showed statistically significant improvements in the Day 7 assessment. This lag points to the importance of measuring effects of anti-phishing interventions beyond the initial effects immediately after an intervention. This lag suggests that participants might not have immediately grasped how role-playing as hackers could be relevant in their professional contexts. Building upon the work of previous role-playing studies [397, 118, 413], our research applies repeated measures to assess the effectiveness of role-playing anti-phishing training, thereby deepening our understanding of its impact.

Self-efficacy is an important step that connects *concordance* and *skills* in the *Security Learning Curve* [321], and is consistently linked to more secure behavior [376]. Previous literature on self-efficacy in cybersecurity often emphasizes measuring employees' self-efficacy, with limited focus on enhancing self-efficacy through cybersecurity training [45, 154]. Two previous studies reported that cybersecurity conference attendance and instructor–led training resulted in higher perceived self-efficacy [298, 215]. Our findings contribute to the research on phishing interventions by offering empirical evidence that both trainings can be effective methods to boost employees' anti-phishing self-efficacy. Notably, the enhancement of self-efficacy did not reach statistical significance compared to the control group. One plausible explanation could be the presence of a "ceiling effect", as a substantial number of employees already possessed high self-efficacy scores prior to the training (see Table 4.3).

4.6.2 Seeking support when encountering phishing

Our findings suggest that group discussion and role-playing training demonstrated statistical significance in increasing employees' support-seeking intention when receiving phishing emails in the immediate and Day 7 assessments. While both trainings worked similarly well in elevating employees' support-seeking intention in the Day 7 assessment, only the role-playing training showed significant improvement compared to the control group (see 4.4.2). Role-playing as hackers helped employees scrutinize their organization's vulnerabilities, collaborate with their colleagues, and critically assess the elements of phishing emails (see 4.5.2). Some employees may have benefited from this process by grasping the ease with which phishing emails can be created and the difficulty of discriminating these emails from legitimate ones (see 4.5.2). Such realizations could help dispel the stigmas and shame often associated with being phished [305], potentially increasing the inclination to seek support when receiving phishing emails.

Furthermore, the role-playing training we experimented with in this study resembles the "Red Teaming" approach, which simulates real-world attacks to fortify organizational resilience against security threats [122]. Thinking like a hacker enables organizations to anticipate potential threats and take proactive risk reduction measures [115]. Role-playing has been found to be an effective approach to improve phishing detection accuracy and confidence [397, 335], enhance phishing awareness, develop phishing knowledge, and stimulate conversations about phishing [30]. However, only a few previous studies engage participants in playing the role of attackers in social engineering/phishing interventions. Role-playing as attackers facilitates employees' experience of social engineering attacks without deception [36], educates students about the harms of excessive online information disclosure and the risks of spear-phishing attacks [118], and improves employees' phishing detection abilities [413]. We investigated whether employees can benefit from adopting a hacker's mindset and whether this shift enhances support-seeking intention. As such, we contribute to the literature by demonstrating that role-playing as hackers is an effective approach to enhance employees' support-seeking intention upon receiving suspicious emails.

We introduced *support-seeking intention* as a useful measurement to assess anti-phishing trainings for several reasons. Given the prevalence of phishing attacks targeting individual email accounts, employees often face these threats in isolation. When employees seek support from their colleagues upon receiving phishing emails, it not only informs the team about the incident but also allows the support-seeker to receive valuable assistance in responding safely to the threat. Furthermore, this support facilitates employees' acquisition of helpful response strategies, developing their ability to handle phishing emails in the future [321]. Lastly, cultivating a strong support-seeking intention can pave the way for collaborative efforts in countering phishing, which is a recognized, effective, and

indispensable approach for mitigating security breaches within organizations [319].

4.6.3 Reporting phishing emails at organizations

Group discussion and role-playing training increase employees' intention and behavior in reporting phishing emails at organizations (see 4.5.1 and 4.4.3). Our study supports the notion that reporting can serve as an effective crowd-sourced strategy to counter phishing attacks [229], as a substantial number of employees reported each phishing test within minutes. Both trainings led to increased numbers of employees mentioning specific counter-phishing practices that they would perform in the future (refer to Figure 4.4). This more elaborate process of evaluating phishing emails has been found to link to a greater likelihood of reporting suspicious emails [52]. Employees considered reporting as a means of detecting phishing attempts, informing the IT team of the attacks, and helping others avoid being phished. Various elements of our trainings might be responsible for these favorable outcomes, such as perceiving the severity of phishing attacks by analyzing real phishing emails, which warrant further investigation. In addition to the training approach, sharing reporting statistics might engage employees with simulated phishing campaigns and motivate them to report suspicious emails [181].

Our study contributes to the domain of mitigating phishing attacks at organizations. It is inevitable that employees will interact with phishing emails, considering advanced phishing tactics such as "spear phishing" [102]. Rather than focusing on the cases where employees interacted with a phishing email, we suggest putting the focus on building a collaborative security culture where employees are encouraged to report incidents proactively. Reporting phishing is one of the most effective methods for the IT team to detect attacks that bypass technical safeguards, and increasing the number of such reports is crucial [20]. However, previous studies indicate that most employees do not participate in reporting [20, 227]. Reporting phishing emails has been used as an indicator to evaluate individuals' responses to phishing and to assess the effectiveness of training [181, 52, 229], but it has rarely been considered as one of the objectives of training. In this study, we created two trainings that effectively encourage employees to report phishing emails to the IT team, and were able to empirically demonstrate the effectiveness of these interventions through simulated in-situ phishing tests.

However, encouraging reporting is insufficient to implement the notion of employees as human firewalls within organizations. Organizations should provide training opportunities to employees who might lack the skills to identify phishing emails. On the other hand, publicly acknowledging reported incidents (e.g., through an employee message board) and validating reported emails are necessary to facilitate reporting as a crowd-sourced defense [200]. Organizations should implement a coherent and consistent protocol for how employees are expected to react to phishing attacks and what response they can expect after reporting. Otherwise, if employees do not perceive the efficacy of their responses,

they might feel demotivated to contribute to organizational security [402, 170].

4.6.4 An enjoyable and effective anti-phishing training

While our study finds group discussion and role-playing are effective anti-phishing trainings regarding the measurements we evaluated, we identified some constraints when implementing the group discussion approach. As detailed in our training design (see 4.3.1), group discussion was planned for groups of 4-6 participants, allowing all participants to exchange and share their experiences. For groups larger than six participants, it is advisable to split them into two subgroups, which necessitates a spacious room and an expert to facilitate and answer questions in each group. In contrast, for the role-playing training, the ideal subgroup size is 3-4 participants. Thus, even 12 attendees can be divided into three subgroups. As long as the necessary computers, fictitious personas, and email accounts are prepared in advance, a single expert can facilitate and address questions for all groups. While group discussions also serve as an effective and interactive training method, they appear to be more demanding in terms of expert facilitation.

We are aware of discussions within the security community on the effectiveness and sustainability of anti-phishing trainings, and on the burden they impose on employees, when not generating hidden costs, or even damage, to a company's productivity (*e.g.*, see [229, 50]). However, the decision to engage employees in training ultimately resides with the company's management, ideally, in cooperation with the security department. In this case, those in favor of the training have to decide which instrument better serves the purpose of providing some level of preparedness against phishing attacks. In this regard, the choice of what security training to rely upon matters; bad training design choices may offer nothing more than a miserable user experience, providing content that is often *perfunctory* and *arcane*, detached from an employee's practices and expertise [338]. In short, they are far from being enjoyable and motivating.

The choice of which training to put employees through can therefore make a significant difference in terms of enjoyability, and consequently, engagement, which may result in varied short-term effectiveness. In line with [30, 118], we find that employees consider role-playing *enjoyable*. More employees in the role-playing group mentioned that the training was interesting than those in the group discussion. When the learning experience is enjoyable, learners are more likely to engage with training and apply what they've learned [190, 338]. Thus, we propose using role-playing as an enjoyable and effective approach for anti-phishing training, and potentially for cybersecurity training. (We provide all materials we used in the Supplementary Material to enable others to use and iterate on our approach.)

4.6.5 Methodological considerations

We employed a mixed-design experiment to evaluate the effects of two anti-phishing trainings in the field. Some methodological considerations are relevant for researchers and practitioners conducting social engineering studies.

- *Self-reported change*: Our study assesses the impact of training on participants' perceptions, focusing on two scales, self-efficacy (SE) and support-seeking (SS), changes in self-reported counter-phishing practices, and perceived usefulness of the training. We examined these factors both immediately after the training session and one week later. To the best of our knowledge, we are the first to adapt and measure support-seeking (SS) in phishing intervention studies.

- *Behavioral change*: We also examined behavioral changes after the training through in-situ phishing tests. We utilized non-clicking and reporting behaviors as indicators of employees' phishing resilience. These behaviors directly impact an organization's resistance to phishing attacks.

- *Informed consent*: Our findings suggest that it is possible to get meaningful results in social engineering studies with obtaining prior informed consent. Prior informed consent is important for ethical research practices, as it ensures that participants are fully aware of the study objectives and potential risks. However, some researchers and practitioners might worry about prior consent influencing self-reported and behavioral measurements, potentially making participants more alert. In the present study, we used two strategies to counteract the potential influence of informed consent: (a) we standardized the information concerning phishing tests provided for all conditions, and (b) we introduced a waiting period ranging from 13 to 30 days between the compensation email and the first simulated phishing test. In a longitudinal setting, it is likely that participants go back to their typical behaviors even if they know they will receive phishing tests at some point over the course of weeks. The daily work tasks claim participants' attention, enabling the observation of natural behavior without deception. We hypothesize that for social engineering research, which often uses deception [103], further longitudinal study designs might lessen the necessity of using deception in certain experimental setups.

4.7 Limitations and future work

Our study has a number of limitations. Given that the retention period of training lasts a maximum of five months [197], further examining residual training efficacy in longitudinal field studies is required to evaluate our training. We measure anti-phishing self-efficacy, support-seeking intention,

non-clicking and reporting, perceived effectiveness, and likelihood of recommending the training in this study. While self-efficacy has been shown to increase the intention to report phishing emails [227], in our study, neither self-efficacy nor support-seeking intention significantly influenced participants' reporting behaviors when all factors were considered in the regression models. Other factors, such as perceived severity and response efficacy, might merit empirical investigation in future phishing intervention studies.

We recognize that emphasizing self-efficacy in anti-phishing training might lead some employees to become overconfident, potentially reducing their accuracy in detecting phishing emails [90]. Therefore, we incorporated real examples of phishing emails and engaged employees in discussing potential consequences to strengthen their vigilance against phishing threats. In future work, we plan to address the potential side effects of overconfidence bias. Additionally, while stories from colleagues are easily understandable and memorable, the quality and accuracy of such information depend on the narrator's security literacy. Insights from previous studies on "folk models" [391] and "security stories sharing" [393] may illuminate strategies to enhance the effectiveness of group discussion as a training method.

Our study was conducted at a single university, which might have a unique context, organizational culture, and demographic composition compared to other types of organizations. The results could differ when applied to corporate settings, government agencies, or non-profit organizations. Further investigation is needed to assess the applicability of our findings to these diverse contexts. It is possible that we unintentionally attracted tech-savvy participants to register for our study. For future studies, a more systematic sampling method should be devised. Prior informed consent might, to some extent, have influenced our observed results of non-clicking and reporting; we suggest future studies to empirically compare our approaches of counteracting the potential influence brought by informed consent with other creative approaches.

There might be richer insights to be gained from analyzing participants' discussions and phishing email designs, which could shed light on the organization's vulnerabilities and improve anti-phishing training. Due to page constraints, we could not present the findings in this paper. We plan to publish these results in a follow-up paper.

Organizations deploy phishing campaigns for mainly three objectives: a) examining organizational vulnerability; b) as a form of awareness training; and c) evaluating the effectiveness of an intervention [383]. The trade-off between the costs of potential successful attacks and the costs of simulated phishing campaigns is a complex concern. It was outside of this paper's scope to evaluate the costs of training and evaluation measures. In our case, the implementation of in-situ tests entailed considerable coordination and effort from our collaborators and participants. This included four formal meetings

and over twenty email correspondences with the Information Security Office to work on the design, testing, and deployment of the simulated phishing emails. This necessitated substantial commitment from the Security Office of both time and expertise, potentially causing interruptions to their workflow. Moreover, we are aware of the time costs and extra workload incurred by our study participants. Future studies should consider these hidden costs associated with deploying phishing tests for evaluation, such as the investment in personnel time and the utilization of IT infrastructure [50].

4.8 Conclusion

In this study, we employed a mixed-design experiment to assess the training effects of group discussion and role-playing, involving measurements such as anti-phishing self-efficacy, support-seeking intention, and responses to phishing attacks. Our findings reveal that both trainings were effective in enhancing perceived self-efficacy and support-seeking intention in the Day 7 assessment. However, only role-playing significantly enhanced support-seeking intention compared to the control group. Both trainings contributed to an increase in reporting simulated phishing emails and safe responses to phishing emails.

Our study contributes to a better understanding of both group discussion and role-playing as effective and interactive anti-phishing training approaches. Also, our study underscores the significance of discussing phishing incidents and sharing anti-phishing practices in the workplace, which can enhance employees' self-efficacy, support-seeking intentions, and vigilance against phishing threats. The study contributes to the field of mitigating phishing attacks by presenting two trainings that effectively prompt employees to report "phishing emails" to the IT team. Furthermore, we introduce support-seeking (SS) as a useful measurement to evaluate phishing interventions and devise a promising novel methodology to examine training effects within and between subjects, measuring both self-reported and behavioral changes in the field. Our study demonstrates the feasibility of obtaining informed consent from research participants for simulated phishing tests while still gaining valuable insights from the results.

4.9 Acknowledgments

Author 1 acknowledges the financial support of the Institute for Advanced Studies at the University of Luxembourg through a Young Academic Grant (2021). We would like to extend our gratitude to Vincent Koenig and the HCI Research Group, who were instrumental in the conception, design, and pretest of this study. Additionally, we appreciate our colleagues from the Information Security Office, Steve Cannivy and Laurent Weber, for their dedication to the successful realization of this study. We

thank the ACs and reviewers for their constructive feedback.

Chapter 5

Empowering Parents to Support Children's Online Security and Privacy: Findings from a Randomized Controlled Trial

To appear: Chen, X., Distler, V., Gordon, C., Yao, Y., & Teuber, Z. Empowering Parents to Support Children's Online Security and Privacy: Findings from a Randomized Controlled Trial. In Proceedings of ACM Conference on Computer and Communications Security (CCS '25). ACM, New York, NY, USA, 15 pages.

Abstract In the ubiquitous computing society, parenting "digital natives" presents unprecedented challenges. Parents often rely on online resources to support and guide their children in security and privacy (*S&P*) related topics. However, the abundance of online resources makes it challenging for parents to find high-quality and relevant resources that align with their S&P needs. Further, the longitudinal development of parental competence and coping strategies in S&P topics remains largely unexplored.

We conducted a formative study with 210 U.S. parents of children (M_{age} = 11.73 years, SD = 3.15) to investigate the challenges parents face in educating children about online S&P topics and to inform the design of a remote intervention program (six short videos). In the main study, we evaluated this intervention's efficacy using a 14-week longitudinal randomized controlled trial, which consisted of 201 U.S. parents, with 113 assigned to the control group and 88 to the intervention group.

We found that short videos significantly enhanced parents' security awareness and their conversation strategies. Notably, parents who initially exhibited lower levels of these measurements benefited the most from the intervention. Moreover, short videos were effective in enhancing parents' self-efficacy in protecting their children from online risks. This study provides valuable insights into various challenges parents face and respective coping strategies that could be implemented to address S&P concerns in family settings. The design and evaluation of the intervention program serve as a foundation for future S&P researchers and educational stakeholders.

5.1 Introduction

Internet usage among children[1] has dramatically increased, with one in three Internet users worldwide being a child, and more than 175,000 children going online for the first time every day [2, 372]. Although the cyber world offers abundant opportunities for learning and creativity, it also presents

[1]Definition of a child: any individual under the age of 18 [373].

various security and privacy (S&P) risks for children, such as harvesting of sensitive personal data and exposure to cyberbullying and harmful content [240]. For instance, about 55% of high school students in the United States have experienced online bullying or harassment [283]. Another German survey revealed that malware, account compromises, online shopping fraud, and data abuse are prevalent security threats among teenagers [172]. These S&P incidents can severely disrupt children's academic and socioemotional development and contribute to mental health issues.

We refer to *online security for children* as the practices, tools, and measures implemented to protect children from threats such as unauthorized access, cyberattacks, or other abnormal activities in digital spaces. Relevant topics include safeguarding children's digital devices, accounts, and communications to ensure confidentiality, integrity, and availability [347, 2]. *Online privacy for children* can be conceptualized as controlling the collection, use, and sharing of children's personal information in digital spaces, focusing on protecting sensitive data (e.g., identity, browsing habits, and communications) from unauthorized access or misuse without consent [342, 247]. Finally, *online safety for children* refers to protecting children from online threats that could harm their physical and mental health [355]. When a child's online security or privacy is compromised, it may lead to safety issues. For example, when smart security cameras installed in children's bedrooms are compromised [355], predators may exploit them to harass children remotely.

For children, including those considered "digital natives" [291], acquiring online S&P skills is a developmental learning process. Teaching these skills requires a shared responsibility between schools and families [242, 10]. Even for young educators, they do not adequately possess S&P knowledge and are not prepared to educate students on these topics [292]. Many educators rely on search engines to find relevant resources, whereas more experienced educators have accumulated trusted resources (e.g., slides, curricula, and videos) or seek recommendations from their colleagues [254]. Teachers might intentionally refrain from teaching certain topics they perceive as sensitive or uncomfortable [254], such as sexting. These points underline the critical role of family members in shaping children's S&P practices [172].

Many parents might themselves have limited S&P literacy, thereby influencing the extent to which they expose children to digital devices and informs their mitigation strategies [355, 242]. Prior studies [141, 14] have examined a number of off-the-shelf tools that enable parents to manage security and privacy in family settings. Parental control apps can enable parents to monitor children's unhealthy behaviors and keep them safe; however, these tools have also been criticized as overly restrictive and, in some cases, an invasion of children's privacy [141]. Further, researchers have indicated that vulnerabilities in parental control tools could be exploited and lead to security issues [14], such as device compromise and account takeover. Parents need to scrutinize the data such apps collect and

evaluate the trustworthiness of their makers [16]. Consequently, the development of parents' S&P literacy is essential for the oversight of children's digital devices and management of off-the-shelf tools.

While considerable efforts have been made to address S&P issues in schools [221, 63], *how to support parents in developing awareness and coping strategies to address S&P risks in a family setting remains largely underexplored* [240, 241]—despite the fact that many S&P threats occur with children's digital devices outside of school premises and hours. The present study aimed to fill this research gap with a mixed-methods approach. We conducted a formative study to inform the design of a remote intervention program and a main study to evaluate the efficacy of the intervention. This work makes the following contributions:

- The creation of our intervention program provides a scalable solution that can be widely implemented to support parents in managing children's S&P in a family setting. We provide the interventions as supplemental materials for other researchers to adapt and use.
- The empirical validation of meaningful metrics for evaluating interventions offers a rigorous approach for future S&P studies, enabling reliable and consistent evaluation of similar interventions. We provide these metrics in the appendix for future use.
- The evaluation offers novel empirical insights into the effectiveness of parental support programs and we provide recommendations on how they can be tailored to address real-world concerns.

In the sections that follow, we review existing work on S&P education in family settings, explore these issues from a developmental psychology perspective, and present our research questions and hypotheses.

5.2 Related Work

5.2.1 Mitigation strategies for S&P risks and available educational resources for parents

Children encounter various online threats, such as cyberbullying, predators, invasion of privacy, inappropriate content, financial scams, hacking, and viruses [267]. Therefore, it is important for parents to have open communication with children about their online activities [139]. Another approach to mitigating these risks is to use parental monitoring applications to limit screen time, access to certain websites, and downloads. However, prior work found that parents had doubts about the feasibility and functionality of monitoring teenagers [294]. Researchers have suggested that parental monitoring may foster a perceived sense of parental control over children's internet use, which

could lead to negative outcomes, such as lowering their guard to potential risks [139]. Researchers suggested that parents should support children in managing S&P issues rather than managing them on behalf of children [220].

Schoolteachers, parents, and governmental bodies are primary stakeholders for children's cyber security education [320]. Children commonly gain cybersecurity knowledge from family, friends, and other trusted individuals [267, 172]. Family conversations about S&P serve as an important mean for shaping children's S&P literacy [10]. Alghythee et al. [10] identified the following five conversation approaches: (a) rule-based, (b) example-based, (c) decision-making process-focused, (d) consequence-based, and (e) contextual conversations. While rule-based conversations were the most commonly reported approach by parents [10], tailoring S&P discussions to specific applications and combining multiple conversational strategies in parent-child interactions may be more effective than relying on a single approach. This requires parents to be knowledgeable about S&P topics and possess effective communication skills [311], which is not always the case. Enhancing parents' technical skills and self-efficacy is essential to overcoming these barriers [240]. Unfortunately, many resources designed to educate S&P topics appear to focus primarily on the school environment/children's needs [240, 412] and are inadequate for use in family settings. One potential solution is the development of literacy programs that specifically support parents in their role as guardians [240], equipping them with communication strategies for S&P conversations. To structure the skills parents needed to oversee digital natives, Romero [311] proposed a Parental Digital Literacy Framework. The framework highlights four sub-sets of skills [311]: (a) basic skills to manage privacy, content and technology, (b) communication skills, (c) creativity, problem-solving, attention and self-regulation skills, and (d) life-long learning.

Interventions aimed at developing parents' S&P literacy are scarce [241]. Prior media literacy interventions, including those targeting parents, have demonstrated effectiveness in enhancing parents' awareness of media influence, critical evaluation of media content, and domain-specific knowledge [205]. Informal sources, including interpersonal stories, news articles, and online content offering security advice [295], play an important role in shaping individuals' security practices. Other approaches to develop parents' S&P literacy include in-person workshops [169], parent-teacher conferences (hosted by school, municipality, or church), peer-to-peer learning, and community-based learning [241]. Additionally, parents can learn about S&P topics through video platforms, online communities, and educational content created by NGOs and governmental bodies [3, 311]. For example, NGOs like Common Sense Media and Family Online Safety Institute (both US-based) deliver curricula and strategies to promote safe and secure digital engagement within families. However, research has identified several unresolved challenges in using these online materials to effectively

develop digital literacy, including difficulties in finding the relevant content, insufficient tailoring to varying literacy levels, and inadequate adaptation to the needs of the target audience [11].

5.2.2 Family security and privacy from a developmental psychology perspective

Building on the definition used in parenting research [150], we define *family S&P education* as parental practices aimed at developing children's S&P awareness in online activities and safeguarding them from S&P risks. Decades of parenting research suggest that parents' support for their children's development in specific domains has a positive impact on overall child development [206]. Specifically, in the context of S&P, emerging studies highlight the pivotal role of family education in equipping children with the necessary skills to navigate digital challenges [240, 293, 151]. We focused on parents with children (in Grade 6-9), a group classified as "early adolescence" [259]. This developmental stage marked by significant biological, cognitive, social, and emotional changes that influence parent-child relationships [259]. During this period, children increasingly seek independence and become more involved in activities outside the family, often accompanied by reduced parental supervision. Parents navigate a renegotiation of authority and relationship boundaries with early adolescents, which can lead to conflicts [47] and pose challenges for parents in providing effective support during this critical stage [317]. On one hand, parents must demonstrate perseverance and consistency to achieve their parenting goals despite setbacks and difficulties. On the other hand, they must adapt to their children's evolving needs while maintaining these goals [146].

Previous quantitative studies have attempted to investigate how parenting behaviors impact their children through the lens of the four classical parenting styles: authoritative, authoritarian, neglecting, and permissive [252, 31]. These styles are defined by the combination of two key parenting dimensions: responsiveness (warmth) and demandingness (behavioral control). Authoritative parenting is characterized by high responsiveness and high demandingness, while permissive is characterized by high responsiveness and low demandingness. Since the 2000s, researchers [149] have acknowledged the critical role of autonomy—the sense of psychological liberty and the perception of self-directed choice, reflecting an individual's ability to act upon their own values and internal will [95]. Parental monitoring as a form of behavioral control may not be appropriate to guide early adolescents who seek for independence and autonomy; thus, parents should adapt their digital parenting strategies and styles according to their children's developmental stages.

The dynamic management of boundaries between teenage technology users, the outside world, and their parents has long been a complex issue. In the human-centered security and privacy community, researchers have investigated children's privacy and security, investigating topics such as connected toys [255], password behaviors [366], misinformation [334], reactions to phishing [231]. Cranor et al.

[82] investigated parents' and teenagers' views of privacy in semi-structured interviews, comparing their differing views on boundaries. The authors found that parents struggled to conceptualize their teenagers' privacy boundaries, for instance, the extent to which text messages and apps were considered private by their kids. Children considered their phones more private than computers. Thus, intervention programs that target parenting strategies in the domain of online S&P in children need to respect their privacy. This aligns with research highlighting the importance of parents having conversations about decision-making thought processes with their children [10]. Parental approaches that balance guidance with autonomy—such as providing rationales rather than enforcing strict rules—help children develop critical thinking and skills in managing online risks.

In summary, it is important for parents to be aware of the various risks their children are exposed to when in their online endeavors. An effective intervention program to support parenting should align with their children's developmental stages. Such a program should not only provide parents with actionable toolkits but also remind parents to respect early adolescents' privacy and support them in developing their S&P competence.

5.3 Research Objectives

The overarching goal of our study was to develop a first version of a S&P intervention aimed at parents and to evaluate its efficacy. In the long run, we hope that this effort supports the research community and practitioners in working towards an evidence-based S&P intervention program for the family environment. To achieve this, we conducted a formative study and a main study. Through the formative study and post-intervention questionnaire, we sought to address the first research question (open and descriptive):

RQ1: What considerations should be taken into account when designing intervention programs to support parents in educating their children about online security and privacy?

The main study aimed to evaluate the effectiveness of the intervention program using a randomized controlled trial. According to the Protection Motivation Theory [309, 244], two key cognitive processes contribute to individuals' behavior change: (a) threat appraisal, which involves parents' awareness of the severity of online risks and children's vulnerability to such risks, and (b) coping appraisal, which refers to parents' belief in the efficacy of protective actions and their ability to perform them. Specifically, within the human-centered security community, Sasse et al. [321] emphasized the indispensable role of self-efficacy in individuals' adoption of new security behaviors. If parents have higher self-efficacy in guiding and protecting their children on S&P topics, they are more likely to implement related S&P practices. Thus, to empower parents in supporting their children's S&P, it is essential to enhance both parental awareness and parenting self-efficacy. This led to the second

research question:

RQ2: How does a remote intervention influence parents' awareness, self-efficacy, and coping practices (e.g., parental mediation strategies and conversation approaches) in educating their children about security and privacy?

To test the intervention's effectiveness, we hypothesized the following:

H1: Parents in the intervention group will demonstrate a greater awareness of their children's online risks than those in the control group.

H2: Parents in the intervention group will report higher self-efficacy in supporting their children with online challenges than those in the control group.

H3: Parents in the intervention group will employ a broader range of mediation strategies and conversation approaches (or use them more frequently) compared to those in the control group.

H4 (exploratory): Levels of parental concerns about their children's online S&P may not differ between groups, as parents in the intervention group, while becoming more aware of potential risks, may also develop self-efficacy in managing them.

5.4 Methods

5.4.1 Formative study

The formative study utilized a questionnaire comprising both open-ended questions and scale items, and it served two key purposes. First, it aimed to identify parents' difficulties and needs in supporting their children's online S&P. This was achieved through qualitative thematic analysis, with the findings directly informing the design of the intervention program to ensure its relevance and practicality. Second, the formative study assessed the measurements intended for use in the main study to evaluate the program's effectiveness. This step was critical, as some of these measurements had not undergone rigorous psychometric validation in prior research (e.g., [113]). Consequently, items on some scales were adapted or removed based on their theoretical alignment and psychometric properties, assessed through internal consistency estimates (Cronbach's alpha and McDonald's omega) and confirmatory factor analysis (*CFA*).

Participants

Parents residing in the United States with at least one child in Grades 6-9 were eligible to participate in this study. In September 2024, a total of 210 U.S. parents (53.33% mothers, n = 112) participated in the formative study via the online platform *Prolific*, with a mean age of 42.46 years (SD = 8.65). Approximately 62% held at least a bachelor's degree, and 80% reported being in a committed

relationship. When answering questions about their children, parents were instructed to refer to their youngest child in grades 6 to 9. Among the referred children, 103 (49%) were girls and 107 (51%) were boys, with a mean age of 11.73 years (SD = 3.15). The socioeconomic status, assessed using the highest parental socio-economic index (HISEI, [276]), was notably higher than the U.S. average (M = 65.32 vs. 55.7; [303]).

Challenges Parents Face

To address RQ1, we asked study participants, *Could you please elaborate on the difficulties you encountered when teaching your child about online security and privacy topics (e.g., online risks and privacy settings of apps)*? After removing ten empty answers, we recorded 200 valid answers. Three authors each familiarized themselves with 50 unique answers and then jointly created a first bottom-up categorization of meaningful codes [48]. Following this, the first author created an integrated coding scheme and coded all answers with MAXQDA [379], no new codes generated in this process. 65% of the answers were double-coded by another two authors; inter-rater reliability was analyzed using Cohen's κ, demonstrating substantial agreement (κ = 0.79). We include the coding scheme in Appendix M. 38 parents (19%) reported they had not encountered any challenges. In the following paragraphs, we present the four main categories of challenges that we identified:

Parental concerns about online risks (n = 67): Parents expressed concerns associated with their children's online activities. Many children were perceived to be overly trusting and did not always heed warnings (P2, P56), believing that strangers they meet online are who they claim to be and will be kind to them (P27, P32, P37). This over-trust might lead to dangerous situations, as children may share personal information or engage with strangers (P9, P21, P32). Additionally, parents worried that children often did not understand the potential dangers of the internet, feeling invincible due to a lack of negative experiences (P8, P68, P151). Parents also struggled with their children's overconfidence and resistance to their advice and stated that children believed they were more tech-savvy than their parents (P35, P96, P162).

Complexity of S&P topics from parents' perspective (n = 49): Many parents did not consider themselves knowledgeable in S&P topics (P64, P79, P200) or lacked confidence in explaining these topics (P13, P112, P135). Some indicated that they were not tech-savvy enough to teach their children, while others found it hard to keep up with new applications their children use and different privacy settings (P35, P85). Many parents also felt overwhelmed by the constantly evolving digital landscape, making it challenging to stay up-to-date on emerging threats (P1, P15, P75, P186). P163 expressed the need for "a reputable resource that stayed on top of things and offer digestible information."

Parenting challenges (n = 62): Many parents struggled to attract their children's attention and

Table 5.1: Summary of topics parents wanted to learn about.

Categories	**Occurrence**	**Exemplary Topics**
Online privacy-related topics	95	Privacy protection, privacy settings, keep personal information safe, sharing personal content/info/nudes online, digital footprints, keeping accounts private, and secure personal data
Online security-related topics	90	General security concerns, and specific threats (including hacking, phishing, scams, email/website account security, false identities, police involvement, passwords, virus, and ransomware)
Parental monitoring and parenting strategies	84	Setting parental controls, track kid's online behavior, location tracking, technology to protect children, websites that kids can access, good online habits, mental mindset, general tips; and how to teach children recognizing/avoiding online threats, securing their accounts, resources for teens to communicate with other youth, online safety for teenagers
Online protective actions	78	Websites should avoid, examples of dealing with specific websites; mental health and peer relationship issues, including smartphone addiction, cyberbullying; and safe social media usage, grooming, catfishing, and identity thief
Online harmful content	63	Pornography, racism, fake news, internet danger, drugs, bots on social media, hoax news, harmful websites, and misinformation
Current trends and emerging risks	12	AI-generated content, deepfake, sexting, latest applications and technologies

interest when having S&P conversations with them (P12, P54, P97). While parents acknowledged the need for parental monitoring, they expressed technical constraints, ethical concerns, and reluctance from their children (P22, P28, P109). Lastly, a few parents found S&P conversations awkward (P114), or they did not feel close enough to their child to discuss such topics (P146); as P44 described "(My child) doesn't want to talk to me about stuff like this. If I push it, she gets irritated and leaves the room."

Parents' critical reflections (n = 11): A few parents found it challenging to balance trusting and monitoring (P76), "teaching them and them tuning you out" (P77), and "fostering independence while ensuring security" (P153). Furthermore, in some cases, the social environment seemed to cause parents difficulties in pursuing S&P in the family; for example, P206 was frustrated by the insecure password policy suggested by their child's school, and the parents of their child's friends were extremely permissive when it came to digital devices, causing serious strain and power struggles between P130 and their child. Lastly, a couple of parents reflected on the challenges posed by tech companies and the lack of proper legislation (e.g., luring the attention of children, P36; lack of privacy protection for teenagers in the US, P29).

Intervention Design

Research indicates that short videos can be an effective training approach to engage trainees with S&P topics [38, 254]. Short videos can deliver a lot of content in a short time, catch viewers' attention, and have both short- and long-term training effects [38]. Short videos are an effective format for explaining complex S&P concepts [211, 340]. Furthermore, short videos are more cost-effective for large-scale deployment than in-person training or games [412, 340], which require more resources and logistics. Lastly, short videos can be accessed via different digital devices and have the potential to seamlessly integrate into parents' daily media consumption. Given these advantages, we selected educational short videos as the intervention approach in this study.

In the formative study, we surveyed the parents: *We're developing a program to help parents protect their children's online security and privacy. If you were to join this program, which three topics would you most like to learn about?* We received valid answers from 208 participants. After removing the irrelevant answers and merging similar topics, we summarize topics that parents want to learn in Table 5.1. Informed by the topics and challenges reported by parents, we developed six themes to address in our video series. It was not feasible to address every topic indicated by parents; therefore, we focused on categories-level and highlighted relevant S&P risks alongside respective coping strategies. Notably, our study participants often mixed online safety with security, prompting us to include certain online safety topics—such as avoiding online predators and harmful content—in our video episodes. Besides parents' expectations, we carefully estimated the potential impact that our intervention might have on their children (primarily "early adolescents"). We verified all the examples prior to including them in the videos with reputable media outlets. When we provided suggestions in the videos, they were based on peer-reviewed publications or relevant institutions (e.g., the Family Online Safety Institute). Considering early adolescence is a stage of seeking increasing levels of autonomy [317], we emphasize open communication in the family environment and parents' role in supporting children to develop their ability to manage their digital devices and accounts. We avoided emphasizing restrictive or authoritarian parenting styles and intentionally presented early adolescents' perspectives (e.g., using sarcasm, criticism, and yelling in serious conversations reduces effective communication [40]).

The production of each episode followed a structured process: defining the goal of each episode (see Appendix N), gathering relevant scientific papers and news stories, outlining the structure, drafting the script, recording audios, and completing visual editing. We used stock image/video/audio from Envato and Pexels, with Final Cut Pro as our editing tool. Two researchers with a background in educational psychology and media production and one media producer collaboratively created the following six short videos (in total 29 minutes):

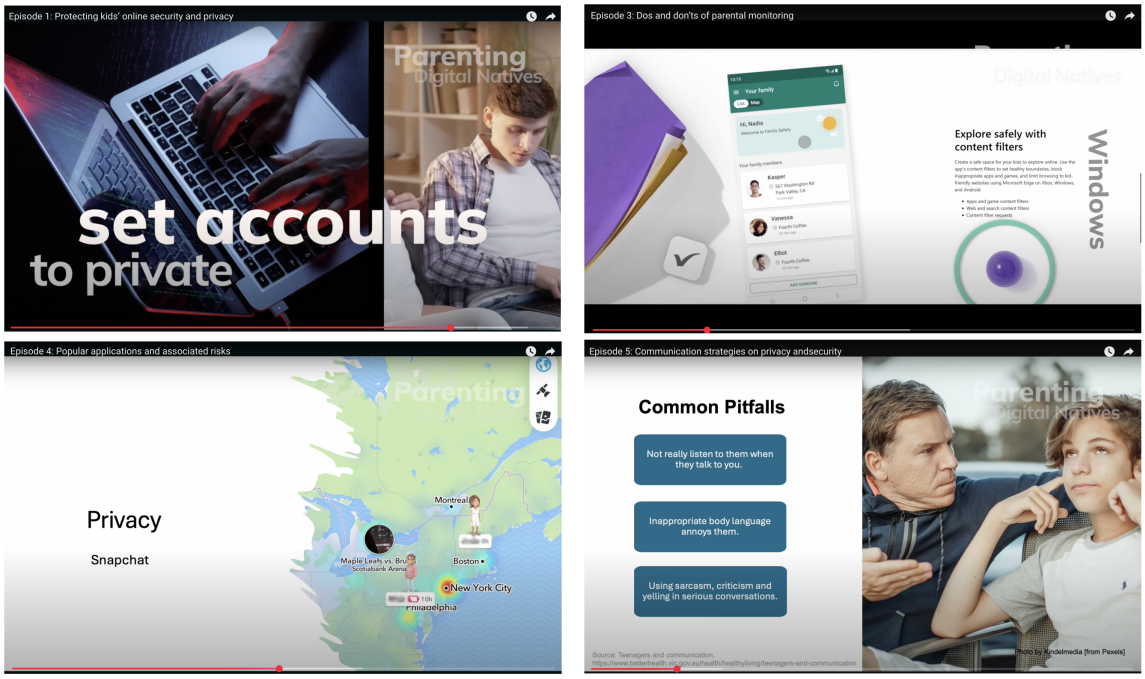

Figure 5.1: Screenshots of Episode 1, 3, 4, and 5.

Episode 1: Protecting kids' online security and privacy (3:51). In the introduction episode, we provide easy-to-understand definitions for online security [2] and online privacy [247] and examples relevant to children's online activities, including social media profile settings, meeting new friends online, online password protection, and downloading free games from unknown websites.

Episode 2: Seven steps to good digital parenting (5:00). We use educational videos created by Family Online Safety Institute (FOSI) as learning material in this episode[2]. The video emphasizes the following key steps to good digital parenting: a) talk with your child about what they are doing online, b) educate yourself to stay updated with technology, c) use parental controls to manage your child's online experience, d) set ground rules and enforce them, e) friend and follow, but don't stalk them, f) explore, share, and celebrate with your child, and g) be a good digital role model.

Episode 3: Dos and don'ts of parental monitoring (4:51). The episode starts with an overview of different ways of monitoring children's online activities, i.e., parental control applications (benefits and risks of using such apps [6]), setting rules and boundaries, and having weekly conversations with their child on their online activities. We introduce the key features of default parental control apps on iOS, Android, macOS, and Windows systems. The video ends with a call for direct communication within the family to support children in navigating the digital world safely.

[2]We received authorization from FOSI to use their videos for research purpose. Link to the video series: `https://www.fosi.org/how-to-be-good-digital-parent`

Episode 4: Popular applications and associated risks (5:23). The episode first lists the 11 most popular apps among American teenagers, as reported by the Pew Research Center [23]. Then, we describe some potential risks associated with using these applications, such as account security threats, cyberbullying, online privacy issues, harmful content, online safety, and digital well-being, referring to various news articles.

Episode 5: Communication strategies on privacy and security (4:55). This episode highlights three strategies that parents can use to talk to their teenagers about online privacy and security (presented in [10]): sharing your own experiences to teach your child about the decision-making thought process, discussing the consequences of online actions, and tailoring advice to what your child actually uses (contextual conversations).

Episode 6: Emerging risks and parental support (5:12). We introduce emerging risks associated with children's usage of generative AI (identified by Yu et al. [410]), the improper use of deepfake technology, and sextortion targeting teenagers. We suggest parents have open and non-judgmental conversations, develop digital literacy alongside their child, and provide reassurance and support to mitigate emerging risks. The video ends with a recap of the key topics we introduced in the past five episodes.

Measurement Validation

Table 5.2 gives an overview of the measurements evaluated in the formative study, including a brief description of each construct, the number of facets and items, and goodness-of-fit. These measurements were informed by prior empirical studies [10, 143, 3] and grounded in established theoretical frameworks [244, 309]. All adapted measurements are included in Appendix O.

Parental security awareness (parental awareness). Parental cybersecurity situational awareness was assessed using a scale developed by Ahmad et al. [3], comprising six items (e.g., "I am aware of what my child is accessing"). Participants responded on a five-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). During analysis, the fourth item ("I realize how difficult it is to control my child's internet usage") displayed negative correlations with all other items and a negative factor loading in the CFA. After removing this item, the scale demonstrated a one-factor structure.

Parental concerns about security and privacy of their children (parental concerns). Parental concerns about children's online security and privacy were assessed using a 13-item scale informed by findings from the Australian Government's eSafety program [113]. The items addressed a range of potential online risks for children, such as "being called insulting names." Responses were recorded

Table 5.2: Overview of the measurements evaluated in the formative study.

Construct	**Description**	**Factors (Items)**	**Response**	α/ω	**CFA Goodness-of-fit**
Parental awareness [3]	It evaluates the level of a parent's awareness to protect their children from online risks.	1 (5)	1-5	.82/.83	$\chi^2(4) = 8.96$, $p = .06$; CFI = .99; RMSEA = .08, 90% CI [.00, .15]; SRMR = .03
Parental concerns [113]	It measures parent's concerns about negative online experiences that might happen to their child online.	1 (13)	1-5	.82/.83	$\chi^2(4) = 8.96$, $p = .06$, CFI = .99, RMSEA = .08, 90% CI for RMSEA [.00, .15], SRMR = .03
Parenting self-efficacy [143]	It measures the extent to which a parent believes they can influence children's online behaviors and prevent them from online risks.	1 (8)	1-5	.93/.93	$\chi^2(19) = 48.04$, $p < .001$; CFI = .95; RMSEA = .09, 90% CI [.06, .11]; SRMR = .04
Parental mediation [143]	It measures the frequency with which a parent uses different mediation strategies to manage their children's online activities.	4 (8)	1-5	.95/.95	$\chi^2(19) = 48.04$, $\chi^2(48) = 97.177$, $p < .001$; CFI = .97; RMSEA = .07, 90% CI [.05, .09]; SRMR = .04
Conversation approaches [10]	It assesses the extent to which a parent applies five conversation approaches when discussing S&P topics with their children.	5 (20)	1-5	.96/.96	$\chi^2(160) = 384.38$, $p < .001$; CFI = .90; RMSEA = .08, 90% CI [.07, .09]; SRMR = .07

on a five-point Likert scale (1 = not at all, 5 = very much). A confirmatory factor analysis indicated an acceptable fit for a one-factor model.

Internet-specific parenting self-efficacy (parenting self-efficacy). We measured parents' confidence in managing their children's exposure to online risks with a scale developed by Glatz et al. [143]. Participants rated their confidence on eight five-point Likert items (e.g., "How confident are you in your ability to prevent your child from coming into contact with dangerous individuals?", 1 = *extremely unconfident* to 5 = *extremely confident*). The results of a CFA indicated a good fit for a one-factor model.

Internet-specific parental mediation (parental mediation). The items were adapted from the parental internet-specific mediation scale [143] to measure the frequency with which parents employ different mediation strategies to manage their children's online activities. The scale encompassed four dimensions. The first dimension, restrictive mediation, referred to parental rules about online activities and the use of monitoring software programs, including items such as, "How often do

you use filtering software installed on your child's devices?" The second dimension, demands child disclosure, assessed how often parents require their children to provide information about their online activities, with items like, "How often do you demand to know which websites your child has visited?" The third dimension, active mediation, measured the frequency of discussions with the child about internet use, as captured by items such as, "How often do you talk to your child about what they are doing on the internet?" Finally, proximity refers to being physically present while the child uses the internet without actively interfering, with items such as, "How often do you sit with your child while they are online?" Answers were rated on a five-point Likert scale (1 = *never*, 5 = *always*). We performed a four-factor CFA model, which showed close fit.

Parental conversation approaches on security and privacy (conversation approaches). To evaluate the various approaches parents use when discussing S&P topics with their children, we developed a scale comprising five dimensions, based on a qualitative study [10]. The first dimension, rule-based conversations, involved parents discussing rules for digital privacy and security with their child (e.g., "I regularly discuss online safety rules with my child"). The second dimension, example-based conversations, focused on using illustrative examples to explain privacy and security concepts (e.g., "I show my child phishing emails I receive to help them recognize online scams"). The third dimension, decision-making thought processes, measured how parents guided their child in making rational decisions about digital privacy (e.g., "I explain to my child how I make choices about keeping our online information safe"). The fourth dimension, consequence-based conversations, emphasized highlighting the outcomes of privacy-related actions to raise the child's awareness (e.g., "I actively discuss with my child the potential risks and consequences of their online actions"). Lastly, contextual conversations involved discussing privacy and security issues specific to particular apps or platforms used by the child (e.g., "I discuss privacy issues with specific apps and platforms my child uses"). Items were evaluated on a five-point Likert (1 = *strongly disagree*, 5 = *strongly agree*). The five-factor CFA model demonstrated an acceptable fit.

Covariates: security attitudes and online privacy concerns Participants' general security attitudes were assessed using a scale developed by Faklaris et al. [117], consisting of six items. An example item is "I seek out opportunities to learn about security measures that are relevant to me." A unidimensional CFA model fit the data well. For online privacy attitudes, we adapted a scale validated by Buchanan et al. [51]. We selected six items related to online activities from the original scale (e.g., "Are you concerned about online organizations not being who they claim they are?"). The fifth item was removed as suggested by both scale analysis and CFA. Thus, the scale showed a one-factor structure with excellent goodness-of-fit.

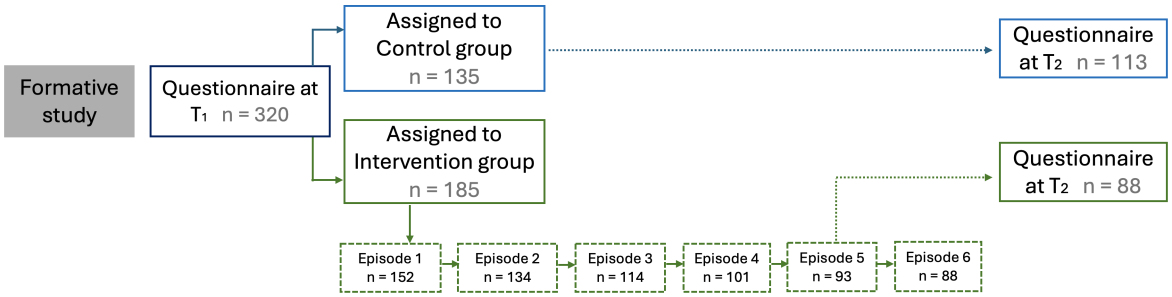

Figure 5.2: Flowchart of our experiment design: 14-week interval between timepoint 1 (T_1) and timepoint 2 (T_2).

5.4.2 Main study

The main study was a randomized controlled trial (*RCT*) with a 2 (Treatment Condition: Intervention or Control) × 2 (Time: pre and post) design conducted over a 14-week period. Thus, it consisted of three phases: a pre-questionnaire, a 12-week remote intervention, a post-questionnaire. During the 12-week interval, participants in the intervention group received a short video and a feedback questionnaire every two weeks via *Prolific*. The intervention was evaluated both during the program (via a short questionnaire after each episode) and afterward (through pre- and post-intervention test comparisons). Refer to Figure 5.2 for an illustration of our experiment design.

Data collection and participants

Before data collection, we estimated the optimal sample size using a prior power analysis with the program *G*power* [119]. With the expectation of a medium effect size of η^2 = .06 as a more conservative estimate, a total sample size of 128 (i.e., 64 in each group; using the F-test) should be sufficient to detect main and interaction effects (1-β = 80). Taking possible attrition into account ([114] shows an attrition rate up to 30% in an in-person intervention program) and recognizing that online interventions assume even higher attrition rates (e.g., due to reduced personal engagement or technical barriers), an oversampling of N = 320 was reasonable.

In September 2024, 320 U.S. parents with at least one child in Grades 6-9 were invited to participate in the study via the online platform *Prolific*. Of these, 135 parents were randomly assigned to the control group, and 185 to the intervention group. All participants completed a pre-questionnaire upon enrollment. Subsequently, parents in the intervention group received a *Qualtrics* survey every two weeks via their Prolific accounts. Each contact point included informed consent, one video episode, a multiple-choice question to review the key points of the episode, feedback questions, and a playlist link to all previous videos. We did not provide specific instructions to parents after videos, as family contexts varied (e.g., device use and family rules). The videos were hosted on YouTube for its wide

accessibility and lack of account restrictions for viewing. In comparison with the alternative approach of asking parents to subscribe to a channel on their preferred video platforms, this controlled setting enabled us to isolate, to a certain extent, contextual variability across platforms, which facilitated evaluations that were less confounded by differences in media platforms.

In Episode 1 of the intervention, 160 out of 185 parents participated; however, the responses of eight parents were excluded due to a completion duration shorter than the length of the video (n = 152). In Episode 2, seven parents dropped out, and the responses of eleven parents were removed for short durations (n = 134). In Episode 3, 16 more participants dropped out, and 17 participants' responses were excluded due to rapid completion times (n = 114). Episode 4 saw eight additional dropouts, along with five responses removed for short completion durations (n = 101). Finally, in Episode 5, five parents dropped out, and the responses of three parents were excluded for the same reason (n = 93). We invited participants who watched all of the first five episodes and all participants assigned to the control group to complete the post-questionnaire. This time, five participants from the intervention group and 21 from the control group dropped out. See Appendix P for the dropout rate table.

Thus, the sample for the quantitative evaluation of the intervention's efficacy consisted of a total of 201 parents (111 mothers), with 88 in the intervention group and 113 in the control group. On average, participants were 44.89 years old (SD = 8.71) and had 2.19 children (SD = 0.96) in their household. The average HISEI score was 64.95 (SD = 17.53), which exceeded the U.S. average level. No statistically significant differences were found between the intervention group and control group in terms of sociodemographics, including age, gender distribution, number of children, and socioeconomic status. For the qualitative feedback analysis, we included participants who took part in each episode. Although some of them dropped out in later episodes, their suggestions remained constructive for improving our intervention.

Measurements

Scale evaluation: To evaluate the efficacy of the intervention, we assessed the following measurements in both pre- and post-questionnaires:

- Parental security awareness
- Parental concerns about S&P of their children
- Internet-specific parenting self-efficacy
- Internet-specific parental mediation
- Parental conversation approaches on S&P
- *Covariates*: general security attitude and privacy concerns

Single question feedback: Participants' subjective evaluation of each episode was assessed using the following indicators: (1) the perceived usefulness and clarity of the video (e.g., "Please rate the usefulness/clarity of the video"), rated on a five-point Likert scale ranging from 1 (*Poor*) to 5 (*Excellent*); and (2) the likelihood of recommending the video to others (e.g., "How likely are you to recommend this video to another parent?"), rated on a five-point Likert scale ranging from 1 (*Very unlikely*) to 5 (*Very likely*).

Open-ended questions: Participants were invited to provide suggestions for improving each episode by responding to an open-ended question: "Do you have any suggestions for improving the video?" Additionally, we asked participants after Episode 6, "Are there any other topics you would like us to include in future Parenting Digital Natives series?" Lastly, in the post-questionnaire, we asked the participants, "Have you tried any of the recommended strategies? If so, could you briefly share your experience with us?"

5.4.3 Ethical considerations

The research project received approval from the Ethics Review Panel of the University of Luxembourg before implementation. Participation was entirely voluntary, and informed consent was obtained from each participant. Detailed information about the study's purpose was provided to participants, and they were reminded of their right to withdraw at any time without needing to give an explanation. For the main study, on average, participants spent 10.88 minutes in the questionnaire at pre-questionnaire; for the questionnaire at post-questionnaire, participants spent 9.40 minutes. We provided $10 to each participant who completed the pre- and post-questionnaires. The intervention group participants further received $2 for attending each session (5-6 min per session). Additionally, a bonus of $5 was awarded to those who attended all of the first five sessions. Participants could also request their data be withdrawn at any stage without facing any negative consequences. All data sets were identifiable only through a pseudo-anonymous identifier, accessible exclusively by *Prolific*. This identifier was deleted after data matching.

5.4.4 Data analysis

In the preregistration, we initially planned to use analysis of variance (ANOVA) with repeated measures. Instead, we decided to use generalized linear regressions to investigate the effect of the intervention. This method is flexible in handling missing values and random effects, accommodates unbalanced groups, and enables the inclusion of covariates and additional predictors. Compared to ANOVA (including non-parametric forms of ANOVA, like ART-ANOVA), it provides richer insights into the effect size of predictors [358]. We used MAXQDA to analyze qualitative data [379]. We

followed an inductive thematic analysis approach [48] to familiarize ourselves with the data, generate meaningful codes, categorize coded segments, and summarize descriptively the key findings. The coding and theme development were driven by the data content, which ensured that the participants' voices remained central in the data analysis.

Table 5.3: Mean (*SD*) of measurements and t-tests of between-group difference. Note: CG for control group; IG for intervention group. * p <.05; ** p <.01; *** p <.001.

Variable	**Pre-questionnaire**			**Post-questionnaire**		
	CG	IG	*t*	CG	IG	*t*
Parental awareness	4.16 (0.61)	4.31 (0.5)	-1.86	4.14 (0.54)	4.3 (0.42)	-2.35*
Parental concerns	3.12 (1.04)	3.01 (1.12)	0.73	3.3 (1)	3.13 (1.13)	1.09
Parenting self-efficacy	3.31 (0.81)	3.5 (0.72)	-1.82	3.27 (0.85)	3.59 (0.73)	-2.83**
Parental mediation	3.13 (0.8)	3.16 (0.91)	-0.21	3.14 (0.81)	3.25 (0.78)	-0.98
Conversation approaches	3.82 (0.74)	3.87 (0.7)	-0.52	3.85 (0.71)	3.97 (0.61)	-1.27
Consequence-based conversations	4.03 (0.86)	4.13 (0.74)	-0.92	4.06 (0.76)	4.21 (0.58)	-1.57
Decision-making thought process	3.82 (0.87)	3.9 (0.86)	-0.63	3.83 (0.81)	4.06 (0.66)	-2.16*
Security attitude	3.8 (0.68)	3.88 (0.78)	-0.81	3.81 (0.69)	3.98 (0.69)	-1.74
Privacy concerns	3.52 (0.96)	3.43 (1.02)	0.64	3.65 (0.96)	3.58 (1.01)	0.55

5.5 Results

5.5.1 Intervention efficacy

In both pre- and post-questionnaires, we examined parental awareness, parental concerns, parenting self-efficacy in preventing children from online risks, parental mediation, and parental conversation approaches. Furthermore, we measured two covariates: general security attitude and privacy concerns. Table 5.3 presents the descriptive statistics for all measurements. The Shapiro-Wilk tests revealed that, except for parental mediation, all other measurements were not normally distributed. *For between-group comparisons*, we used t-tests. Notably, the measurements at pre-questionnaire showed no statistically significant differences between the control group and the intervention group. At post-questionnaire, the intervention group demonstrated significantly higher levels of parental awareness, parenting self-efficacy, and decision-making thought processes (a key dimension of parental S&P conversation approaches) compared to the control group. *For within-group comparisons*, we conducted Wilcoxon signed-rank tests. Among our target outcomes, while the increase of parental awareness, parental concerns, and parenting self-efficacy in the intervention group did not reach statistical significance, decision-making thought processes demonstrated significance (V = 637.5, p = .03). In terms of covariates, we observed a significant increase in privacy concerns from pre-questionnaire to post-questionnaire for both the control group (V = 1671.5, p = .02) and the intervention group (V = 798, p < .05).

We used generalized linear regressions to investigate the effect of the intervention. In all regressions, each target outcome at the post-questionnaire was predicted by its (a) respective value

at pre-questionnaire, (b) general privacy concerns at pre-questionnaire, (c) security attitude at pre-questionnaire, (d) condition/group assignment (i.e., 0 = *CG*, 1 = *IG*), (e) the interaction a*d, and (f) whether any strategies learned in the intervention were applied (1 = *applied*, 0 = not *applied*; *63* indicated applied). Furthermore, we controlled for sex and age.

The results showed that the main effect of condition was not significant for parental concerns, overall parental mediation, or general conversation approaches, indicating that being in the intervention group did not influence these factors. However, significant main effects of the intervention were observed for parental awareness, parenting self-efficacy, and consequence-based conversations and decision-making thought processes—two specific aspects of conversation approaches. Regarding the main effects, the intervention increased parental awareness (B = 1.38, SE = 0.42, t = 3.26, $p < .01$), parenting self-efficacy (B = 1.18, SE = 0.46, t = 2.53, $p < .05$), consequence-based conversations (B = 0.82, SE = 0.37, t = 2.22, $p < .05$) and decision-making thought processes (B = 0.87, SE = 0.35, t = 2.53, $p < .05$). In terms of time × condition interaction effects, a significant negative effect was found between pre-questionnaire parental awareness and intervention (B = -0.31, SE = 0.10, t = -3.11, $p < .01$); also between pre-questionnaire consequence-based conversations and intervention (B = -0.20, SE = 0.09, t = -2.22, $p < .05$). Additionally, a significant negative interaction was identified between pre-questionnaire decision-making thought processes and the intervention (B = -0.20, SE = 0.08, t = -2.41, $p < .05$). See Appendix Q for regression tables. These findings indicate that the intervention had a diminishing effect on gains in awareness, consequence-based conversations, and decision-making thought processes as baseline levels increased, suggesting a stronger impact on parents with lower initial scores and a leveling effect at higher initial scores.

Key results. The video series effectively increased parental awareness and parenting self-efficacy. While it did not significantly impact overall parental mediation or conversation approaches, a closer examination of specific facets revealed that the intervention increased the frequency of consequence-based conversations and sharing decision-making thought processes.

Table 5.4: Mean (SD) of feedback questions of each episode. Note: "-" indicates the metric was not measured for the episode.

	Episode 1	Episode 2	Episode 3	Episode 4	Episode 5	Episode 6
Clarity	-	4.39 (0.72)	-	4.38 (0.68)	4.46 (0.65)	4.39 (0.70)
Usefulness	4.01 (0.84)	-	4.12 (0.78)	-	-	-
Likelihood to recommend	4.05 (0.98)	3.95 (1.10)	4.00 (1.04)	4.17 (0.93)	4.17 (0.97)	4.25 (0.93)

5.5.2 Intervention feedback

The mean value of clarity/usefulness for all episodes reached a score above 4, indicating "very good" in delivering the planned topics of the episode. With a mean of 4.46, participants of Episode 5 generally found the video to be between "Very good" and "Excellent" in terms of clarity in explaining the different communication strategies for privacy and security. Regarding the likelihood to recommend, except for Episode 2, all other episodes achieved a score of 4 or above, indicating they would "likely" recommend the respective episode to other parents. We include the mean (SD) of feedback questions of each episode in Table 4.4.4.

The high scores in clarity, usefulness, and likelihood to recommend align with the predominantly positive feedback we received in response to the open-ended question, "Do you have any suggestions for improving the video?" Of the 544 responses, 352 (65%) indicated that they had no suggestions for improvement. The most frequently mentioned keywords in the positive responses were: "great" (45 times), "informative" (29 times), "high quality" (27 times), "excellent" (17 times), "useful/helpful" (16 times), "concise/clear/straightforward" (14 times), and "easy to understand" (14 times). Additionally, in 76 instances, participants provided suggestions for improving the video series, focusing on the following aspects:

More engaging narrative: Most participants praised the audio for being clear and understandable. In all episodes, we had a native American male from the Bay Area as the narrator. However, some participants suggested that the pacing could be faster (E4P16, E4P85, E5P92[3]). A few participants criticized the monotone narrative style as lacking engagement (E1P63, E2P9). E4P64 specifically recommended a female voice to better engage the audience. Surprisingly, E1P93 misperceived the narrator as an AI voice, noting that "the pronunciation came off weird quite a few times, as well as the inflection." These suggestions highlight the important role of the narrator in engaging the audience.

Better visual presentation: Several participants suggested improving the visual presentation with animations (E1P110) or high-quality images (E1P144). We used stock videos/photos from Envato and Pexels for the visual elements, but some scene compositions were less polished than others in the video (E4P77). E1P111 suggested using a more casual and personable presentation manner to make the video more relatable and engaging. Further, regarding content accessibility and retention, E3P72 recommended providing a text summary of key points at the end of the video for easier reference. E2P84 suggested adding descriptive words during the narrative to emphasize key points could guide viewers' attention to important aspects of the video.

Further in-depth content: Several participants recommended creating videos with more actionable advice, particularly in areas like addressing children's antagonistic or hostile attitude when

[3]E5P92: Participant 92 of Episode 5

establishing rules on internet use (E2P81), as well as easier-to-follow tutorials (e.g., "digital parenting for dummies", E4P91). Participants also preferred videos that address specific concerns, such as the dangers of online predators (E2P69), how to respond when "seeing something bad" (E2P99), and children's interaction with specific apps (E1P91), in dedicated videos. Overall, they called for videos that not only offer general advice but also provide in-depth, actionable, and application-specific guidance.

Broader target audience: The intervention program aimed to support parents in having security and privacy conversations within a family setting. Many participants perceived the program as successful, with E2P103 noting that "it did a great job of showing it can be easy and fun to get involved with my kids and what they are doing online." Some participants reported that they were already applying what they had learned from the videos and sharing these educative materials (E5P68). However, some parents suggested the creation of videos specifically designed for children (E1P100), as well as videos that parents and children could watch together. Such resources could help bridge the knowledge gap between parents and children (E5P83).

When it comes to *topics for future episodes*, most participants considered the six episodes covered well the topics that they were interested in (E6P26, E6P46, E6P61). Meanwhile, participants expressed that the videos should be an evolving program, as "online landscape is always changing" (E6P36) and "threats keep emerging time to time" (E6P48). Participants, especially, expected more GAI-related content (E6P66, E6P30) and videos about the consequences of neglecting security and privacy (E6P67). Further, participants wanted to receive suggestions on where to find relevant resources (E6P65). Lastly, participants called for tutorial videos, showing them the parental controls available on a phone (E6P29), how to adjust settings on various digital devices (E6P68), and how to break screen time habits for both parents and children (E6P17).

Key results. Parents found the short videos"great," "informative," and of "high quality," and expressed their high likelihood of recommending the video series to other parents. Meanwhile, they suggested a more engaging narrative, better visual presentation, further in-depth content, and a broader target audience for future series. The video intervention should be an evolving program tailored to parents' specific needs.

5.6 Discussion

5.6.1 Anticipating user needs to inform intervention design

The formative inquiry laid a solid foundation for designing our intervention. A key challenge in designing the intervention was to ensure the content remained relevant to target users. Informed by User Experience (*UX*) scholars [381, 230], we used open-ended questions to collect parents' needs (difficulties in teaching S&P topics) and their anticipation of the intervention program (topics that they want to learn). This step informed the thematic topics that guided our intervention goals for each episode. Although most of these topics have already been highlighted as important in recent S&P literature [10, 410, 172], without empirical inquiry, we cannot match them to our target parents (of children in Grades 6–9). It is important to note that one limitation of this formative inquiry is that parents could not anticipate the S&P risks they were unaware of or not engaging with (e.g., in the Character.ai case [284], parents could not predict the harm of conversational agents they were not familiar with). In addition to the qualitative survey, other methods (e.g., in-depth interviews [182] and focus groups [70]) can also serve as formative inquiry methods.

Although the video series received positive feedback from parents, practitioners can apply various strategies to further engage the audience. We primarily used the following strategies to create our videos: translating scientific knowledge into relatable language, explaining S&P topics with news stories, matching image cues with narratives, and presenting key points in a structured way (similar to academic presentations). The video series was produced with a singular narrative, aligning speech with online stock visuals. Additional visual design and communication techniques could be adopted in future video interventions to enhance audience engagement (see Chapter 5.5.2). Additionally, when combined with storytelling techniques [190, 392], short videos can serve as an effective format for teaching S&P topics.

5.6.2 Short video as a scalable and flexible intervention approach

Short videos can serve as an effective intervention approach for addressing S&P topics within the family environment. Both quantitative and qualitative data from the current study supported the effectiveness of the video format, with participants rating all intervention videos as "very good" in terms of clarity and usefulness. Further, the videos increased parental awareness, parenting self-efficacy and two of the targeted conversation approaches in relation to security and privacy issues. There are several reasons why short videos may work as an effective intervention for parents. People learn best when complementary information is presented simultaneously to both the auditory and visual systems [251, 310]. Another key advantage of videos is their asynchronous nature, which provides greater control

over their learning process [274]. Parents can pause videos to take notes, skip less relevant sections, or revisit information as needed. These features may be particularly beneficial for parents with limited time, as they help reduce cognitive load and facilitate information retention. Additionally, short videos are a relatively low-cost intervention, as they can be easily and widely distributed once produced, and recent work [174] suggest to explore it as an innovative format to educate the public of emerging scams.

Some of our targeted outcomes did not achieve statistical significance through the intervention program, including parental mediation strategies (the frequency with which parents employ specific mediation strategies to manage their children's online activities) and contextualizing S&P conversations. The following three factors might cause this. First, our intervention might not have delivered the various mediation strategies as effectively as it did for other planned topics (see Chapter 5.4.1). As suggested by participants, in-depth or step-by-step videos on specific mediation strategies might address this flaw. Second, we postulate a group difference between measurements. Significant results related to attitudes and beliefs (parental awareness/self-efficacy), while non-significant measurements captured the frequency of mediation. Third, although 63 participants reported applying strategies, the extent varied, and the behavioral impact might take an extended time to reach statistical significance. This highlights the value of longitudinal designs and triangulating scale measurements with other data sources.

5.6.3 Adding a control group and evaluating over time makes a difference

Even without intervention, the control group showed a statistically significant increase in general privacy concerns from pre- to post-questionnaires. This unexpected finding has several important implications. First, the act of responding to items related to privacy concerns may have prompted parents to reflect on their attitudes and behaviors and heightened their awareness of privacy-related issues. This phenomenon, known as the mere measurement effect [266], suggests that simply engaging with privacy topics can influence cognitive processes and inspire self-reflection. Second, the rise in privacy concerns suggests that these concerns may evolve with time, for instance, influenced by exposure to news reports on the topic or experiences in the personal environment of the participant. Without a control group, it may be difficult to discern whether observed changes result from the intervention itself or other external factors, such as natural development, stimuli from external events, or participant expectations. By isolating the treatment effect, the control group provides a clearer, more reliable measure of the intervention's true impact [337]. Together, these implications highlight the critical importance of including a control group to differentiate the effects of the intervention from potential influences of self-reflection triggered by measurement and the passage of time.

Lessons learned from a 14-week experiment: *dropout rate, application*, and *those who stayed*. Evaluating S&P interventions presents significant challenges for the research and practitioner community [180]. Lack of longitudinal and in-the-field evaluations limits the ecological validity and generalizability of findings [72, 66]. Further, the design and implementation of longitudinal experiments necessitate careful consideration during interpretation [62]. 16% of the participants from the control group dropped out over the period of 14 weeks, and 5-15% of the participants from the intervention group dropped out at each contact point (excluding the ones declined our study invitation). Future researchers may consider our dropout rate as a point of reference when designing longitudinal experiments. Further, the longitudinal design allows parents to digest and apply what we delivered in the video series, as evidenced by 63 of them indicating they had applied the learned strategies in the post-questionnaire. Additionally, we postulate that the contact points can serve as biweekly reminders of learning, similar to "triggers" in Fogg's behavior change model [126], which prompted them to pay more attention to S&P topics in their daily life. It is noteworthy that, although our intervention received relatively high ratings in the feedback questions (see Table 4.4.4), the rising positive ratings between Episodes 2 and 6 require scrutiny, considering that parents who did not like our interventions might have dropped out.

5.6.4 Collaborative approaches to developing S&P literacy in families

Mitigating online S&P risks in families requires a collaborative approach involving both parents and their children. Our study not only validates the importance of this collaborative approach [281, 6] but also contributes insights into how we can form this collaborative approach in practice. Research on mitigating S&P risks in smart homes [355] emphasizes that transparency between parents and children can facilitate knowledge sharing and open communication, thus supporting joint learning and shared oversight. When parents adopt a more open yet constructive attitude toward their children's online behaviors, they can foster a positive emotional climate at home and have a positive impact on their children's digital habits and decision-making. Further, Alghythee et al. [10] suggested designing privacy literacy interfaces that support the interaction between parents and children in the learning process. This aligns with the suggestion from our study participants to create video series suitable for parents and children to learn together, creating occasions that they can exchange on various S&P topics.

In our study, some parents described tension with their children when addressing S&P topics and found it difficult to initiate conversations. Indeed, a study investigating the experiences of "security adepts" (referring to adults) [137] found that challenges included a lack of interest, a feeling of judging others' behaviors, and fear of being perceived as paranoid. The authors recommended the

use of conversation starters (e.g., media coverage, action days, movies; see also [299]) to break the ice. When parents frequently use rule-based conversations, they may encounter resistance from their children [10]. However, diversifying S&P conversation approaches requires parents to update their digital literacy, actively collect materials, and learn about the applications their children are using. 81% of parents in the formative study expressed varying degrees of difficulty in addressing S&P topics. As most security interventions predominantly focus on student demographics [357], parents would appreciate more structured intervention programs to support the development of their S&P literacy. Our intervention program only dived into three conversation strategies and provided overviews for S&P concepts, and good digital parenting. Many other topics that parents wanted to learn (see Table 5.1) can be included to develop future intervention programs to empower parents.

5.6.5 Creating hassle-free interventions for families

We need an interdisciplinary approach to address the security and privacy challenges posed to the general public. Whenever a new technology is deployed in society, malicious players find ways to exploit these technologies and put people's security and privacy at risk [167]. As technologies have assimilated into families, organizations, and government bodies, S&P literacy has become necessary for the general public. We combined expertise in digital literacy, security and privacy intervention design, and developmental psychology to address this challenge. So far, cybersecurity is still an emerging theme for the media literacy community [116]. This work can draw the attention of the media literacy community to integrate family S&P topics into their research scope. Furthermore, the developmental needs of children require further consideration in prioritizing safety and privacy by design from technology makers [162]. For instance, platforms should take more responsibility for protecting children from harmful content and online predators (see Table 5.1), reducing the burdens parents experience in their everyday lives. Lastly, the development of S&P literacy among the public, in tandem with technological deployment, is imperative for mitigating various risks posed by adversarial players. How to design relevant and hassle-free interventions for everyday technology users remains an important area for further exploration. Our positive feedback highlights that interventions which adopt formats already embedded in individuals' daily lives, are tailored to relevant topics, and impose minimal costs on target users are likely to be accepted by them.

When an intervention is embedded into the existing workflows and routines of target users, it creates less friction. In contrast, some intervention formats, e.g., in-person workshops [72] or online courses [412], require dedicated time slots, pulling users away from their daily routines. A widely adopted, though controversial, example of workplace embedded training is the simulated phishing campaign. However, such an approach is not easily transferable to non-work environments. Moreover, researchers

have criticized it for being less effective than assumed [183] and for inducing negative emotions in recipients [327, 70]. In light of this, low cost refers not only to individual time investment but also to emotional costs [385]. Furthermore, individuals present varying levels of security and privacy literacy; when an intervention does not align with users' skill levels, they may perceive it as having low task value and disengage from it [73]. Another promising direction to make interventions more relevant is to personalize them based on users' differing levels of proficiency [326]. In addition to tailored short videos, researchers can further explore how to utilize existing household devices to create low-effort and emotionally positive interventions [384]. Opportunities such as "learning interfaces" [268, 10] may also act as stimuli to raise S&P awareness in families.

5.6.6 Limitations and future directions

The present study has some limitations. First, while it is one of few longitudinal study designs in human-centered security, our dependent variables were measured at two points of time, two weeks before and after the intervention. This does not allow us to investigate the sustainability of the intervention, which could be addressed in future research, for example, by incorporating follow-up assessments after five months [38]. Second, our collected data relied on self-reports, which could be triangulated with teenager self-reported data, or even observational insights in the future. Conversely, we were interested in parents' concerns, beliefs, and attitudes, which are intrapsychic constructs and can be ideally measured through self-reports [253]. We did not investigate teenagers' views in the present study due to the limitation of data collection through a crowdsourcing platform. While parents might feel more aware and self-efficacious after the intervention, we did not investigate how these changes were perceived by their children. Third, the sample in the main study consisted of parents of schoolchildren in grades 6-9 and had a higher socio-economic status compared to the U.S. average, limiting the generalizability of the results to the broader population of U.S. parents or parents from other regions [174]. Furthermore, we did not include a knowledge quiz or attempt to assess parents' new knowledge. Future work might set more specific learning objectives and assess whether these were reached. Parents in the intervention group may have developed increased curiosity about S&P topics and sought additional resources. Future studies might include it as a possible confounding variable, as some observed outcomes may be attributable to participants' independent learning rather than the intervention content alone.

Last but not least, although the video format offers several advantages for parenting interventions, a potential limitation is its reliance on internet availability and a certain level of digital competency. Short videos have become a popular medium consumed by the general public, offering opportunities to deploy educational content across different platforms. By aligning with individuals' existing media

consumption habits, short videos can reduce cognitive barriers to engagement and potentially foster S&P awareness in a more contextually relevant and accessible manner. However, parents with lower digital skills may be less likely to access or engage with video interventions, even though they may have a greater need for such resources due to a wider gap between their own and their child's digital competencies. Future research could explore which parent groups may be underserved by video interventions and investigate alternative formats (e.g., community-based learning [241]) to ensure these groups are also supported effectively.

5.7 Conclusion

In this study, we aimed to bridge the gap between research and practice by developing an evidence-based intervention program to support parenting in the context of protecting children from online S&P risks. Through a 14-week randomized controlled trial, the evaluation revealed that short videos can be an effective approach to enhance parental awareness, develop parenting self-efficacy, and diversify conversation approaches.

The feedback from our intervention points to promising directions for future research in the family environment to protect children from security and privacy risks. Some key implications include: (a) anticipating target users' needs to inform intervention design, and suggesting various strategies researchers can apply to further engage the audience with short video format interventions; (b) short videos can be a scalable and flexible intervention approach in families, and future in-depth or tutorial videos on specific topics can further support parents; (c) it is important to include a control group to differentiate the effects of the intervention from potential influences of self-reflection and the passage of time, and to interpret longitudinal results with scrutiny; and (d) short videos can be a suitable medium for learning and exchange on S&P topics between parents and children; practitioners and researchers could explore topics not covered in our study in future intervention programs to support parents. As an additional resource associated with this paper, we provide the video series and the validated measurement scales for future research and application.

Data Availability Statement

The preregistration [365], Appendixes, intervention program, anonymized datasets generated in the main study, and R scripts used for analysis are included in the OSF repository: `osf.io/x3ust`.

Acknowledgments

Author 1 gratefully acknowledges the financial support of the Institute for Advanced Studies (IAS) at the University of Luxembourg through a 'Young Academics' grant awarded in 2021. The study was supported by the User Lab of the University of Luxembourg. We thank Christine Schiltz (Cognitive Science and Assessment Institute) and Sascha Helsper (MEDIACentre) for their support in producing the video series. We thank Family Online Safety Institute for allowing us to include *Seven steps to good digital parenting* in our video series. We also thank Marek for his contribution to video design and voiceovers.

Chapter 6

Concluding Remarks

> [A theory] *is rather more like a map (an analogy Kaplan also makes) of a partially explored territory. Its function is often heuristic, that is, to guide the explorer in further discovery. The way theories make a difference in the world is thus not that they answer questions, but that they guide and stimulate intelligent search.*
>
> -Joseph Weizenbaum [396, p.142]

In this chapter, I synthesize the key findings and describe the interdependencies among this dissertation's components in Chapter 6.1. I then discuss this dissertation's contributions in Chapter 6.2, followed by its limitations and implications for future research in Chapter 6.3. Subsequently, I propose directions for advancing the field through theory-informed research in human-centered security and privacy in Chapter 6.4 and conclude with a final remark in Chapter 6.4.

6.1 Synthesis of results

This doctoral dissertation examines approaches that can be employed to improve the design and evaluation of security and privacy interventions. I choose an interdisciplinary approach that combines motivation theories, User Experience, and experiments in real-world settings to investigate (1) how autonomous motivation influences individuals' security behaviors, (2) how motivation theories can be applied to design interventions for specific demographic groups, and (3) how to evaluate the effectiveness of interventions in real-world settings. I will now synthesize the results.

6.1.1 How autonomous motivation influences individuals' security behaviors

The first objective of this dissertation was to examine how autonomous motivation influences individuals' security behaviors (Chapter 2). We conducted a preregistered systematic literature review in organizational security contexts to address this research objective. By analyzing the definitions, measurements (when provided), and references of all autonomous motivators from reviewed papers (n = 45), we identified 17 unique autonomous motivators that related to security behaviors. We found that neither the SDT Motivation Continuum [313] nor the SDT Extra-role Taxonomy [131] could accommodate the 17 autonomous motivators identified. Therefore, we introduced two core constructs — intrinsic (task) value and expectation, from Expectancy-Value Theory [111] — into the SDT Extra-role Taxonomy and integrated psychological needs fulfillment as an additional reason for

behavior. Thus, we proposed a refined taxonomy of autonomous motivation related to organizational security behaviors.

Furthermore, we found three categories of security behaviors/intentions are related to autonomous motivators. The most examined security behaviors in the reviewed studies were ISP compliance behavior/intention (n = 24), followed by extra-role security behaviors (n = 22). Additionally, three studies investigated employees' ISP violation behaviors. *Autonomous motivators were not only positively associated with compliance behaviors and extra-role security behaviors but also negatively related to violation behaviors.* 24 different theoretical frameworks related to autonomous motivators were employed to study these security behaviors. SDT was the most frequently cited (n = 16), followed by Protection Motivation Theory (n = 7), Theory of Planned Behavior (n = 6), and Deterrence Theory (n = 3). While previous studies aiming to explain security behavior understood as compliance often made use of theories focusing on threats and deterrence, our review suggests that there is a shift towards SDT as the most frequently applied theory in studies that focused on what motivates employees to ensure security.

Most of our reviewed studies chose deductive approaches; only five used design approaches, and three were exploratory studies. Authors of reviewed papers reflected on their study limitations and suggested future study opportunities. We categorized these suggestions into the following four themes:

- Testing and refining theoretical frameworks (both established theories and conceptual models) is essential for advancing cybersecurity research.
- To enhance methodological rigor, future research should incorporate longitudinal designs, increase the use of qualitative approaches, improve measurement techniques and sampling strategies, include relevant control variables, and conduct replication studies.
- To mitigate potential biases in findings, it is essential to study underinvestigated sectors and demographics to extend beyond heavily regulated industries and traditional geographic regions.
- Authors called for more *theory-informed interventions* and practical applications. Organizations should create security interventions with more *culturally aligned*, *autonomous*, and *empowerment-focused* approaches.

6.1.2 Guiding the design of interventions with User Experience and motivation theories

The second objective of this dissertation was to explore how motivation theories can be applied to design interventions for specific demographic groups. Through User Experience evaluation, we examine users' interaction with existing interventions and their anticipation of future interventions. Meanwhile, motivation theories—such as Self-Determination Theory and

Expectancy-Value Theory—provide overarching guidelines for designing specific interventions. To achieve this goal, we investigated an organizational context, as well as the family settings.

In the organizational context, we conducted seven focus groups with 34 employees to explore factors influencing employees' engagement with current phishing interventions (see Chapter 3). The focus groups examined general opinions, *confidence*, *goal setting*, and *role identification*, as well as their *perceived costs and benefits* related to phishing interventions. These discussion questions were informed by the core concepts within the EVT framework that affect an individual's choices and performance. Additionally, we asked employees to brainstorm as if they were the new chief information security officer to design engaging interventions to raise phishing awareness. We revealed a spectrum of factors influencing employees' engagement. The perceived value of phishing interventions influences employees' participation. Although the expectation of mitigation and fear of consequences can motivate employees, lack of feedback and communication, worries, and privacy concerns discourage them from reporting phishing emails. We also found that the expectancy–value framework provides a unique lens for explaining how organizational culture, social roles, and the influence of colleagues and supervisors foster proactive responses to phishing attacks. These constructs are not visible in commonly applied theoretical frameworks for security behavior research, such as Protection Motivation Theory and the Theory of Planned Behavior. Further, we documented a range of improvements to phishing interventions proposed by employees.

Based on our empirical findings, we chose SDT as the backbone of our intervention design. Specifically, we aimed to create interventions that promote social interactions among employees to satisfy their need for relatedness and to incorporate elements of interest and enjoyment. Subsequently, we developed two training programs: a *group discussion* and a *role-playing training* (see Chapter 4). The group discussion aimed to promote the exchange of phishing experiences among employees and how they responded to such attacks. In the role-playing condition, we divided employees into two groups and asked them to play the role of hackers aiming to infiltrate the organization. They began with a discussion on suspicious elements and attack techniques used in real phishing emails, followed by collaboratively creating one phishing email to phish the other group.

In the family settings, we utilized a qualitative questionnaire to identify the difficulties and needs of parents (of children in Grade 6-9) face in supporting their children's online S&P (see Chapter 5). We found that parents were concerned about their children's online activities. Many did not consider themselves knowledgeable in S&P topics or lacked confidence in explaining these topics. Additionally, many parents struggled to attract their children's attention and interest when having S&P conversations with them. Lastly, a few parents found it challenging to balance trust and monitoring, instructing and disengagement, and allowing independence while ensuring their child's security.

Further, we surveyed parents on the S&P topics they want to learn about. We summarized reported topics from 208 parents into six categories: (1) Online privacy: topics included managing privacy settings, protecting personal information, understanding digital footprints, and safe sharing practices, including sensitive content. (2) Online security: covered general and specific threats such as hacking, phishing, scams, account protection, and malware (e.g., viruses and ransomware). (3) Parental monitoring and parenting strategies: included parental controls, tracking tools, safe browsing guidelines, how to teach children about threats, and fostering secure digital habits. (4) Online protective actions: included avoiding risky websites, managing social media use, and dealing with issues like cyberbullying, grooming, addiction, and peer pressure. (5) Online harmful content: includes exposure to pornography, racism, misinformation, fake news, drugs, and manipulative online actors (e.g., bots). And (6) Emerging risks: AI-generated content, deepfake, sexting, and the latest applications and technologies.

Considering parents' packed schedules and time constraints, we selected EVT as the guiding framework for our intervention design. We aimed to develop an intervention characterized by high perceived benefits and low cost to optimize the benefit-to-cost ratio. Subsequently, we chose short videos as the medium of intervention. Short videos can deliver a lot of content in a short time, catch viewers' attention, and have both short- and long-term training effects. We focused on category-level and highlighted relevant S&P risks alongside respective coping strategies. Following a structured approach, we created a program consisting of *six short videos* (totaling 29 minutes) with episodes that focus on: general online security and privacy, good digital practices, parental monitoring approaches, risks associated with popular applications, parental communication strategies, and emerging risks.

6.1.3 Evaluating the effectiveness of interventions in real-world settings with field experiments

To evaluate the effectiveness of group discussion and role-playing training, we conducted a pre-registered *mixed-design experiment* (N = 105), incorporating repeated measures at three time points, a control group, and three simulated phishing tests between 20 and 50 days after the training (see Chapter 4). We found that group discussion significantly improved self-reported *anti-phishing self-efficacy* in both the immediate and Day 7 assessments. The role-playing training did not reach a significant improvement in the immediate assessment, but demonstrated significant improvement for the Day 7 assessment. However, these improvements in both trainings did not reach statistical significance when compared to the control group. Further, group discussion and role-playing training significantly enhanced *support-seeking intention* in the immediate and Day 7 assessments. Both interventions worked similarly well when comparing their effects on Day 7 (no statistically significant

difference in effects between the two trainings). However, only role-playing training demonstrated statistical significance when compared to the control group.

Regarding the *reporting* and *nonclicking* response of simulated phishing tests, employees in both trainings reported phishing tests statistically more often than employees in the control condition. We did not find a difference in nonclicking between groups, but clicking numbers might have been too low to detect differences. In terms of qualitative results, both trainings enhanced employees' vigilance towards phishing emails, increased reporting intention, and promoted their intention to interact with colleagues when receiving phishing emails. Both trainings were perceived as highly effective and were highly recommended by the employees. More employees from the role-playing training mentioned that the training was "interesting" or "fun" than those from the group discussion.

To evaluate the effectiveness of the video series, we used a 14-week *longitudinal randomized controlled trial*, which consisted of 201 U.S. parents, with 113 assigned to the control group and 88 to the intervention group (see Chapter 5). We assessed the following measurements in both pre- and post-questionnaires: *parental security awareness*, *Parental concerns about S&P of their children*, *Internet-specific parenting self-efficacy*, *Internet-specific parental mediation*, and *parental conversation approaches on S&P*. We also collected parents' general *security attitudes* and *privacy concerns* as covariates. We found that the video series effectively increased parental awareness and parenting self-efficacy. While it did not significantly impact overall parental mediation or conversation approaches, a closer examination of specific facets revealed that the intervention increased the frequency of consequence-based conversations and sharing decision-making thought processes. Surprisingly, even without intervention, the control group showed a statistically significant increase in general privacy concerns from pre- to post-questionnaires. Furthermore, parents reported the short videos "great," "informative," and of "high quality," and expressed their high likelihood of recommending the video series to other parents. Meanwhile, they suggested a more engaging narrative, better visual presentation, further in-depth content, and a broader target audience for future series. The video intervention should be an evolving program tailored to parents' specific needs.

6.2 Contributions

This dissertation makes several contributions that can be structured across theoretical, practical, and methodological aspects.

6.2.1 Theoretical contributions

First, we developed a refined taxonomy of autonomous motivation related to organizational security behaviors through a systematic review of relevant empirical studies (see Chapter 2). The taxonomy

clarifies a popular, yet inconsistently applied, construct—autonomous motivation—in security behavioral studies. This taxonomy facilitates future research to examine and develop human-centered security policies and interventions.

Second, we contextualized Expectancy-Value Theory within the security domain and demonstrated its relevance as a valuable framework for future exploration with focus groups (see Chapter 3). Our research emphasizes some core constructs that have been largely ignored by studies that examine employees' acceptance of security interventions, i.e., subjective task value, personal development, and expectation. This contribution is relevant to both researchers and practitioners who want to improve users' acceptance of S&P interventions.

Third, we proposed integrating new constructs into the Expectancy-Value framework, specifically adding "organizational culture" and "colleague and supervisor behavior", to better capture social and cultural influences on employees' security behaviors (see Chapter 3). Theories from other fields are useful for examining security behaviors; however, they require contextualization and empirical validation. Our work lays the foundation for future scholars who want to further develop the expectancy-value framework in the security context.

6.2.2 Practical contributions

We highlight the merits of guiding intervention design with established theories. Specifically, we reveal that when an intervention satisfies users' needs for relatedness and personal enjoyment (role-playing as hackers in training), it can induce positive attitudes and improve safe responses. When an intervention adopts formats already embedded in users' daily lives, is tailored to relevant topics, and imposes minimal costs on target users, it will likely be accepted by them. These results emphasize the value of grounding intervention design in well-established theories to enhance user acceptance and engagement.

Additionally, our S&P interventions can be implemented across contexts and demonstrate flexibility. We provided all our workshop materials and protocols for group discussions and role-playing trainings, which are adaptable for deployment in different organizations. Moreover, the short video intervention was designed with scalability and flexibility in mind; for example, the short educational videos can be easily integrated into users' media consumption habits [68]. These interventions can be further improved to deliver S&P training for various demographic groups.

Finally, this dissertation provides empirical evidence of the effectiveness of target interventions in influencing users' S&P attitudes and behaviors. Our results demonstrate that interventions can support users in navigating real-world S&P challenges (e.g., safe responses to suspicious emails), offering actionable recommendations for addressing S&P risks (e.g., communication strategies for

parents on S&P topics). Together, these practical contributions bridge the gap between theoretical knowledge and practical application and offer a foundation for future scalable, contextually relevant S&P intervention design.

6.2.3 Methodological contributions

The field experiments in this dissertation can guide future researchers in conducting real-world evaluations in human-centered security and privacy. We provided an example of combining self-reported and behavioral data to evaluate training efficacy with a mixed-design experiment. We shared our experience of collecting behavioral data related to security, obtaining informed consent prior to in-situ phishing tests, and addressing data protection and ethical considerations when conducting field experiments. Additionally, through a longitudinal randomized controlled trial, we highlight the critical importance of including a control group to differentiate the effects of the intervention from potential influences of self-reflection triggered by measurement and the passage of time. Furthermore, we emphasized that scholars require careful consideration regarding the dropout rate, interpretation of findings, and whether participants applied learned content in the longitudinal field evaluation.

Furthermore, through detailed documentation of our adaptation and validation process, we offer a number of meaningful measurements to the community for future evaluations. For instance, we adapted the *Instrumental Support Seeking* scale to evaluate participants' intention to seek support when receiving suspicious emails (see Chapter 4), and we developed the *Parental conversation approaches on S&P* scale to assess the extent to which a parent applies five conversation approaches when discussing S&P topics with their children. These proposed measurements in this dissertation are useful for future empirical investigations into these S&P constructs.

Last but not least, this dissertation promotes open science practices in human-centered security and privacy. We preregistered our experimental designs, hypotheses, data collection process, and analysis methods to enhance the transparency of our findings. To support reproducibility and facilitate future replication, we made our study protocols, training materials, and measurements publicly available and published all non-sensitive data and analysis scripts. These practices uphold transparency and enable other researchers to build upon our work, develop our interventions, and verify our results. We contribute to the broader movement toward methodological openness and transparency in security and privacy research.

6.3 Limitations and future work

We proposed a refined taxonomy of autonomous motivation related to security behaviors and contextualized the Expectancy-Value framework within the security domain in this dissertation.

However, these theoretical constructions require further empirical validation. Although we carefully aligned our taxonomy and framework with existing empirical investigations, they require step-by-step model testing to establish their validity and generalizability. Additionally, the participants included in this dissertation were limited to individuals residing in Europe and North America. Consequently, the generalizability of our findings to other regions warrants further investigation [174]. The group discussion and role-playing trainings were designed and evaluated within an educational institution in Western Europe. This particular cultural and institutional context may involve distinct factors that shape employees' engagement with, and acceptance of, such training approaches. Therefore, future studies should explore the applicability and effectiveness of these interventions across different organizational settings, including corporate, governmental, and nonprofit sectors.

Furthermore, our field experiments do not exhaustively address all potential risks to study participants. For instance, simulated phishing campaigns, while ecologically valid, may induce stress or anxiety among employees [50, 383]—a concern noted in prior literature. Future research should incorporate mechanisms to assess and mitigate potential psychological harm, such as debriefing protocols or opt-out mechanisms. In addition, future work could benefit from incorporating longitudinal designs to investigate the persistence of attitude and behavior changes of S&P interventions over time. Triangulating physiological or behavioral sensing data (e.g., eye gaze data [152] and body motion [18]) may also provide a more nuanced understanding of how users develop safe responses with S&P interventions in the real world.

Due to the scope and limited duration of this dissertation, we only briefly mentioned the topic of generative AI and its associated risks in the short video series. The wide adoption of generative AI applications also poses new threats to users' security and privacy [238]. For instance, generative AI has enhanced social engineering attacks, rendering some security heuristics obsolete, such as the assumption that suspicious emails typically contain grammatical errors. Further, generative AI tools enable attackers to create attack ecosystems more efficiently, allowing them to create fake platforms that mimic legitimate websites faster than ever, even for attackers lacking coding proficiency. Additionally, using generative AI tools by individuals may lead to accidental leakage of input data, which may contain personal and confidential information [260]. Finally, domain expertise is required to evaluate the output of generative AIs; while users may be able to generate code with prompts, they may lack the necessary knowledge to identify malicious scripts hidden in the output. These emerging risks warrant further investigation to develop effective mitigation strategies to protect users' security and privacy.

6.4 Moving forward with theory-informed research

The deliberate examination, application, and integration of theories formed a thread guiding the development of this dissertation. Through the lens of Self-Determination Theory [73], I consolidated the scattered findings regarding the role of autonomous motivation in organizational security behavior studies and proposed a refined taxonomy for future examination. I also utilized the established expectancy-value framework from educational psychology to explore the factors that influence employees' engagement with phishing interventions [70]. Further, the propositions of Self-Determination Theory informed the design of interventions that fulfill employees' needs for relatedness and intrinsic motivation (interest and joy) [72]. Extending beyond the workplace, I developed interventions that target parents in family settings with propositions from the expectancy-value framework (optimization of benefit and cost ratio). Subsequently, I examined the effectiveness of theory-informed interventions across contexts through field experiments. Compared to the widely used problem-solving approach [279, 388], this dissertation's theory-informed approach underscores the value of grounding human-centered security and privacy research in established theoretical frameworks.

Nevertheless, I do not argue that theory is a prerequisite for all empirical inquiries; rather, its relevance and utility depend on the research objectives and methodological constraints. Empirical investigations into various security and privacy behaviors have produced an expanding body of findings; hence, meta-syntheses become essential, and such syntheses are likely to benefit from theoretical frameworks as conceptual tools. By mapping existing findings onto different behaviors, e.g., compliance and extra-role security behaviors, scholars can discern convergent and divergent patterns across contexts [73]. A structured analysis schema informed by established theories would enable cross-study comparisons [234], illuminating conceptual "new grounds" where new theoretical construction is needed for human-centered security and privacy.

Finally, examining the interaction between individuals' fundamental needs for security and privacy and other psychological dimensions represents an underinvestigated frontier. For instance, through the Self-Determination Theory lens [73], future research can explore how basic psychological needs and personal values interact with characteristics of tasks and organizational norms. Likewise, Maslow's hierarchy of needs model [250] invites inquiry into the interplay between safety needs and higher-order desires such as belonging, esteem, and self-actualization in interacting with technologies. In doing so, S&P researchers engage in dialogue with existing theories and contribute to continuous theoretical development. Systematic investigation of these multi-dimensional interactions will coalesce scattered empirical findings into a cohesive, theory-informed paradigm for building bridges between human-centered security and privacy and other HCI communities.

6.5 Final remarks

We live in a society characterized by ubiquitous computing and pervasive services. Individuals who are not aware of the various security and privacy risks entangled within digital technologies need to be informed of these risks. Thus, interventions remain necessary to advise individuals of the risks associated with technology usage and to empower them to address these risks. The overarching research objective of this dissertation was to improve the design and evaluation of security and privacy interventions. Specifically, we adopted an interdisciplinary approach by conducting user experience evaluations to anticipate users' needs, applying propositions from motivation theories to design interventions, and evaluating their effectiveness through longitudinal field experiments.

The intervention programs developed and evaluated in this dissertation support organizational employees in responding safely to suspicious emails and empower parents to help their children address security and privacy risks in family settings. Our interdisciplinary approach could be replicated and scaled up, supported by the openness and transparency of our study materials and publicly available data. We highlight the value of grounding security and privacy research in established theoretical frameworks and advocate for advancing theory-informed research in human-centered security and privacy.

References

References

[1] Anne Adams and Martina Angela Sasse. 1999. Users are not the enemy. *Communications of the ACM*, 42, 12, 40–46.

[2] Wasyihun Sema Admass, Yirga Yayeh Munaye, and Abebe Abeshu Diro. 2024. Cyber security: state of the art, challenges and future directions. *Cyber Security and Applications*, 2, 100031.

[3] Nazilah Ahmad, Umi Asma'Mokhtar, Wan Fariza Paizi Fauzi, Zulaiha Ali Othman, Yusri Hakim Yeop, and Siti Norul Huda Sheikh Abdullah. 2018. Cyber security situational awareness among parents. In *2018 cyber resilience conference (crc)*. IEEE, 1–3.

[4] Icek Ajzen. 1991. The theory of planned behavior. *Organizational behavior and human decision processes*, 50, 2, 179–211.

[5] Icek Ajzen. 2020. The theory of planned behavior: frequently asked questions. *Human behavior and emerging technologies*, 2, 4, 314–324.

[6] Mamtaj Akter, Amy J Godfrey, Jess Kropczynski, Heather R Lipford, and Pamela J Wisniewski. 2022. From parental control to joint family oversight: can parents and teens manage mobile online safety and privacy as equals? *Proceedings of the ACM on Human-Computer Interaction*, 6, CSCW1, 1–28.

[7] Saad Alahmari, Karen Renaud, and Inah Omoronyia. 2023. Moving beyond cyber security awareness and training to engendering security knowledge sharing. *Information Systems and e-Business Management*, 21, 1, 123–158.

[8] Eirik Albrechtsen and Jan Hovden. 2010. Improving information security awareness and behaviour through dialogue, participation and collective reflection. an intervention study. *Computers & Security*, 29, 4, 432–445.

[9] Ahmed Aleroud and Lina Zhou. 2017. Phishing environments, techniques, and countermeasures: a survey. *Computers & Security*, 68, 160–196.

[10] Kenan Kamel A Alghythee, Adel Hrncic, Karthik Singh, Sumanth Kunisetty, Yaxing Yao, and Nikita Soni. 2024. Towards understanding family privacy and security literacy conversations at home: design implications for privacy literacy interfaces. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–12.

[11] Badr A Alharbi, Usama M Ibrahem, Mahmoud A Moussa, Mona A Alrashidy, and Sameh F Saleh. 2023. Parents' digital skills and their development in the context of the corona pandemic. *Humanities and Social Sciences Communications*, 10, 1, 1–10.

[12] Yasser Alhelaly, Gurpreet Dhillon, and Tiago Oliveira. 2023. When expectation fails and motivation prevails: the mediating role of awareness in bridging the expectancy-capability gap in mobile identity protection. *Computers & Security*, 134, 103470.

[13] Rao Faizan Ali, PDD Dominic, Syed Emad Azhar Ali, Mobashar Rehman, and Abid Sohail. 2021. Information security behavior and information security policy compliance: a systematic literature review for identifying the transformation process from noncompliance to compliance. *Applied Sciences*, 11, 8, 3383.

[14] Suzan Ali, Mounir Elgharabawy, Quentin Duchaussoy, Mohammad Mannan, and Amr Youssef. 2020. Betrayed by the guardian: security and privacy risks of parental control solutions. In *Proceedings of the 36th annual computer security applications conference*, 69–83.

[15] Zainab Alkhalil, Chaminda Hewage, Liqaa Nawaf, and Imtiaz Khan. 2021. Phishing attacks: a recent comprehensive study and a new anatomy. *Frontiers in Computer Science*, 3, 563060.

[16] Khalid Alkhattabi, Ahmed Alshehri, and Chuan Yue. 2020. Security and privacy analysis of android family locator apps. In *Proceedings of the 25th ACM Symposium on Access Control Models and Technologies*, 47–58.

[17] Rawan A Alsharida, Bander Ali Saleh Al-rimy, Mostafa Al-Emran, and Anazida Zainal. 2023. A systematic review of multi perspectives on human cybersecurity behavior. *Technology in society*, 73, 102258.

[18] Florian Alt, Mariam Hassib, and Verena Distler. 2023. Human-centered behavioral and physiological security. In *Proceedings of the 2023 New Security Paradigms Workshop* (NSPW '23). Association for Computing Machinery, , Segovia, Spain, 48–61. ISBN: 9798400716201. DOI: 10.1145/3633500.3633504.

[19] Steven Alter. 2014. Theory of workarounds. *Communications of the Association for Information Systems*, 34, 1, 55. http://aisel.aisnet.org/cais/vol34/iss1/55.

[20] Kholoud Althobaiti, Adam DG Jenkins, and Kami Vaniea. 2021. A case study of phishing incident response in an educational organization. *Proceedings of the ACM on Human-Computer Interaction*, 5, CSCW2, 1–32.

[21] Ahmed Alzahrani, Chris Johnson, and Saad Altamimi. 2018. Information security policy compliance: investigating the role of intrinsic motivation towards policy compliance in the organisation. In *2018 4th international conference on information management (ICIM)*. IEEE, New York, NY, USA, 125–132.

[22] Ahmed Alzahrani and Christopher Johnson. 2019. Ahp-based security decision making: how intention and intrinsic motivation affect policy compliance. *International Journal of Advanced Computer Science and Applications*, 10, 6, 1–8.

[23] Monica Anderson, Michelle Faverio, and Jeffrey Gottfried. 2023. Teens, social media and technology 2023. https://www.pewresearch.org/internet/2023/12/11/teens-social-media-and-technology-2023/. Accessed: 10-27-2024. (2023).

[24] APWG. 2023. Phishing activity trends report. https://apwg.org/trendsreports/.

[25] Salim Awudu and Sotirios Terzis. 2023. Investigating staff information security policy compliance in electronic identity systems: the ghanaian national identity system. In *International Conference for International Association for Development of the Information Society (IADIS): Proceedings of International Conferences e-society and Mobile Learning*. ERIC, USA, 68–75.

[26] Maria Bada, Angela M Sasse, and Jason RC Nurse. 2019. Cyber security awareness campaigns: why do they fail to change behaviour? *arXiv preprint arXiv:1901.02672*.

[27] Sikha Bagui, Debarghya Nandi, Subhash Bagui, and Robert Jamie White. 2021. Machine learning and deep learning for phishing email classification using one-hot encoding. *Journal of Computer Science*, 17, 610–623.

[28] Aurélien Baillon, Jeroen De Bruin, Aysil Emirmahmutoglu, Evelien Van De Veer, and Bram Van Dijk. 2019. Informing, simulating experience, or both: a field experiment on phishing risks. *PloS one*, 14, 12, e0224216.

[29] Albert Bandura and Dale H Schunk. 1981. Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of personality and social psychology*, 41, 3, 586.

[30] Malak Baslyman and Sonia Chiasson. 2016. "smells phishy?": an educational game about online phishing scams. In *2016 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, Toronto, Canada, 1–11.

[31] Diana Baumrind. 1991. The influence of parenting style on adolescent competence and substance use. *The journal of early adolescence*, 11, 1, 56–95.

[32] Piers Bayl-Smith, Ronnie Taib, Kun Yu, and Mark Wiggins. 2022. Response to a phishing attack: persuasion and protection motivation in an organizational context. *Information & Computer Security*, 30, 1, 63–78.

[33] Adam Beautement, Ingolf Becker, Simon Parkin, Kat Krol, and M. Angela Sasse. 2016. Productive security: a scalable methodology for analysing employee security behaviours. In *Proceedings of the Twelfth USENIX Conference on Usable Privacy and Security* (SOUPS '16). USENIX Association, Denver, CO, USA, 253–270. ISBN: 9781931971317.

[34] Adam Beautement, M Angela Sasse, and Mike Wonham. 2008. The compliance budget: managing security behaviour in organisations. In *Proceedings of the 2008 new security paradigms workshop*, 47–58.

[35] Ingolf Becker, Simon Parkin, and M Angela Sasse. 2017. Finding security champions in blends of organisational culture. In *Proceedings of the 2nd European Workshop on Usable Security*. Vol. 11. Internet Society, Paris, France, 124.

[36] Kristian Beckers and Sebastian Pape. 2016. A serious game for eliciting social engineering security requirements. In *2016 IEEE 24th International Requirements Engineering Conference (RE)*. IEEE, Beijing, China, 16–25.

[37] Daniel Bennett and Elisa Mekler. 2024. Beyond intrinsic motivation: the role of autonomous motivation in user experience. *ACM Transactions on Computer-Human Interaction*, 1, 1, 1–44.

[38] Benjamin M. Berens, Mattia Mossano, and Melanie Volkamer. 2024. Taking 5 minutes protects you for 5 months: evaluating an anti-phishing awareness video. *Comput. Secur.*, 137, C, (Apr. 2024). DOI: 10.1016/j.cose.2023.103620.

[39] Benjamin Maximilian Berens, Florian Schaub, Mattia Mossano, and Melanie Volkamer. 2024. Better together: the interplay between a phishing awareness video and a link-centric phishing support tool. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–60.

[40] Better Health Channel. n.d. Teenagers and communication. https://www.betterhealth.vic.gov.au/health/healthyliving/teenagers-and-communication. Accessed: 2025-01-17. (n.d.).

[41] John F. Binning. 2016. Construct. https://www.britannica.com/science/construct. Encyclopedia Britannica. (Feb. 2016).

[42] John M Blythe and Lynne Coventry. 2018. Costly but effective: comparing the factors that influence employee anti-malware behaviours. *Computers in Human Behavior*, 87, 87–97.

[43] John M Blythe, Lynne Coventry, and Linda Little. 2015. Unpacking security policy compliance: the motivators and barriers of employees' security behaviors. In *Eleventh Symposium On Usable Privacy and Security ({SOUPS} 2015)*. Usenix, Berkeley, CA, USA, 103–122.

[44] Susanne Bødker. 2006. When second wave hci meets third wave challenges. In *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles* (NordiCHI '06). Association for Computing Machinery, Oslo, Norway, 1–8. ISBN: 1595933255. DOI: 10.1145/1182475.1182476.

[45] Nele Borgert, Luisa Jansen, Imke Böse, Jennifer Friedauer, M. Angela Sasse, and Malte Elson. 2024. Self-efficacy and security behavior: results from a systematic review of research methods. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (CHI '24). Association for Computing Machinery, Honolulu, HI, USA. ISBN: 9798400703300. DOI: 10.1145/3613904.3642432.

[46] Nele Borgert, Oliver D. Reithmaier, Luisa Jansen, Larina Hillemann, Ian Hussey, and Malte Elson. 2023. Home is where the smart is: development and validation of the cybersecurity self-efficacy in smart homes (cysesh) scale. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (CHI '23). Association for Computing Machinery, Hamburg, Germany. ISBN: 9781450394215. DOI: 10.1145/3544548.3580860.

[47] Susan Branje. 2018. Development of parent–adolescent relationships: conflict interactions as a mechanism of change. *Child development perspectives*, 12, 3, 171–176.

[48] Virginia Braun and Victoria Clarke. 2021. Thematic analysis: a practical guide to understanding and doing. 1. *Thousand Oaks*.

[49] Sharon S Brehm and Jack W Brehm. 2013. *Psychological reactance: A theory of freedom and control*. Academic Press, New York, USA.

[50] Lina Brunken, Annalina Buckmann, Jonas Hielscher, and M Angela Sasse. 2023. "to do this properly, you need more resources": the hidden costs of introducing simulated phishing campaigns. In *32nd USENIX Security Symposium (USENIX Security 23)*. USENIX Association, Anaheim, USA, 4105–4122.

[51] Tom Buchanan, Carina Paine, Adam N Joinson, and Ulf-Dietrich Reips. 2007. Development of measures of online privacy concern and protection for use on the internet. *Journal of the American society for information science and technology*, 58, 2, 157–165.

[52] J Buckley, D Lottridge, JG Murphy, and PM Corballis. 2023. Indicators of employee phishing email behaviours: intuition, elaboration, attention, and email typology. *International Journal of Human-Computer Studies*, 172, 102996.

[53] Burcu Bulgurcu, Hasan Cavusoglu, and Izak Benbasat. 2010. Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. *MIS quarterly*, 34, 3, 523–548.

[54] Jan-Willem Bullee and Marianne Junger. 2020. How effective are social engineering interventions? a meta-analysis. *Information & Computer Security*, 28, 5, 801–830.

[55] Pavlo Burda, Luca Allodi, and Nicola Zannone. 2020. Don't forget the human: a crowdsourced approach to automate response and containment against spear phishing attacks. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, New York, NY, USA, 471–476.

[56] Pavlo Burda, Luca Allodi, and Nicola Zannone. 2024. Cognition in social engineering empirical research: a systematic literature review. *ACM Transactions on Computer-Human Interaction*, 31, 2, 1–55.

[57] AJ Burns, M Eric Johnson, and Deanna D Caputo. 2019. Spear phishing in a barrel: insights from a targeted phishing campaign. *Journal of Organizational Computing and Electronic Commerce*, 29, 1, 24–39.

[58] AJ Burns, Tom L Roberts, Clay Posey, Paul Benjamin Lowry, and Bryan Fuller. 2023. Going beyond deterrence: a middle-range theory of motives and controls for insider computer abuse. *Information Systems Research*, 34, 1, 342–362.

[59] Barbara M Byrne. 2013. *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. routledge, New York.

[60] David C. 2022. Telling users to 'avoid clicking bad links' still isn't working. https://www.ncsc.gov.uk/blog-post/telling-users-to-avoid-clicking-bad-links-still-isnt-working.

[61] Anthony Carella, Murat Kotsoev, and Traian Marius Truta. 2017. Impact of security awareness training on phishing click-through rates. In *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, Boston, MA, USA, 4458–4466.

[62] Edward Joseph Caruana, Marius Roman, Jules Hernández-Sánchez, and Piergiorgio Solli. 2015. Longitudinal studies. *Journal of thoracic disease*, 7, 11, E537.

[63] Jake Chanenson, Brandon Sloane, Navaneeth Rajan, Amy Morril, Jason Chee, Danny Yuxing Huang, and Marshini Chetty. 2023. Uncovering privacy and security challenges in k-12 schools. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (CHI '23). Association for Computing Machinery, Hamburg, Germany. ISBN: 9781450394215. DOI: 10.1145/3544548.3580777.

[64] Christina Ling-Hsing Chang, Victor Chen, Gary Klein, and James J Jiang. 2011. Information system personnel career anchor changes leading to career changes. *European Journal of Information Systems*, 20, 1, 103–117.

[65] Gary Charness, Uri Gneezy, and Michael A Kuhn. 2012. Experimental methods: between-subject and within-subject design. *Journal of economic behavior & organization*, 81, 1, 1–8.

[66] Sunil Chaudhary, Vasileios Gkioulos, and Sokratis Katsikas. 2022. Developing metrics to assess the effectiveness of cybersecurity awareness program. *Journal of Cybersecurity*, 8, 1, tyac006.

[67] Hao Chen and Wenli Li. 2019. Understanding commitment and apathy in is security extra-role behavior from a person-organization fit perspective. *Behaviour & Information Technology*, 38, 5, 454–468.

[68] Xiaowei Chen, Verena Distler, Chloe Gordon, Yaxing Yao, and Ziwen Teuber. 2025. Empowering parents to support children's online security and privacy: findings from a randomized controlled trial. In *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security (CCS '25)*. ACM, Taipei, Taiwan, (Oct. 2025), 1–15. DOI: 10.1145/3719027.3765214.

[69] Xiaowei Chen, Sophie Doublet, and Verena Distler. 2024. Making motivation theories accessible: introducing motivation cards to map motivators for security and privacy education. In *S&PEI Workshop of the Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*.

[70] Xiaowei Chen, Sophie Doublet, Anastasia Sergeeva, Gabriele Lenzini, Vincent Koenig, and Verena Distler. 2024. What motivates and discourages employees in phishing interventions: an exploration of {expectancy-value} theory. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*, 487–506.

[71] Xiaowei Chen, Anders Hedman, Verena Distler, and Vincent Koenig. 2023. Do persuasive designs make smartphones more addictive?-a mixed-methods study on chinese university students. *Computers in Human Behavior Reports*, 10, 100299.

[72] Xiaowei Chen, Margault Sacré, Gabriele Lenzini, Samuel Greiff, Verena Distler, and Anastasia Sergeeva. 2024. The effects of group discussion and role-playing training on self-efficacy, support-seeking, and reporting phishing emails: evidence from a mixed-design experiment. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (CHI '24). Association for Computing Machinery, Honolulu, HI, USA. ISBN: 9798400703300. DOI: 10.1145/3613904.3641943.

[73] Xiaowei Chen, Lorin Schöni, Verena Distler, and Verena Zimmermann. 2025. Beyond deterrence: a systematic review of the role of autonomous motivation in organizational security behavior studies. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, Yokohama. ISBN: 9798400713941. DOI: 10.1145/3706598.3713122.

[74] Xiaowei Chen, Verena Zimmermann, Lorin Schöni, and Verena Distler. 2024. Systematic literature review on autonomous motivation in organizational cybersecurity behaviors. https://osf.io/jxtk9. (Mar. 2024).

[75] Gokul CJ, Sankalp Pandit, Sukanya Vaddepalli, Harshal Tupsamudre, Vijayanand Banahatti, and Sachin Lodha. 2018. Phishy - a serious game to train enterprise users on phishing awareness. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (CHI PLAY '18 Extended Abstracts). Association for Computing Machinery, Melbourne, Australia, 169–181. ISBN: 9781450359689. DOI: 10.1145/3270316.3273042.

[76] Victoria Clarke and Virginia Braun. 2017. Thematic analysis. *The journal of positive psychology*, 12, 3, 297–298.

[77] Dan Conway, Ronnie Taib, Mitch Harris, Shlomo Berkovsky, Kun Yu, and Fang Chen. 2017. A qualitative investigation of bank employee experiences of information security and phishing. In *Proceedings of the Thirteenth USENIX Conference on Usable Privacy and Security* (SOUPS '17). USENIX Association, Santa Clara, CA, USA, 115–129. ISBN: 9781931971393.

[78] David A Cook and Anthony R Artino Jr. 2016. Motivation to learn: an overview of contemporary theories. *Medical education*, 50, 10, 997–1014.

[79] Gregory W Corder and Dale I Foreman. 2011. Nonparametric statistics for non-statisticians. (2011).

[80] W Alec Cram, John D'arcy, and Jeffrey G Proudfoot. 2019. Seeing the forest and the trees: a meta-analysis of the antecedents to information security policy compliance. *MIS quarterly*, 43, 2, 525–554.

[81] Lorrie F Cranor. 2008. A framework for reasoning about the human in the loop.

[82] Lorrie Faith Cranor, Adam L. Durity, Abigail Marsh, and Blase Ur. 2014. Parents' and Teens' Perspectives on Privacy In a Technology-Filled World. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*. USENIX Association, Menlo Park, CA, (July 2014), 19–35. ISBN: 978-1-931971-13-3. https://www.usenix.org/conference/soups2014/proceedings/presentation/cranor.

[83] Russell Cropanzano and Marie S Mitchell. 2005. Social exchange theory: an interdisciplinary review. *Journal of management*, 31, 6, 874–900.

[84] Robert E Crossler, Allen C Johnston, Paul Benjamin Lowry, Qing Hu, Merrill Warkentin, and Richard Baskerville. 2013. Future directions for behavioral information security research. *computers & security*, 32, 90–101.

[85] John D'Arcy, Anat Hovav, and Dennis Galletta. 2009. User awareness of security countermeasures and its impact on information systems misuse: a deterrence approach. *Information systems research*, 20, 1, 79–98.

[86] Joseph Da Silva and Rikke Bjerg Jensen. 2022. " cyber security is a dark art": the ciso as soothsayer. *Proceedings of the ACM on Human-Computer Interaction*, 6, CSCW2, 1–31.

[87] Adele Da Veiga, Liudmila V Astakhova, Adéle Botha, and Marlien Herselman. 2020. Defining organisational information security culture—perspectives from academia and industry. *Computers & Security*, 92, 101713.

[88] Laila Dahabiyeh. 2021. Factors affecting organizational adoption and acceptance of computer-based security awareness training tools. *Information & Computer Security*, 29, 5, 836–849.

[89] Duy Dang-Pham, Karlheinz Kautz, Ai-Phuong Hoang, and Siddhi Pittayachawan. 2022. Identifying information security opinion leaders in organizations: insights from the theory of social power bases and social network analysis. *Computers & Security*, 112, 102505.

[90] Sanchari Das, Christena Nippert-Eng, and L Jean Camp. 2022. Evaluating user susceptibility to phishing attacks. *Information & Computer Security*, 30, 1, 1–18.

[91] Sauvik Das, Laura A. Dabbish, and Jason I. Hong. 2019. A typology of perceived triggers for end-user security and privacy behaviors. In *Proceedings of the Fifteenth USENIX Conference on Usable Privacy and Security* (SOUPS '19). USENIX Association, Santa Clara, CA, USA, 97–115. ISBN: 9781939133052.

[92] Joshua M Davis, Deepti Agrawal, and Xiang Guo. 2023. Enhancing users' security engagement through cultivating commitment: the role of psychological needs fulfilment. *European Journal of Information Systems*, 32, 2, 195–206.

[93] Joshua M Davis, Deepti Agrawal, and Obi Ogbanufe. 2025. Shaping extra-role security behaviors through employee-agent relations: a dual-channel motivational perspective. *International Journal of Information Management*, 80, 102833.

[94] Marco De Bona and Federica Paci. 2020. A real world study on employees' susceptibility to phishing attacks. In *Proceedings of the 15th International Conference on Availability, Reliability and Security* (ARES '20). Association for Computing Machinery, Virtual Event, Ireland. ISBN: 9781450388337. DOI: 10.1145/3407023.3409179.

[95] Edward L Deci and Richard M Ryan. 2000. The" what" and" why" of goal pursuits: human needs and the self-determination of behavior. *Psychological inquiry*, 11, 4, 227–268.

[96] Edward L Deci and Richard M Ryan. 2008. Facilitating optimal motivation and psychological well-being across life's domains. *Canadian psychology/Psychologie canadienne*, 49, 1, 14.

[97] Edward L Deci and Richard M Ryan. 2008. Self-determination theory: a macrotheory of human motivation, development, and health. *Canadian psychology/Psychologie canadienne*, 49, 3, 182.

[98] Edward L Deci and Richard M Ryan. 2014. The importance of universal psychological needs for understanding motivation in the workplace. *The Oxford handbook of work engagement, motivation, and self-determination theory*, 13, 13–32.

[99] Gurpreet Dhillon, Yurita Yakimini Abdul Talib, and Winnie Ng Picoto. 2020. The mediating role of psychological empowerment in information security compliance intentions. *Journal of the Association for Information Systems*, 21, 1, 5.

[100] Antonio Díaz Andrade, Monideepa Tarafdar, Robert M Davison, Andrew Hardin, Angsana A Techatassanasoontorn, Paul Benjamin Lowry, Sutirtha Chatterjee, and Gerhard Schwabe. 2023. The importance of theory at the information systems journal. *Information Systems Journal*, 33, 693–702.

[101] Verena Distler. 2021. *The Experience of Security in Human-Computer Interactions: Understanding Security Perceptions Through the Concept of User Experience*. Doctoral Dissertation. University of Luxembourg, Luxembourg.

[102] Verena Distler. 2023. The influence of context on response to spear-phishing attacks: an in-situ deception study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (CHI '23). Association for Computing Machinery, Hamburg, Germany. ISBN: 9781450394215. DOI: 10.1145/3544548.3581170.

[103] Verena Distler, Matthias Fassl, Hana Habib, Katharina Krombholz, Gabriele Lenzini, Carine Lallemand, Lorrie Faith Cranor, and Vincent Koenig. 2021. A systematic literature review of empirical methods and risk representation in usable privacy and security research. *ACM Transactions on Computer-Human Interaction*, 28, 6, 1–50.

[104] Verena Distler, Matthias Fassl, Hana Habib, Katharina Krombholz, Gabriele Lenzini, Carine Lallemand, Vincent Koenig, and Lorrie Faith Cranor. 2023. Empirical research methods in usable privacy and security. In *Human Factors in Privacy Research*. Springer International Publishing Cham, 29–53.

[105] Verena Distler, Gabriele Lenzini, Carine Lallemand, and Vincent Koenig. 2020. The framework of security-enhancing friction: how ux can help users behave more securely. In *Proceedings of the New Security Paradigms Workshop 2020*, 45–58.

[106] Ronald C Dodge Jr, Curtis Carver, and Aaron J Ferguson. 2007. Phishing for user security awareness. *computers & security*, 26, 1, 73–80.

[107] P Drogkaris and A Bourka. 2019. Cybersecurity culture guidelines: behavioural aspects of cybersecurity. *European Union Agency for Network and Information Security (ENISA)*.

[108] Jacquelynne S Eccles. 1983. Expectancies, values and academic behaviors. *Achievement and achievement motives*.

[109] Jacquelynne S Eccles and Allan Wigfield. 1995. In the mind of the actor: the structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and social psychology bulletin*, 21, 3, 215–225.

[110] Jacquelynne S Eccles and Allan Wigfield. 2002. Motivational beliefs, values, and goals. *Annual review of psychology*, 53, 1, 109–132.

[111] Jacquelynne S Eccles and Allan Wigfield. 2020. From expectancy-value theory to situated expectancy-value theory: a developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary educational psychology*, 61, 101859.

[112] Jacquelynne S Eccles and Allan Wigfield. 2024. The development, testing, and refinement of eccles, wigfield, and colleagues' situated expectancy-value model of achievement performance and choice. *Educational Psychology Review*, 36, 2, 1–29.

[113] eSafetyCommissioner. 2019. Parenting in the digital age. Research Report. (2019). `https://www.esafety.gov.au/research/parenting-digital-age`.

[114] Heike Eschenbeck et al. 2019. School-based mental health promotion in children and adolescents with stressos using online or face-to-face interventions: study protocol for a randomized controlled trial within the prohead consortium. *Trials*, 20, 1–12.

[115] Jose Esteves, Elisabete Ramalho, and Guillermo De Haro. 2017. To improve cybersecurity, think like a hacker. *MIT Sloan Management Review*, 53, 3, 71–77.

[116] Francisco Javier Rocha Estrada, Carlos Enrique George-Reyes, and Leonardo David Glasserman-Morales. 2022. Security as an emerging dimension of digital literacy for education: a systematic literature review. *Journal of E-Learning and Knowledge Society*, 18, 2, 22–33.

[117] Cori Faklaris, Laura A Dabbish, and Jason I Hong. 2019. A self-report measure of end-user security attitudes (SA-6). In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX, Santa Clara, USA, 61–77.

[118] Rubia Fatima, Affan Yasin, Lin Liu, and Jianmin Wang. 2019. How persuasive is a phishing email? a phishing game for phishing awareness. *Journal of Computer Security*, 27, 6, 581–612.

[119] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39, 2, 175–191.

[120] FBI. 2023. Business email compromise: the $50 billion scam. https://www.ic3.gov/Media/Y2023/PSA230609.

[121] FBI. 2023. Internet crime report 2022. https://www.ic3.gov/Media/PDF/AnnualReport/2022_IC3Report.pdf. Accessed: 10-02-2024. (2023).

[122] Mike Fenton. 2016. Restoring executive confidence: red team operations. *Network security*, 2016, 11, 5–7.

[123] Ana Ferreira, Lynne Coventry, and Gabriele Lenzini. 2015. Principles of persuasion in social engineering and their use in phishing. In *Human Aspects of Information Security, Privacy, and Trust: Third International Conference, HAS 2015, Held as Part of HCI International 2015, August 2-7, 2015. Proceedings 3*. Springer, Los Angeles, CA, USA, 36–47.

[124] Ana Ferreira and Gabriele Lenzini. 2015. An analysis of social engineering principles in effective phishing. In *2015 Workshop on Socio-Technical Aspects in Security and Trust*. IEEE, Verona, Italy, 9–16.

[125] Andy Field. 2013. *Discovering statistics using IBM SPSS statistics*. sage, Los Angeles.

[126] Brian J Fogg. 2009. A behavior model for persuasive design. In *Proceedings of the 4th international Conference on Persuasive Technology*. Association for Computing Machinery, New York, NY, USA, 1–7.

[127] Paul Formosa, Michael Wilson, and Deborah Richards. 2021. A principlist framework for cybersecurity ethics. *Computers & Security*, 109, 102382.

[128] Forschung and Lehre. 2023. Hochschulen im visier von cyberkriminellen. Accessed: 10-02-2024. (2023). https://www.forschung-und-lehre.de/management/hochschulen-im-visier-von-cyberkriminellen-5541.

[129] Muriel Frank and Clara Ament. 2021. How motivation shapes the sharing of information security incident experience. In *Proceedings of the 54th Hawaii International Conference on System Sciences*. ScholarSpace, Hawaii, USA, 4528–4537.

[130] Muriel Frank, Lennart Jaeger, and Lukas Manuel Ranft. 2022. Contextual drivers of employees' phishing susceptibility: insights from a field study. *Decision Support Systems*, 160, 113818.

[131] Muriel Frank and Vanessa Kohn. 2023. Understanding extra-role security behaviors: an integration of self-determination theory and construal level theory. *Computers & Security*, 132, 103386.

[132] Anjuli Franz. 2022. Why do employees report cyber threats? comparing utilitarian and hedonic motivations to use incident reporting tools. In *ICIS 2022 Proceedings*, 1–13.

[133] Anjuli Franz, Verena Zimmermann, Gregor Albrecht, Katrin Hartwig, Christian Reuter, Alexander Benlian, and Joachim Vogt. 2021. Sok: still plenty of phish in the sea—a taxonomy of user-oriented phishing interventions and avenues for future research. In *Proceedings of the Seventeenth USENIX Conference on Usable Privacy and Security* (SOUPS'21). USENIX Association, USA. ISBN: 978-1-939133-25-0.

[134] Damjan Fujs, Anže Mihelič, and Simon LR Vrhovec. 2019. The power of interpretation: qualitative methods in cybersecurity research. In *Proceedings of the 14th International Conference on Availability, Reliability and Security*, 1–10.

[135] Trevor Gabriel and Steven Furnell. 2011. Selecting security champions. *Computer Fraud & Security*, 2011, 8, 8–12.

[136] Yotamu Gangire, Adéle Da Veiga, and Marlien Herselman. 2021. Assessing information security behaviour: a self-determination theory perspective. *Information & Computer Security*, 29, 4, 625–646.

[137] Nina Gerber and Karola Marky. 2022. The nerd factor: the potential of S&P adepts to serve as a social resource in the user's quest for more secure and Privacy-Preserving behavior. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. USENIX Association, Boston, MA, (Aug. 2022), 57–76. ISBN: 978-1-939133-30-4. https://www.usenix.org/conference/soups2022/presentation/gerber.

[138] Cornelia Gerdenitsch, Daniela Wurhofer, and Manfred Tscheligi. 2023. Working conditions and cybersecurity: time pressure, autonomy and threat appraisal shaping employees' security behavior. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 17, 4, 19.

[139] Michaela Geržičáková, Lenka Dedkova, and Vojtěch Mỳlek. 2023. What do parents know about children's risky online experiences? the role of parental mediation strategies. *Computers in Human Behavior*, 141, 107626.

[140] Adam Kavon Ghazi-Tehrani and Henry N Pontell. 2021. Phishing evolves: analyzing the enduring cybercrime. *Victims & Offenders*, 16, 3, 316–342.

[141] Arup Kumar Ghosh, Karla Badillo-Urquiola, Shion Guha, Joseph J LaViola Jr, and Pamela J Wisniewski. 2018. Safety vs. surveillance: what children have to say about mobile apps for parental control. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, 1–14.

[142] Dennis A Gioia, Kevin G Corley, and Aimee L Hamilton. 2013. Seeking qualitative rigor in inductive research: notes on the gioia methodology. *Organizational research methods*, 16, 1, 15–31.

[143] Terese Glatz, Elizabeth Crowe, and Christy M Buchanan. 2018. Internet-specific parental self-efficacy: developmental differences and links to internet-specific mediation. *Computers in human behavior*, 84, 8–17.

[144] Diksha Goel and Ankit Kumar Jain. 2018. Mobile phishing attacks and defence mechanisms: state of art and open research challenges. *computers & security*, 73, 519–544.

[145] Shakthidhar Gopavaram, Jayati Dev, Marthie Grobler, DongInn Kim, Sanchari Das, and L Jean Camp. 2021. Cross-national study on phishing resilience. In *Proceedings of the Workshop on Usable Security and Privacy (USEC)*. Internet Society, Auckland, New Zealand, 1–11.

[146] Christa L Green, Joan MT Walker, Kathleen V Hoover-Dempsey, and Howard M Sandler. 2007. Parents' motivations for involvement in children's education: an empirical test of a theoretical model of parental involvement. *Journal of educational psychology*, 99, 3, 532.

[147] Esther Greenglass, Ralf Schwarzer, Dagmara Jakubiec, Lisa Fiksenbaum, and Steffen Taubert. 1999. The proactive coping inventory (pci): a multidimensional research instrument. In *20th international conference of the stress and anxiety research society (STAR)*. Vol. 12. FPUW, Cracow, Poland, 14.

[148] Frank L Greitzer, Wanru Li, Kathryn B Laskey, James Lee, and Justin Purl. 2021. Experimental investigation of technical and human factors related to phishing susceptibility. *ACM Transactions on Social Computing*, 4, 2, 1–48.

[149] Shayl F Griffith and Wendy S Grolnick. 2014. Parenting in caribbean families: a look at parental control, structure, and autonomy support. *Journal of Black Psychology*, 40, 2, 166–190.

[150] Wendy S Grolnick. 2016. Parental involvement and children's academic motivation and achievement. In *Building autonomous learners: Perspectives from research and practice using self-determination theory*. Springer, 169–183.

[151] Moritz Gruber, Christian Höfig, Maximilian Golla, Tobias Urban, and Matteo Große-Kampmann. 2022. “we may share the number of diaper changes”: a privacy and security analysis of mobile child care applications. *Proceedings on Privacy Enhancing Technologies*.

[152] Zhiwei Guan, Shirley Lee, Elisabeth Cuddihy, and Judith Ramey. 2006. The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ’06). Association for Computing Machinery, Montréal, Québec, Canada, 1253–1262. ISBN: 1595933727. DOI: 10.1145/1124772.1124961.

[153] David E Guest and Neil Conway. 2002. Communicating the psychological contract: an employer perspective. *Human resource management journal*, 12, 2, 22–38.

[154] Erik Urdal Gundersen. 2022. *Self-efficacy in organizations cybersecurity training*. Master’s thesis. University of Agder.

[155] Ken H Guo, Yufei Yuan, Norman P Archer, and Catherine E Connelly. 2011. Understanding nonmalicious security violations in the workplace: a composite behavior model. *Journal of management information systems*, 28, 2, 203–236.

[156] Marco Gutfleisch, Markus Schöps, Stefan Albert Horstmann, Daniel Wichmann, and M. Angela Sasse. 2023. Security champions without support: results from a case study with owasp samm in a large-scale e-commerce enterprise. In *Proceedings of the 2023 European Symposium on Usable Security* (EuroUSEC ’23). Association for Computing Machinery, Copenhagen, Denmark, 260–276. ISBN: 9798400708145. DOI: 10.1145/3617072.3617115.

[157] Marco Gutfleisch, Markus Schöps, Sibel Sayin, Frederic Wende, and Martina Angela Sasse. 2022. Putting security on the table: the digitalisation of security tabletop games and its challenging aftertaste. In *Proceedings of the ACM/IEEE 44th International Conference on Software Engineering: Software Engineering Education and Training* (ICSE-SEET ’22). Association for Computing Machinery, Pittsburgh, Pennsylvania, 217–222. ISBN: 9781450392259. DOI: 10.1145/3510456.3514139.

[158] Steffi Haag, Mikko Siponen, and Fufan Liu. 2021. Protection motivation theory in information systems security research: a review of the past and a road map for the future. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 52, 2, 25–67.

[159] JR Hackman. 1976. Motivation through the design work: test of the theory. *Organizational Behavior and Human Performance*, 16, 250–279.

[160] Neal R Haddaway, Matthew J Page, Chris C Pritchard, and Luke A McGuinness. 2022. Prisma2020: an r package and shiny app for producing prisma 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis. *Campbell systematic reviews*, 18, 2, e1230.

[161] Felix Haeussinger and Johann Kranz. 2013. Information security awareness: its antecedents and mediating effects on security compliant behavior. In *Thirty Fourth International Conference on Information Systems*. Citeseer, Milan, 1–16.

[162] Andrew Hale, Barry Kirwan, and Urban Kjellén. 2007. Safe by design: where are we now? *Safety science*, 45, 1-2, 305–327.

[163] JinYoung Han, Yoo Jung Kim, and Hyungjin Kim. 2017. An integrative model of information security policy compliance with psychological contract: examining a bilateral perspective. *Computers & Security*, 66, 52–65.

[164] Julie M Haney and Wayne G Lutters. 2019. Motivating cybersecurity advocates: implications for recruitment and retention. In *Proceedings of the 2019 on Computers and People Research Conference*. Association for Computing Machinery, New York, NY, USA, 109–117.

[165] Stephen Hart, Andrea Margheri, Federica Paci, and Vladimiro Sassone. 2020. Riskio: a serious game for cyber security awareness and education. *Computers & Security*, 95, 101827.

[166] Marc Hassenzahl and Noam Tractinsky. 2006. User experience-a research agenda. *Behaviour & information technology*, 25, 2, 91–97.

[167] Joseph M Hatfield. 2018. Social engineering in cybersecurity: the evolution of a concept. *Computers & Security*, 73, 102–113.

[168] John Hattie, Flaviu A Hodis, and Sean HK Kang. 2020. Theories of motivation: integration and ways forward. *Contemporary Educational Psychology*, 61, 101865.

[169] Cristyne Hébert, Kurt Thumlert, and Jennifer Jenson. 2022. # Digital parents: intergenerational learning through a digital literacy workshop. *Journal of Research on Technology in Education*, 54, 1, 34–91.

[170] Tejaswini Herath and H Raghav Rao. 2009. Encouraging information security behaviors in organizations: role of penalties, pressures and perceived effectiveness. *Decision support systems*, 47, 2, 154–165.

[171] Tejaswini Herath and H Raghav Rao. 2009. Protection motivation and deterrence: a framework for security policy compliance in organisations. *European Journal of information systems*, 18, 106–125.

[172] Franziska Herbert, Steffen Becker, Annalina Buckmann, Marvin Kowalewski, Jonas Hielscher, Yasemin Acar, Markus Dürmuth, Yixin Zou, and M Angela Sasse. 2024. Digital security—a question of perspective a large-scale telephone survey with four at-risk user groups. In *2024 IEEE Symposium on Security and Privacy (SP)*. IEEE, 697–716.

[173] Franziska Herbert, Steffen Becker, Leonie Schaewitz, Jonas Hielscher, Marvin Kowalewski, Angela Sasse, Yasemin Acar, and Markus Dürmuth. 2023. A world full of privacy and security (mis) conceptions? findings of a representative survey in 12 countries. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–23.

[174] Franziska Herbert, Collins W Munyendo, Jonas Hielscher, Steffen Becker, and Yixin Zou. 2025. Digital security perceptions and practices around the world: a weird versus non-weird comparison. In *USENIX Security*. USENIX.

[175] Cormac Herley. 2009. So long, and no thanks for the externalities: the rational rejection of security advice by users. In *Proceedings of the 2009 Workshop on New Security Paradigms Workshop* (NSPW '09). Association for Computing Machinery, Oxford, United Kingdom, 133–144. ISBN: 9781605588452. DOI: 10.1145/1719030.1719050.

[176] Jonas Hielscher, Annette Kluge, Uta Menges, and M Angela Sasse. 2021. "taking out the trash": why security behavior change requires intentional forgetting. In *Proceedings of the 2021 New Security Paradigms Workshop*. Association for Computing Machinery, 108–122.

[177] Jonas Hielscher, Uta Menges, Simon Parkin, Annette Kluge, and M Angela Sasse. 2023. {"employees} who {don't} accept the time security takes are not aware {enough"}: the {ciso} view of {human-centred} security. In *32nd USENIX Security Symposium (USENIX Security 23)*. Anaheim, USA, 2311–2328.

[178] Jonas Hielscher and Simon Parkin. 2024. " what keeps people secure is that they met the security team": deconstructing drivers and goals of organizational security awareness. In *33nd USENIX Security Symposium (USENIX Security 23)*. USENIX, Philadephia, USA, 3295–3312.

[179] Jonas Hielscher, Markus Schöps, Uta Menges, Marco Gutfleisch, Mirko Helbling, and M. Angela Sasse. 2023. Lacking the tools and support to fix friction: results from an interview study with security managers. In *Proceedings of the Nineteenth USENIX Conference on Usable Privacy and Security* (SOUPS '23). USENIX Association, Anaheim, CA, USA. ISBN: 978-1-939133-36-6.

[180] Jonas Hielscher, Markus Schöps, Jens Opdenbusch, Felix Reichmann, Marco Gutfleisch, Karola Marky, and Simon Parkin. 2024. Selling satisfaction: a qualitative analysis of cybersecurity awareness vendors' promises. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, 2666–2680.

[181] Doron Hillman, Yaniv Harel, and Eran Toch. 2023. Evaluating organizational phishing awareness training on an enterprise scale. *Computers & Security*, 132, 103364.

[182] Myat Pan Hmone, Michael J Dibley, Mu Li, and Ashraful Alam. 2016. A formative study to inform mhealth based randomized controlled trial intervention to promote exclusive breastfeeding practices in myanmar: incorporating qualitative study findings. *BMC Medical Informatics and Decision Making*, 16, 1–10.

[183] Grant Ho et al. 2025. Understanding the efficacy of phishing training in practice. In *2025 IEEE Symposium on Security and Privacy (SP)*. IEEE, 37–54.

[184] Allegra Hobbs. 2021. The colonial pipeline hack: exposing vulnerabilities in us cybersecurity. In *SAGE Business Cases*. SAGE Publications: SAGE Business Cases Originals.

[185] Duncan Hodges and Oliver Buckley. 2017. Its not all about the money: self-efficacy and motivation in defensive and offensive cyber security professionals. In *Human Aspects of Information Security, Privacy and Trust: 5th International Conference, HAS 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings 5*. Springer, Berlin/Heidelberg, Germany, 494–506.

[186] Yuxiang Hong and Mengyi Xu. 2021. Autonomous motivation and information security policy compliance: role of job satisfaction, responsibility, and deterrence. *Journal of Organizational and End User Computing (JOEUC)*, 33, 6, 1–17.

[187] Maryam Hosseini, Neda Abdolvand, and Saeedeh Rajaee Harandi. 2022. Two-dimensional analysis of customer behavior in traditional and electronic banking. *Digital Business*, 2, 2, 100030.

[188] Siqi Hu, Carol Hsu, and Zhongyun Zhou. 2022. Security education, training, and awareness programs: literature review. *Journal of Computer Information Systems*, 62, 4, 752–764.

[189] Yue Huang, Marthie Grobler, Lauren S Ferro, Georgia Psaroulis, Sanchari Das, Jing Wei, and Helge Janicke. 2025. Systemization of knowledge (sok): goals, coverage, and evaluation in cybersecurity and privacy games. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–27.

[190] David Michael Hull, Sebastian Walter Schuetz, and Paul Benjamin Lowry. 2023. Tell me a story: the effects that narratives exert on meaningful-engagement outcomes in antiphishing training. *Computers & Security*, 129, 103252.

[191] Princely Ifinedo. 2014. Information systems security policy compliance: an empirical study of the effects of socialisation, influence, and cognition. *Information & Management*, 51, 1, 69–79.

[192] Sitwala Imenda. 2014. Is there a conceptual difference between theoretical and conceptual frameworks? *Journal of social sciences*, 38, 2, 185–195.

[193] Philip G. Inglesant and M. Angela Sasse. 2010. The true cost of unusable password policies: password use in the wild. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10). Association for Computing Machinery, Atlanta, Georgia, USA, 383–392. ISBN: 9781605589299. DOI: 10.1145/1753326.1753384.

[194] Microsoft Threat Intelligence. 2023. Midnight blizzard conducts targeted social engineering over microsoft teams. https://www.microsoft.com/en-us/security/blog/2023/08/02/midnight-blizzard-conducts-targeted-social-engineering-over-microsoft-teams/.

[195] Lennart Jaeger and Andreas Eckhardt. 2021. Eyes wide open: the role of situational information security awareness for security-related behaviour. *Information Systems Journal*, 31, 3, 429–472.

[196] Mohieddin Jafari and Naser Ansari-Pour. 2019. Why, when and how to adjust your p values? *Cell Journal (Yakhteh)*, 20, 4, 604.

[197] Daniel Jampen, Gürkan Gür, Thomas Sutter, and Bernhard Tellenbach. 2020. Don't click: towards an effective anti-phishing training. a comparative literature review. *Human-centric Computing and Information Sciences*, 10, 1, 1–41.

[198] K Jansson and Rossouw von Solms. 2013. Phishing for phishing awareness. *Behaviour & information technology*, 32, 6, 584–593.

[199] Jeffrey L Jenkins, Alexandra Durcikova, Grayson Ross, and Jay F Nunamaker Jr. 2010. Encouraging users to behave securely: examining the influence of technical, managerial, and educational controls on users' secure behavior. In *ICIS 2010 Proceedings. 150.*

[200] Matthew Jensen, Alexandra Durcikova, and Ryan Wright. 2017. Combating phishing attacks: a knowledge management approach. In *Hawaii International Conference on System Sciences (HICSS)*. IEEE, Honolulu, USA, 4288–4297.

[201] Soohyun Jeon and Anat Hovav. 2015. Empowerment or control: reconsidering employee security policy compliance in terms of authorization. In *2015 48th Hawaii International Conference on System Sciences*. IEEE, New York, NY, USA, 3473–3482.

[202] Soohyun Jeon, Anat Hovav, Jinyoung Han, and Steven Alter. 2018. Rethinking the prevailing security paradigm: can user empowerment with traceability reduce the rate of security policy circumvention? *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 49, 3, 54–77.

[203] Soohyun Jeon, Insoo Son, and Jinyoung Han. 2020. Exploring the role of intrinsic motivation in issp compliance: enterprise digital rights management system case. *Information Technology & People*, 34, 2, 599–616.

[204] Soohyun Jeon, Insoo Son, and Jinyoung Han. 2023. Understanding employee's emotional reactions to issp compliance: focus on frustration from security requirements. *Behaviour & Information Technology*, 42, 13, 2093–2110.

[205] Se-Hoon Jeong, Hyunyi Cho, and Yoori Hwang. 2012. Media literacy interventions: a meta-analytic review. *Journal of communication*, 62, 3, 454–472.

[206] William H Jeynes. 2024. A meta-analysis: the association between relational parental involvement and student and parent outcome variables. *Education and Urban Society*, 56, 5, 564–600.

[207] James S Jones. 1987. Participatory teaching methods in computer science. *ACM SIGCSE Bulletin*, 19, 1, 155–160.

[208] Karl G Jöreskog and Dag Sörbom. 1993. *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Scientific software international, Skokie, USA.

[209] Heidi Julien, Jen JL Pecoskie, and Kathleen Reed. 2011. Trends in information behavior research, 1999–2008: a content analysis. *Library & Information Science Research*, 33, 1, 19–24.

[210] Kristian Kannelønning and Sokratis K Katsikas. 2023. A systematic literature review of how cybersecurity-related behavior has been assessed. *Information & Computer Security*, 31, 4, 463–477.

[211] Smirity Kaushik, Yaxing Yao, Pierre Dewitte, and Yang Wang. 2021. "how i know for sure": people's perspectives on solely automated decision-making (SADM). In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, 159–180.

[212] Herbert C Kelman. 1958. Compliance, identification, and internalization three processes of attitude change. *Journal of conflict resolution*, 2, 1, 51–60.

[213] Deborah Kendzierski and Daniel J Whitaker. 1997. The role of self-schema in linking intentions with behavior. *Personality and Social Psychology Bulletin*, 23, 2, 139–147.

[214] Leon Kersten, Pavlo Burda, Luca Allodi, and Nicola Zannone. 2022. Investigating the effect of phishing believability on phishing reporting. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 117–128.

[215] Naurin Farooq Khan, Naveed Ikram, Hajra Murtaza, and Mehwish Javed. 2023. Evaluating protection motivation based cybersecurity awareness training on kirkpatrick's model. *Computers & Security*, 125, 103049.

[216] Iacovos Kirlappos, Simon Parkin, and M Angela Sasse. 2014. Learning from "shadow security": why understanding non-compliance provides the basis for effective security. In *Proceedings of Workshop on Usable Security 2014*.

[217] Rex B Kline. 2011. Principles and practice of structural equation modeling (3. baskı). *New York, NY: Guilford*, 14, 1497–1513.

[218] Johann Kranz and Felix Haeussinger. 2014. Why deterrence is not enough: the role of endogenous motivations on employees' information security behavior. In *Thirty Fifth International Conference on Information Systems*. Association for Information Systems, Atlanta, GA, USA, 1–14.

[219] Udo Kuckartz and Stefan Rädiker. 2019. *Analyzing qualitative data with MAXQDA*. Springer, Switzerland.

[220] Priya C Kumar, Fiona O'Connell, Lucy Li, Virginia L Byrne, Marshini Chetty, Tamara L Clegg, and Jessica Vitak. 2023. Understanding research related to designing for children's privacy and security: a document analysis. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*, 335–354.

[221] Priya C. Kumar, Marshini Chetty, Tamara L. Clegg, and Jessica Vitak. 2019. Privacy and security considerations for digital technology use in elementary schools. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19). Association for Computing Machinery, Glasgow, Scotland Uk, 1–13. ISBN: 9781450359702. DOI: 10.1145/3290605.3300537.

[222] Ponnurangam Kumaraguru, Justin Cranshaw, Alessandro Acquisti, Lorrie Cranor, Jason Hong, Mary Ann Blair, and Theodore Pham. 2009. School of phish: a real-world evaluation of anti-phishing training. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (SOUPS '09). Association for Computing Machinery, Mountain View, California, USA. ISBN: 9781605587363. DOI: 10.1145/1572532.1572536.

[223] Ponnurangam Kumaraguru, Yong Rhee, Steve Sheng, Sharique Hasan, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong. 2007. Getting users to pay attention to anti-phishing education: evaluation of retention and transfer. In *Proceedings of the Anti-Phishing Working Groups 2nd Annual ECrime Researchers Summit* (eCrime '07). Association for Computing Machinery, Pittsburgh, Pennsylvania, USA, 70–81. ISBN: 9781595939395. DOI: 10.1145/1299015.1299022.

[224] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong. 2008. Lessons from a real world evaluation of anti-phishing training. In *2008 eCrime Researchers Summit*. IEEE, Atlanta, USA, 1–12.

[225] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong. 2010. Teaching johnny not to fall for phish. *ACM Transactions on Internet Technology (TOIT)*, 10, 2, 1–31.

[226] Kuang-Ming Kuo, Paul C Talley, and Chi-Hsien Huang. 2020. A meta-analysis of the deterrence theory in security-compliant and security-risk behaviors. *Computers & Security*, 96, 101928.

[227] Youngsun Kwak, Seyoung Lee, Amanda Damiano, and Arun Vishwanath. 2020. Why do users not report spear phishing emails? *Telematics and Informatics*, 48, 101343.

[228] Dominika Kwasnicka, Stephan U Dombrowski, Martin White, and Falko Sniehotta. 2016. Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. *Health psychology review*, 10, 3, 277–296.

[229] Daniele Lain, Kari Kostiainen, and Srdjan Čapkun. 2022. Phishing in organizations: findings from a large-scale and long-term study. In *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Francisco, USA, 842–859.

[230] Carine Lallemand and Guillaume Gronier. 2015. *Méthodes de design UX: 30 méthodes fondamentales pour concevoir et évaluer les systèmes interactifs*. Editions Eyrolles.

[231] Elmer Lastdrager, Inés Carvajal Gallardo, Pieter Hartel, and Marianne Junger. 2017. How Effective is Anti-Phishing Training for Children? In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. USENIX Association, 229–239. ISBN: 978-1-931971-39-3. https://www.usenix.org/conference/soups2017/technical-sessions/presentation/lastdrager.

[232] Richard S Lazarus. 1991. Cognition and motivation in emotion. *American psychologist*, 46, 4, 352.

[233] Benedikt Lebek, Jörg Uffen, Michael H Breitner, Markus Neumann, and Bernd Hohler. 2013. Employees' information security awareness and behavior: a literature review. In *2013 46th Hawaii International Conference on System Sciences*. IEEE, Hawaii, USA, 2978–2987.

[234] Benedikt Lebek, Jörg Uffen, Markus Neumann, Bernd Hohler, and Michael H. Breitner. 2014. Information security awareness and behavior: a theory-based literature review. *Management Research Review*, 37, 12, 1049–1092.

[235] Daeun Lee, Harjinder Singh Lallie, and Nadine Michaelides. 2023. The impact of an employee's psychological contract breach on compliance with information security policies: intrinsic and extrinsic motivation. *Cognition, Technology & Work*, 25, 2, 273–289.

[236] Han Li, Rathindra Sarathy, Jie Zhang, and Xin Luo. 2014. Exploring the effects of organizational justice, personal ethics and sanction on internet use policy compliance. *Information Systems Journal*, 24, 6, 479–502.

[237] Ling Li, Wu He, Li Xu, Ivan Ash, Mohd Anwar, and Xiaohong Yuan. 2019. Investigating the impact of cybersecurity policy awareness on employees' cybersecurity behavior. *International Journal of Information Management*, 45, 13–24.

[238] Megan Li, Wendy Bickersteth, Ningjing Tang, Jason Hong, Lorrie Cranor, Hong Shen, and Hoda Heidari. 2025. A closer look at the existing risks of generative ai: mapping the who, what, and how of real-world incidents. *arXiv preprint arXiv:2505.22073*.

[239] Yaojie Li, Thomas F Stafford, Bryan Fuller, and Selwyn Ellis. 2017. Beyond compliance: empowering employees' extra-role security behaviors in dynamic environments. In *AMCIS*.

[240] Ann-Kristin Lieberknecht. 2024. Exploring determinants of parental engagement in online privacy protection: a qualitative approach. In *Proceedings of the 2024 European Symposium on Usable Security*, 94–111.

[241] Ann-Kristin Lieberknecht and Aline Melanie Ochs. 2024. Safeguarding children's digital privacy: exploring design requirements for effective literacy training for parents. In *IFIP World Conference on Information Security Education*. Springer, 111–126.

[242] Lanjing Liu, Lan Gao, Nikita Soni, and Yaxing Yao. 2024. Exploring design opportunities for family-based privacy education in informal learning spaces. *Proceedings on Privacy Enhancing Technologies*.

[243] Rachid Ait Maalem Lahcen, Bruce Caulkins, Ram Mohapatra, and Manish Kumar. 2020. Review and insight on the behavioral aspects of cybersecurity. *Cybersecurity*, 3, 1–18.

[244] James E Maddux and Ronald W Rogers. 1983. Protection motivation and self-efficacy: a revised theory of fear appeals and attitude change. *Journal of experimental social psychology*, 19, 5, 469–479.

[245] Steve Mansfield-Devine. 2017. Raising awareness: people are your last line of defence. *Computer Fraud & Security*, 2017, 11, 10–14.

[246] Ioana Andreea Marin, Pavlo Burda, Nicola Zannone, and Luca Allodi. 2023. The influence of human factors on the intention to report phishing emails. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (CHI '23). Association for Computing Machinery, Hamburg, Germany. ISBN: 9781450394215. DOI: 10.1145/3544548.3580985.

[247] Karola Marky, Sarah Prange, Florian Krell, Max Mühlhäuser, and Florian Alt. 2020. "you just can't know about everything": privacy perceptions of smart home visitors. In *Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia*, 83–95.

[248] John Marsden, Zachary Albrecht, Paula Berggren, Jessica Halbert, Kyle Lemons, Anthony Moncivais, and Matthew Thompson. 2020. Facts and stories in phishing training: a replication and extension. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI EA '20). Association for Computing Machinery, Honolulu, USA, 1–6. ISBN: 9781450368193. DOI: 10.1145/3334480.3381435.

[249] Florin Martius, Luisa Jansen, Lukas Struck, Arthi Arumugam, Lisa Geierhaas, Anna-Marie Ortloff, Matthew Smith, and Christian Tiefenau. 2025. Out of sight, out of mind? exploring data protection practices for personal data in usable security & privacy studies. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–16.

[250] Abraham Harold Maslow. 1943. A theory of human motivation. *Psychological review*, 50, 4, 370.

[251] Richard E Mayer. 2008. Applying the science of learning: evidence-based principles for the design of multimedia instruction. *American psychologist*, 63, 8, 760.

[252] EE McCoby. 1983. Socialization in the context of the family: parent-child interaction. *Handbook of child psychology*, 4, 1–101.

[253] Jennifer Dodorico McDonald. 2008. Measuring personality constructs: the advantages and disadvantages of self-reports, informant reports and behavioural assessments. *Enquire*, 1, 1, 1–19.

[254] Joy McLeod, Leah Zhang-Kennedy, and Elizabeth Stobert. 2024. Comparing teacher and creator perspectives on the design of cybersecurity and privacy educational resources. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*, 587–603.

[255] Emily McReynolds, Sarah Hubbard, Timothy Lau, Aditya Saraf, Maya Cakmak, and Franziska Roesner. 2017. Toys That Listen: A Study of Parents, Children, and Internet-Connected Toys. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5197–5207. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025735.

[256] Elisa D Mekler and Kasper Hornbæk. 2019. A framework for the experience of meaning in human-computer interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, 1–15.

[257] Philip Menard, Gregory J Bott, and Robert E Crossler. 2017. User motivations in protecting information security: protection motivation theory versus self-determination theory. *Journal of Management Information Systems*, 34, 4, 1203–1230.

[258] Uta Menges, Jonas Hielscher, Laura Kocksch, Annette Kluge, and M Angela Sasse. 2023. Caring not scaring-an evaluation of a workshop to train apprentices as security champions. In *Proceedings of the 2023 European Symposium on Usable Security*, 237–252.

[259] Laurie L Meschke, Christina Renee Peter, and Suzanne Bartholomae. 2012. Developmentally appropriate practice to promote healthy adolescent development: integrating research and practice. In *Child & Youth Care Forum*. Vol. 41. Springer, 89–108.

[260] Anna Metreveli, Xiaowei Chen, Anders Hedman, and Anastasia Sergeeva. 2025. "who will be left behind?": a swedish case of learning ai in vocational education. *International Journal of Educational Research*, 133, 102697.

[261] John P Meyer, Natalie J Allen, and Catherine A Smith. 1993. Commitment to organizations and occupations: extension and test of a three-component conceptualization. *Journal of applied psychology*, 78, 4, 538.

[262] Marianne Miserandino. 1996. Children who do well in school: individual differences in perceived competence and autonomy in above-average children. *Journal of educational psychology*, 88, 2, 203.

[263] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, Prisma Group, et al. 2010. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *International journal of surgery*, 8, 5, 336–341.

[264] Gregory D Moody, Mikko Siponen, and Seppo Pahnila. 2018. Toward a unified model of information security policy compliance. *MIS quarterly*, 42, 1, 285–A22.

[265] Frederick P Morgeson and Stephen E Humphrey. 2006. The work design questionnaire (wdq): developing and validating a comprehensive measure for assessing job design and the nature of work. *Journal of applied psychology*, 91, 6, 1321.

[266] Vicki G Morwitz, Eric Johnson, and David Schmittlein. 1993. Does measuring intent change behavior? *Journal of consumer research*, 20, 1, 46–61.

[267] Kate Muir and Adam Joinson. 2020. An exploratory study into the negotiation of cyber-security within the family home. *Frontiers in psychology*, 11, 424.

[268] Doruntina Murtezaj, Viktorija Paneva, Verena Distler, and Florian Alt. 2024. Public security user interfaces: supporting spontaneous engagement with it security. In *Proceedings of the New Security Paradigms Workshop*, 56–70.

[269] Alena Naiakshina, Anastasia Danilova, Eva Gerlitz, and Matthew Smith. 2020. On conducting security developer studies with cs students: examining a password-storage study with cs students, freelancers, and company developers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20). Association for Computing Machinery, Honolulu, HI, USA, 1–13. ISBN: 9781450367080. DOI: 10.1145/3313831.3376791.

[270] Alena Naiakshina, Anastasia Danilova, Eva Gerlitz, Emanuel von Zezschwitz, and Matthew Smith. 2019. "if you want, i can store the encrypted password": a password-storage field study with freelance developers. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19). Association for Computing Machinery, Glasgow, Scotland Uk, 1–12. ISBN: 9781450359702. DOI: 10.1145/3290605.3300370.

[271] Jeanne Nakamura, Mihaly Csikszentmihalyi, et al. 2009. Flow theory and research. *Handbook of positive psychology*, 195, 206.

[272] Boon-Yuen Ng, Atreyi Kankanhalli, and Yunjie Calvin Xu. 2009. Studying users' computer security behavior: a health belief perspective. *Decision Support Systems*, 46, 4, 815–825.

[273] NIST. 2023. Phishing. https://www.nist.gov/itl/smallbusinesscyber/guidance-topic/phishing.

[274] Michael Noetel, Shantell Griffith, Oscar Delaney, Taren Sanders, Philip Parker, Borja del Pozo Cruz, and Chris Lonsdale. 2021. Video improves learning in higher education: a systematic review. *Review of educational research*, 91, 2, 204–236.

[275] Jakub Štěpán Novák, Jan Masner, Petr Benda, Pavel Šimek, and Vojtěch Merunka. 2024. Eye Tracking, Usability, and User Experience: A Systematic Review. *International Journal of Human–Computer Interaction*, 40, 17, (Sept. 2024), 4484–4500. DOI: 10.1080/10447318.2023.2221600.

[276] OECD. 2023. Pisa 2022 results (volume ii): learning during–and from–disruption. https://doi.org/10.1787/a97db61c-en. (2023).

[277] Obi Ogbanufe and Ling Ge. 2023. A comparative evaluation of behavioral security motives: protection, intrinsic, and identity motivations. *Computers & Security*, 128, 103136.

[278] Obi Ogbanufe, Russell Torres, and Katia Guerra. 2023. Byoa and security: examining perspective-taking and self-determination. *Journal of Computer Information Systems*, 2023, 1–17.

[279] Antti Oulasvirta and Kasper Hornbæk. 2016. Hci research as problem-solving. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). Association for Computing Machinery, San Jose, California, USA, 4956–4967. ISBN: 9781450333627. DOI: 10.1145/2858036.2858283.

[280] Keshnee Padayachee. 2012. Taxonomy of compliant information security behavior. *Computers & Security*, 31, 5, 673–680.

[281] Jinkyung Katie Park, Mamtaj Akter, Pamela Wisniewski, and Karla Badillo-Urquiola. 2024. It's still complicated: from privacy-invasive parental control to teen-centric solutions for digital resilience. *IEEE Security & Privacy*.

[282] Minjung Park and Sangmi Chai. 2018. Internalization of information security policy and information security practice: a comparison with compliance. In *51st Hawaii International Conference on System Sciences*. University of Hawai'i, Hawai'i, USA, 4723–4731.

[283] J. W. Patchin and S. Hinduja. 2024. 2023 cyberbullying data. cyberbullying research center. https://cyberbullying.org/2023-cyberbullying-data. Accessed: 10-23-2024. (2024).

[284] Kate Payne. 2025. In lawsuit over teen's death, judge rejects arguments that ai chatbots have free speech rights. https://apnews.com/article/ai-lawsuit-suicide-artificial-intelligence-free-speech-ccc77a5ff5a84bda753d2b044c83d4b6. Accessed: 2025-07-12. AP News, (2025).

[285] Douglas D Perkins and Marc A Zimmerman. 1995. Empowerment theory, research, and application. *American journal of community psychology*, 23, 569–579.

[286] Clay Posey, Tom Roberts, Paul Benjamin Lowry, Becky Bennett, and James Courtney. 2010. Insiders' protection of organizational information assets: a multidimensional scaling study of protection-motivated behaviors. In *Roode Workshop on IS Security Research*. SSRN, Boston, MA, USA, 233–277.

[287] Clay Posey, Tom Roberts, Paul Benjamin Lowry, James Courtney, and Becky Bennett. 2011. Motivating the insider to protect organizational information assets: evidence from protection motivation theory and rival explanations. In *The Dewald Roode workshop in information systems security*. SSRN, Kennesaw, GA, 1–51.

[288] Clay Posey, Tom L Roberts, and Paul Benjamin Lowry. 2015. The impact of organizational commitment on insiders' motivation to protect organizational information assets. *Journal of Management Information Systems*, 32, 4, 179–214.

[289] Clay Posey, Tom L Roberts, Paul Benjamin Lowry, Rebecca J Bennett, and James F Courtney. 2013. Insiders' protection of organizational information assets: development of a systematics-based taxonomy and theory of diversity for protection-motivated behaviors. *Mis Quarterly*, 37, 4, 1189–1210.

[290] Travis C Pratt, Francis T Cullen, Kristie R Blevins, Leah E Daigle, and Tamara D Madensen. 2006. The empirical status of deterrence theory: a meta-analysis. In *Taking stock: The status of criminological theory*. Transaction Publishers, New Jersey, USA, 367–395.

[291] Marc Prensky. 2001. Digital natives, digital immigrants part 2: do they really think differently? *On the horizon*, 9, 6, 1–6.

[292] Portia Pusey and William A Sadera. 2011. Cyberethics, cybersafety, and cybersecurity: preservice teacher knowledge, preparedness, and the need for teacher education to make a difference. *Journal of Digital Learning in Teacher Education*, 28, 2, 82–85.

[293] Farzana Quayyum. 2023. Collaboration between parents and children to raise cybersecurity awareness. In *Proceedings of the 2023 European Interdisciplinary Cybersecurity Conference*, 149–152.

[294] Farzana Quayyum, Jonas Bueie, Daniela S Cruzes, Letizia Jaccheri, and Juan Carlos Torrado Vidal. 2021. Understanding parents' perceptions of children's cybersecurity awareness in norway. In *Proceedings of the Conference on Information Technology for Social Good*, 236–241.

[295] Emilee Rader and Rick Wash. 2015. Identifying patterns in informal sources of security information. *Journal of Cybersecurity*, 1, 1, 121–144.

[296] Emilee Rader, Rick Wash, and Brandon Brooks. 2012. Stories as informal lessons about security. In *Proceedings of the Eighth Symposium on Usable Privacy and Security* (SOUPS '12). Association for Computing Machinery, Washington, D.C. ISBN: 9781450315326. DOI: 10.1145/2335356.2335364.

[297] Sara Rahimi and Marzieh khatooni. 2024. Saturation in qualitative research: an evolutionary concept analysis. *International Journal of Nursing Studies Advances*, 6, 100174. DOI: https://doi.org/10.1016/j.ijnsa.2024.100174.

[298] Ellen M Raineri and Jessica Resig. 2020. Evaluating self-efficacy pertaining to cybersecurity for small businesses. *Journal of Applied Business & Economics*, 22, 12, 13–23.

[299] Maike M. Raphael, Aikaterini Kanta, Rico Seebonn, Markus Dürmuth, and Camille Cobb. 2024. Batman hacked my password: a Subtitle-Based analysis of password depiction in movies. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*. USENIX Association, Philadelphia, PA, (Aug. 2024), 199–218. ISBN: 978-1-939133-42-7. https://www.usenix.org/conference/soups2024/presentation/raphael.

[300] Johnmarshall Reeve, Edward L Deci, and Richard M Ryan. 2004. Self-determination theory: a dialectical framework for understanding sociocultural influences on student. *Big theories revisited*, 4, 31.

[301] Andrew Reeves, Dragana Calic, and Paul Delfabbro. 2023. "generic and unusable" 1: understanding employee perceptions of cybersecurity training and measuring advice fatigue. *Computers & Security*, 128, 103137.

[302] Benjamin Reinheimer, Lukas Aldag, Peter Mayer, Mattia Mossano, Reyhan Duezguen, Bettina Lofthouse, Tatiana Von Landesberger, and Melanie Volkamer. 2020. An investigation of phishing awareness and education over time: when and how to best remind users. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*. USENIX Association, Virtual Event, USA, 259–284.

[303] Kristina Reiss, Mirjam Weis, Eckhard Klieme, and Olaf Köller. 2019. *PISA 2018: Grundbildung im internationalen Vergleich*. Waxmann Verlag.

[304] Karen Renaud and Stephen Flowerday. 2017. Contemplating human-centred security & privacy research: suggesting future directions. *Journal of Information Security and Applications*, 34, 76–81.

[305] Karen Renaud, Rosalind Searle, and Marc Dupuis. 2021. Shame in cyber security: effective behavior modification tool or counterproductive foil? In *New Security Paradigms Workshop* (NSPW '21). Association for Computing Machinery, Virtual Event, USA, 70–87.

[306] David B Resnik and Peter R Finn. 2018. Ethics and phishing experiments. *Science and engineering ethics*, 24, 1241–1252.

[307] Hyeun-Suk Rhee, Cheongtag Kim, and Young U Ryu. 2009. Self-efficacy in information security: its influence on end users' information security practice behavior. *Computers & security*, 28, 8, 816–826.

[308] Fabio Rizzoni, Sabina Magalini, Alessandra Casaroli, Pasquale Mari, Matt Dixon, and Lynne Coventry. 2022. Phishing simulation exercise in a large hospital: a case study. *Digital Health*, 8, 20552076221081716.

[309] Ronald W Rogers. 1975. A protection motivation theory of fear appeals and attitude change1. *The journal of psychology*, 91, 1, 93–114.

[310] Vivien E Rolfe and Douglas Gray. 2011. Are multimedia resources effective in life science education? a meta-analysis. *Bioscience education*, 18, 1, 1–14.

[311] Margarida Romero. 2014. Digital literacy for parents of the 21st century children. *Elearning Papers*, 38, 32–40.

[312] Benjamin D Rosenberg and Jason T Siegel. 2018. A 50-year review of psychological reactance theory: do not read this article. *Motivation Science*, 4, 4, 281.

[313] Richard M Ryan and Edward L Deci. 2000. Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemporary educational psychology*, 25, 1, 54–67.

[314] Richard M Ryan, Edward L Deci, et al. 2002. Overview of self-determination theory: an organismic dialectical perspective. *Handbook of self-determination research*, 2, 3-33, 36.

[315] Richard M Ryan and Edward L Deci. 2017. *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Press, New York, USA.

[316] Richard M Ryan and Edward L Deci. 2020. Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions. *Contemporary educational psychology*, 61, 101860.

[317] Richard M Ryan, Edward L Deci, Wendy S Grolnick, and Jennifer G La Guardia. 2015. The significance of autonomy and autonomy support in psychological development and psychopathology. *Developmental psychopathology: Volume one: Theory and method*, 795–849.

[318] Nader Sohrabi Safa and Rossouw Von Solms. 2016. An information security knowledge sharing model in organizations. *Computers in Human Behavior*, 57, 442–451.

[319] Nader Sohrabi Safa, Rossouw Von Solms, and Lynn Futcher. 2016. Human aspects of information security in organisations. *Computer Fraud & Security*, 2016, 2, 15–18.

[320] Rahime Belen Sağlam, Vincent Miller, and Virginia NL Franqueira. 2023. A systematic literature review on cyber security education for children. *IEEE Transactions on Education*, 66, 3, 274–286.

[321] M Angela Sasse, Jonas Hielscher, Jennifer Friedauer, and Annalina Buckmann. 2022. Rebooting it security awareness–how organisations can encourage and sustain secure behaviours. In *European Symposium on Research in Computer Security*. Springer, 248–265.

[322] M Angela Sasse and Awais Rashid. 2021. The cyber security body of knowledge—human factors knowledge area v 1.0. the university of bristol (2019). (2021).

[323] Yukiko Sawaya, Mahmood Sharif, Nicolas Christin, Ayumu Kubota, Akihiro Nakarai, and Akira Yamada. 2017. Self-confidence trumps knowledge: a cross-cultural study of security behavior. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Denver, USA, 2202–2214.

[324] Thomas A Schmitt, Daniel A Sass, Wayne Chappelle, and William Thompson. 2018. Selecting the "best" factor structure and moving measurement validation forward: an illustration. *Journal of personality assessment*, 100, 4, 345–362.

[325] Lorin Schöni, Victor Carles, Martin Strohmeier, Peter Mayer, and Verena Zimmermann. 2024. You know what?-evaluation of a personalised phishing training based on users' phishing knowledge and detection skills. In *The 2024 European Symposium on Usable Security*. Association for Computing Machinery, Karlstad, Sweden, 1–14.

[326] Lorin Schöni, Neele Roch, Hannah Sievers, Martin Strohmeier, Peter Mayer, and Verena Zimmermann. 2025. It's a match - enhancing the fit between users and phishing training through personalisation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery. ISBN: 9798400713941. DOI: 10.1145/3706598.3713845.

[327] Markus Schöps, Marco Gutfleisch, Eric Wolter, and M Angela Sasse. 2024. Simulated stress: a case study of the effects of a simulated phishing campaign on employees' perception, stress and Self-Efficacy. In *33rd USENIX Security Symposium (USENIX Security 24)*, 4589–4606.

[328] Michiel Schotten, Wim JN Meester, Susanne Steiginga, Cameron A Ross, et al. 2017. A brief history of scopus: the world's largest abstract and citation database of scientific literature. In *Research analytics*. Auerbach Publications, Boca Raton, FL, USA, 31–58.

[329] Dale H Schunk and Maria K DiBenedetto. 2020. Motivation and social cognitive theory. *Contemporary educational psychology*, 60, 101832.

[330] Scott E Seibert, Gang Wang, and Stephen H Courtright. 2011. Antecedents and consequences of psychological and team empowerment in organizations: a meta-analytic review. *Journal of applied psychology*, 96, 5, 981.

[331] Anastasia Sergeeva, Björn Rohles, Verena Distler, and Vincent Koenig. 2023. "we need a big revolution in email advertising": users' perception of persuasion in permission-based advertising emails. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (CHI '23). Association for Computing Machinery, Hamburg, Germany. ISBN: 9781450394215. DOI: 10.1145/3544548.3581163.

[332] Ahmad Bakhtiyari Shahri, Zuraini Ismail, and Shahram Mohanna. 2016. The impact of the security competency on "self-efficacy in information security" for effective health information security in iran. *Journal of medical systems*, 40, 1–9.

[333] Susan P Shapiro. 2005. Agency theory. *Annu. Rev. Sociol.*, 31, 1, 263–284.

[334] Filipo Sharevski and Jennifer Vander Loop. 2023. Children, Parents, and Misinformation on Social Media. en. arXiv:2312.09359 [cs]. (Dec. 2023). Retrieved Oct. 2, 2024 from http://arxiv.org/abs/2312.09359.

[335] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elizabeth Nunge. 2007. Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In *Proceedings of the 3rd Symposium on Usable Privacy and Security* (SOUPS '07). Association for Computing Machinery, Pittsburgh, Pennsylvania, USA, 88–99. ISBN: 9781595938015. DOI: 10.1145/1280680.1280692.

[336] Alireza Shojaifar, Samuel A Fricker, and Martin Gwerder. 2020. Automating the communication of cybersecurity knowledge: multi-case study. In *Information Security Education. Information Security in Action. WISE 2020. IFIP Advances in Information and Communication Technology*. Springer, Cham, Switzerland, 110–124.

[337] Bonnie Sibbald and Martin Roland. 1998. Understanding controlled trials. why are randomised controlled trials important? *BMJ: British Medical Journal*, 316, 7126, 201.

[338] Mario Silic and Paul Benjamin Lowry. 2020. Using design-science based gamification to improve organizational security training and compliance. *Journal of management information systems*, 37, 1, 129–161.

[339] Gavin R Slemp, Mark A Lee, and Lara H Mossman. 2021. Interventions to support autonomy, competence, and relatedness needs in organizations: a systematic review with recommendations for research and practice. *Journal of Occupational and Organizational Psychology*, 94, 2, 427–457.

[340] Garrett Smith et al. 2024. "I Know I'm Being Observed:" video interventions to educate users about targeted advertising on facebook. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–27.

[341] Grant Solomon and Irwin Brown. 2021. The influence of organisational culture and information security culture on employee compliance behaviour. *Journal of Enterprise Information Management*, 34, 4, 1203–1228.

[342] Daniel J Solove. 2005. A taxonomy of privacy. *University of Pennsylvania Law Review*, 154, 477.

[343] Teodor Sommestad, Henrik Karlzén, and Jonas Hallberg. 2019. The theory of planned behavior and information security policy compliance. *Journal of Computer Information Systems*, 59:4, 344–353.

[344] Jai-Yeol Son. 2011. Out of fear or desire? toward a better understanding of employees' motivation to follow is security policies. *Information & Management*, 48, 7, 296–302.

[345] Paul E Spector. 1982. Behavior in organizations as a function of employee's locus of control. *Psychological bulletin*, 91, 3, 482.

[346] Gretchen M Spreitzer. 1995. Psychological empowerment in the workplace: dimensions, measurement, and validation. *Academy of management Journal*, 38, 5, 1442–1465.

[347] William Stallings and Lawrie Brown. 2015. *Computer security: principles and practice*. Pearson.

[348] Jeffrey M Stanton, Kathryn R Stam, Paul Mastrangelo, and Jeffrey Jolton. 2005. Analysis of end user security behaviors. *Computers & security*, 24, 2, 124–133.

[349] IBM SPSS Statistics. 2020. Transforming different likert scales to a common scale. (2020).

[350] Michelle Steves, Kristen Greene, and Mary Theofanos. 2020. Categorizing human phishing difficulty: a phish scale. *Journal of Cybersecurity*, 6, 1, tyaa009.

[351] Simon Stockhardt, Benjamin Reinheimer, Melanie Volkamer, Peter Mayer, Alexandra Kunz, Philipp Rack, and Daniel Lehmann. 2016. Teaching phishing-security: which way is best? In *ICT Systems Security and Privacy Protection: 31st IFIP TC 11 International Conference, SEC 2016, Ghent, Belgium, May 30-June 1, 2016, Proceedings 31*. Springer, Ghent, Belgium, 135–149.

[352] Dan N Stone, Edward L Deci, and Richard M Ryan. 2009. Beyond talk: creating autonomous motivation through self-determination theory. *Journal of general management*, 34, 3, 75–91.

[353] Noor Suhani Sulaiman, Muhammad Ashraf Fauzi, Walton Wider, Jegatheesan Rajadurai, Suhaidah Hussain, and Siti Aminah Harun. 2022. Cyber–information security compliance and violation behaviour in organisations: a systematic review. *Social Sciences*, 11, 9, 386.

[354] Alex Sumner, Xiaohong Yuan, Mohd Anwar, and Maranda McBride. 2022. Examining factors impacting the effectiveness of anti-phishing trainings. *Journal of Computer Information Systems*, 62, 5, 975–997.

[355] Kaiwen Sun, Yixin Zou, Jenny Radesky, Christopher Brooks, and Florian Schaub. 2021. Child safety in the smart home: parents' perceptions, needs, and mitigation strategies. *Proceedings of the ACM on Human-Computer Interaction*, 5, CSCW2, 1–41.

[356] Robert I Sutton and Barry M Staw. 1995. What theory is not. *Administrative science quarterly*, 40:3, 371–384.

[357] Valdemar Švábenskỳ, Jan Vykopal, and Pavel Čeleda. 2020. What are cybersecurity education papers about? a systematic literature review of sigcse and iticse conferences. In *Proceedings of the 51st ACM technical symposium on computer science education*, 2–8.

[358] Barbara G Tabachnick, Linda S Fidell, and Jodie B Ullman. 2013. *Using multivariate statistics*. Vol. 6. Pearson Boston, MA.

[359] Maja Tadić Vujčić, Wido GM Oerlemans, and Arnold B Bakker. 2017. How challenging was your work today? the role of autonomous work motivation. *European Journal of Work and Organizational Psychology*, 26, 1, 81–93.

[360] Mohammad Tahaei, Alisa Frik, and Kami Vaniea. 2021. Privacy champions in software teams: understanding their motivations, strategies, and challenges. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15.

[361] Yurita Abdul Talib and Gurpreet Dhillon. 2015. Employee isp compliance intentions: an empirical test of empowerment. In *Thirty Sixth International Conference of Information Systems,* Association for Information Systems, Fort Worth, USA, 1–19.

[362] Anne Clara Tally, Jacob Abbott, Ashley M Bochner, Sanchari Das, and Christena Nippert-Eng. 2023. Tips, tricks, and training: supporting anti-phishing awareness among mid-career office workers based on employees' current practices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (CHI '23). Association for Computing Machinery, Hamburg, Germany. ISBN: 9781450394215. DOI: 10.1145/3544548.3580650.

[363] Alireza Tamjidyamcholo, Mohd Sapiyan Bin Baba, Nor Liyana Mohd Shuib, and Vala Ali Rohani. 2014. Evaluation model for knowledge sharing in information security professional virtual community. *Computers & Security*, 43, 19–34.

[364] Gurvirender PS Tejay and Zareef A Mohammed. 2023. Cultivating security culture for information security success: a mixed-methods study based on anthropological perspective. *Information & Management*, 60, 3, 103751.

[365] Ziwen Teuber and Xiaowei Chen. 2025. Parenting digital natives: a randomized controlled trial on security and privacy education in families. https://doi.org/10.17605/OSF.IO/ZQY7B. OSF Preregistration. (Aug. 2025).

[366] Mary Theofanos, Yee-Yin Choong, and Olivia Murphy. 2021. 'passwords keep me safe' – understanding what children think about passwords. In *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, (Aug. 2021), 19–35. ISBN: 978-1-939133-24-3. https://www.usenix.org/conference/usenixsecurity21/presentation/theofanos.

[367] Asha Thomas and Vikas Gupta. 2022. The role of motivation theories in knowledge sharing: an integrative theoretical reviews and future research agenda. *Kybernetes*, 51, 1, 116–140.

[368] Kenneth W Thomas and Betty A Velthouse. 1990. Cognitive elements of empowerment: an "interpretive" model of intrinsic task motivation. *Academy of management review*, 15, 4, 666–681.

[369] April Tyack and Elisa D. Mekler. 2020. Self-determination theory in hci games research: current uses and open questions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20). Association for Computing Machinery, Honolulu, HI, USA, 1–22. ISBN: 9781450367080. DOI: 10.1145/3313831.3376723.

[370] UChicago. 2024. Latest phishing scams. https://security.uchicago.edu/phishing/latest/. Accessed: 10-02-2024. (2024).

[371] Daniel Udo-Akang. 2012. Theoretical constructs, concepts, and applications. *American International Journal of Contemporary Research*, 2, 9, 89–97.

[372] UNICEF. 2018. More than 175,000 children go online for the first time every day, tapping into great opportunities. https://www.unicef.org/eca/press-releases. Accessed: 12-01-2024. (2018).

[373] UNICEF. 2023. Convention on the rights of the child - children's version. https://www.unicef.org/child-rights-convention/convention-text-childrens-version. Accessed: 12-01-2024. (2023).

[374] René Van Bavel, Nuria Rodríguez-Priego, José Vila, and Pam Briggs. 2019. Using protection motivation theory in the design of nudges to improve online security behavior. *International Journal of Human-Computer Studies*, 123, 29–39.

[375] O Van den Akker et al. 2020. Generalized systematic review registration form. (2020).

[376] Anthony Vance, Mikko Siponen, and Seppo Pahnila. 2012. Motivating is security compliance: insights from habit and protection motivation theory. *Information & Management*, 49, 3-4, 190–198.

[377] Ali Vedadi, Merrill Warkentin, Detmar W Straub, and Jordan Shropshire. 2024. Fostering information security compliance as organizational citizenship behavior. *Information & Management*, 61, 5, 103968.

[378] Antje C. Venjakob and Claudia R. Mello-Thoms. 2015. Review of prospects and challenges of eye tracking in volumetric imaging. *Journal of Medical Imaging*, 3, 1, (Sept. 2015), 011002. Publisher: SPIE. DOI: 10.1117/1.JMI.3.1.011002.

[379] VERBIsoftware. 2024. Maxqda. https://www.maxqda.com/. Accessed: 10-02-2024. (2024).

[380] Silas Formunyuy Verkijika. 2019. "if you know what to do, will you take action to avoid mobile phishing attacks": self-efficacy, anticipated regret, and gender. *Computers in Human Behavior*, 101, 286–296.

[381] Arnold POS Vermeeren, Effie Lai-Chong Law, Virpi Roto, Marianna Obrist, Jettie Hoonhout, and Kaisa Väänänen-Vainio-Mattila. 2010. User experience evaluation methods: current state and development needs. In *Proceedings of the 6th Nordic conference on human-computer interaction: Extending boundaries*, 521–530.

[382] Melanie Volkamer, Karen Renaud, Benjamin Reinheimer, Philipp Rack, Marco Ghiglieri, Peter Mayer, Alexandra Kunz, and Nina Gerber. 2018. Developing and evaluating a five minute phishing awareness video. In *Trust, Privacy and Security in Digital Business: 15th International Conference, TrustBus 2018, Regensburg, Germany, September 5–6, 2018, Proceedings 15*. Springer, 119–134.

[383] Melanie Volkamer, Martina Angela Sasse, and Franziska Boehm. 2020. Analysing simulated phishing campaigns for staff. In *Computer Security: ESORICS 2020 International Workshops, DETIPS, DeSECSys, MPS, and SPOSE, September 17–18, 2020, Revised Selected Papers 25*. Springer, Guildford, UK, 312–328. DOI: https://doi.org/10.1007/978-3-030-66504-3_19.

[384] Alexandra von Preuschen, Carolin Benda, Monika Christine Schuhmacher, and Verena Zimmermann. 2025. Fear, fun or none: a qualitative quest towards unlocking cybersecurity attitudes. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–24.

[385] Alexandra von Preuschen, Monika C Schuhmacher, and Verena Zimmermann. 2024. Beyond fear and frustration-towards a holistic understanding of emotions in cybersecurity. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*. USENIX Association, Philadephia, USA, 623–642.

[386] Maximilian von Welck, Manuel Trenz, Tina Blegind Jensen, and Daniel Veit. 2017. Empowerment and byox: towards improved is security compliance. In *38th International Conference on Information Systems: Transforming Society with Digital Innovation, ICIS 2017: Transforming Society with Digital Innovation*. Association for Information Systems, Atlanta, GA, USA, 1–11.

[387] Joan IJ Wagner, Greta Cummings, Donna L Smith, Joanne Olson, Lynn Anderson, and Sharon Warren. 2010. The relationship between structural empowerment and psychological empowerment for nurses: a systematic review. *Journal of nursing management*, 18, 4, 448–462.

[388] René Walendy et al. 2024. I see an ic: a mixed-methods approach to study human problem-solving processes in hardware reverse engineering. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (CHI '24). Association for Computing Machinery, Honolulu, HI, USA. ISBN: 9798400703300. DOI: 10.1145/3613904.3642837.

[389] Jeffrey D Wall, Prashant Palvia, and Paul Benjamin Lowry. 2013. Control-related motivations and information security policy compliance: the role of autonomy and efficacy. *Journal of Information Privacy and Security*, 9, 4, 52–79.

[390] Merrill Warkentin, Allen C Johnston, and Jordan Shropshire. 2011. The influence of the informal social learning environment on information privacy policy compliance efficacy and intention. *European Journal of Information Systems*, 20, 3, 267–284.

[391] Rick Wash. 2010. Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security* (SOUPS '10). Association for Computing Machinery, Redmond, Washington, USA. ISBN: 9781450302647. DOI: 10.1145/1837110.1837125.

[392] Rick Wash and Molly M. Cooper. 2018. Who provides phishing training? facts, stories, and people like me. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18). Association for Computing Machinery, Montreal QC, Canada, 1–12. ISBN: 9781450356206. DOI: 10.1145/3173574.3174066.

[393] Rick Wash and Emilee Rader. 2011. Influencing mental models of security: a research agenda. In *Proceedings of the 2011 New Security Paradigms Workshop* (NSPW '11). Association for Computing Machinery, Marin County, California, USA, 57–66. ISBN: 9781450310789. DOI: 10.1145/2073276.2073283.

[394] Patrickson Weanquoi, Jaris Johnson, and Jinghua Zhang. 2018. Using a game to improve phishing awareness. *Journal of Cybersecurity Education, Research and Practice*, 2018, 2, 2.

[395] Bradley W Weaver, Adam M Braly, and David M Lane. 2021. Training users to identify phishing emails. *Journal of Educational Computing Research*, 59, 6, 1169–1183.

[396] Joseph Weizenbaum. 1976. *Computer Power and Human Reason: From Judgment to Calculation*. W. H. Freeman and Company, San Francisco.

[397] Zikai Alex Wen, Zhiqiu Lin, Rowena Chen, and Erik Andersen. 2019. What. hack: engaging anti-phishing training through a role-playing phishing simulation game. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Glasgow, Scotland, Uk, 1–12.

[398] Alma Whitten and J. D. Tygar. 1999. Why johnny can't encrypt: a usability evaluation of pgp 5.0. In *Proceedings of the 8th Conference on USENIX Security Symposium - Volume 8* (SSYM'99). USENIX Association, Washington, D.C., 14.

[399] Allan Wigfield and Jacquelynne S Eccles. 2000. Expectancy–value theory of achievement motivation. *Contemporary educational psychology*, 25, 1, 68–81.

[400] Allan Wigfield, Stephen Tonks, and Susan Lutz Klauda. 2009. Expectancy-value theory. *Handbook of motivation at school*, 2, 55–74.

[401] Sue Wilkinson. 1998. Focus group methodology: a review. *International journal of social research methodology*, 1, 3, 181–203.

[402] Emma J Williams, Joanne Hinds, and Adam N Joinson. 2018. Exploring susceptibility to phishing in the workplace. *International Journal of Human-Computer Studies*, 120, 1–13.

[403] Emma J Williams and Adam N Joinson. 2020. Developing a measure of information seeking about phishing. *Journal of Cybersecurity*, 6, 1, tyaa001.

[404] Penny Williams. 2022. Organisational culture: definitions, distinctions and functions. *Handbook of research methods for organisational culture*, 5–22.

[405] Robert Willison and Merrill Warkentin. 2013. Beyond deterrence: an expanded view of employee computer abuse. *MIS quarterly*, 37, 1, 1–20.

[406] Rogier Woltjer, Jiri Trnka, Jonas Lundberg, and Björn Johansson. 2006. Role-playing exercises to strengthen the resilience of command and control systems. In *Proceedings of the 13th Eurpoean conference on Cognitive ergonomics: trust and control in complex socio-technical systems*. ECCE '06, Zurich, Switzerland, 71–78.

[407] Michael Workman, William H. Bommer, and Detmar Straub. 2008. Security lapses and the omission of information security measures: a threat control model and empirical test. *Comput. Hum. Behav.*, 24, 6, (Sept. 2008), 2799–2816. DOI: 10.1016/j.chb.2008.04.005.

[408] Ning Yang, Tripti Singh, and Allen Johnston. 2020. A replication study of user motivation in protecting information security using protection motivation theory and self determination theory. *AIS Transactions on Replication Research*, 6, 1, 10.

[409] William Yeoh, He Huang, Wang-Sheng Lee, Fadi Al Jafari, and Rachel Mansson. 2022. Simulated phishing attack and embedded training campaign. *Journal of Computer Information Systems*, 62, 4, 802–821.

[410] Yaman Yu, Tanusree Sharma, Melinda Hu, Justin Wang, and Yang Wang. 2024. Exploring parent-child perceptions on safety in generative ai: concerns, mitigation strategies, and design implications. English. In *Proceedings of 2025 IEEE Symposium on Security and Privacy (SP)*. San Francisco, US.

[411] Jie Zhang, Xin Luo, Somasheker Akkaladevi, and Jennifer Ziegelmayer. 2009. Improving multiple-password recall: an empirical study. *European Journal of Information Systems*, 18, 2, 165–176.

[412] Leah Zhang-Kennedy and Sonia Chiasson. 2021. A systematic review of multimedia tools for cybersecurity awareness and education. *ACM Comput. Surv.*, 54, 1, (Jan. 2021). DOI: 10.1145/3427920.

[413] Sarah Ying Zheng and Ingolf Becker. 2023. Phishing to improve detection. In *Proceedings of the 2023 European Symposium on Usable Security* (EuroUSEC '23). Association for Computing Machinery, Copenhagen, Denmark, 334–343. ISBN: 9798400708145. DOI: 10.1145/3617072.3617121.

[414] Verena Zimmermann and Karen Renaud. 2019. Moving from a 'human-as-problem" to a 'human-as-solution" cybersecurity mindset. *International Journal of Human-Computer Studies*, 131, 169–187.

[415] Verena Zimmermann, Lorin Schöni, Thierry Schaltegger, Benjamin Ambuehl, Melanie Knieps, and Nico Ebert. 2024. Human-centered cybersecurity revisited: from enemies to partners. *Commun. ACM*, 67, 11, (Oct. 2024), 72–81. DOI: 10.1145/3665665.

[416] Yixin Zou, Khue Le, Peter Mayer, Alessandro Acquisti, Adam J Aviv, and Florian Schaub. 2024. Encouraging users to change breached passwords using the protection motivation theory. *ACM Transactions on Computer-Human Interaction*, 1, 1, 1–45.

[417] Yixin Zou, Abraham H Mhaidli, Austin McCall, and Florian Schaub. 2018. " i've got nothing to lose": consumers' risk perceptions and protective actions after the equifax data breach. In *Fourteenth Symposium on Usable Privacy and security (soups 2018)*, 197–216.

[418] Mary Ellen Zurko and Richard T Simon. 1996. User-centered security. In *Proceedings of the 1996 workshop on New security paradigms*, 27–33.

Appendices

Chapter A

Glossary of Theoretical Frameworks applied in security behavior studies

Agency Theory: Agency Theory "is mainly concerned with the efforts provided by the individual members and of motivating them to obtain the desired effort input. An agency relationship exists whenever one party (principal) entrusts some decision-making authority to another party (agent)" [170, p.155]. Agency theory has been applied in the fields of economics, law, political science, and sociology [333].

Cognitive Model of Empowerment: "Individuals' judgments and behavior regarding tasks also are shaped by cognitions that go beyond verifiable reality. Such interpretive cognitions go beyond the perception of facts to provide additional, needed meaning for an individual" [368, p.669]. Psychological empowerment is formed based on individuals' assessments of a task regarding: impact, competence, meaningfulness, and choice [368]. The original model was published in an organizational science venue. Talib and Dhillon [361] referred to the model as the "intrinsic motivation model" in their work.

Deterrence Theory: "Individuals are less likely to commit a deviant activity when the risks of getting caught and the severity of the punishment increase" [344, p.297]. There are two central constructs in the theory: *Deterrent certainty* refers to the high likelihood of sanctions for violations of policies or rules, whereas *deterrent severity* refers to the harshness of the sanctions. The theory is rooted in the classical school of criminology [290].

Expectancy-Value Theory: "Individuals' choice, persistence, and performance can be explained by their beliefs about how well they will do in the activity and the extent to which they value the activity" [399, p.68]. Expectancy-Value Theory is a popular motivation theory in education contexts, but it has rarely been applied in information security studies [70].

Flow Theory: Flow Theory describes a state of deep immersion and engagement in an activity, where individuals experience intense focus, a sense of control, and intrinsic enjoyment [338, 271]. This state occurs when the challenge of the task is aligned with the individual's skills, leading to a balance between challenge and ability [271]. In flow, individuals lose awareness of time and external distractions, becoming fully absorbed in the task at hand [271].

Gaming "Theory": Gamification involves "applying game-like design artifacts and system processes to strengthen employees' motivations to encourage learning, efficacy, and increased employee compliance with organizational security initiatives" [338, p.131]. Silic and Lowry [338] conducted a design-science research project. Gaming theory and flow theory were used to guide the intervention design in the study.

Kanter's Model of Structural Empowerment: This model posits that power within organizations originates from two key systemic sources: formal and informal power [387]. Formal power is associated with roles that are highly visible, central to the organization's operations, and require autonomous decision-making [387]. By contrast, informal power is derived from relationships and alliances with superiors, peers, and subordinates [387]. These two forms of power facilitate access to job-related empowerment structures [387]: support (feedback and guidance), information (data, technical knowledge, and expertise), resources (time, materials, money, supplies, and equipment), and opportunity (autonomy, growth potential, sense of challenge, and learning opportunities).

Motive-Control Theory of Insider Computer Abuse: This theory "distinguishes between the influences of expressive and instrumental motives on insider computer abuse and explains how intrinsic (i.e., self-control) and extrinsic (i.e., organizational deterrence) controls moderate these relationships" [58, p.3]. Burns et al. [58] proposed this middle-range theory to focus on understanding the inherent tension between insider motivations and organizational controls.

Organizational Justice "Theory": There are four dimensions of employees' perceived fairness/unfairness in organizations or what is interchangeably termed justice/injustice: distributive justice, procedural justice, interpersonal justice, informational justice [405]. Li et al. [236] referred to this theoretical summary as "organizational justice theory." Can we cite a summary of previous findings, in this case the four dimensions of employees' perception of fairness/justice, as a theory? In which condition can we name a conceptual summary as a theory?

Organismic Integration Theory: When an individual internalizes external regulations (e.g., ISP), they will autonomously comply with these regulations [218]. Ryan and Deci examined "what motivates individuals to engage in behaviors and practices that are not necessarily intrinsically interesting" [315, p.179]. They propose that "supports for the basic needs for competence, relatedness, and autonomy facilitate the internalization and integration of non-intrinsically motivated behaviors" [315, p.203]. Organismic Integration Theory is a sub-theory of Self-Determination Theory [315].

Person Organization Fit Theory: An employee's behavior results from interactions between the individual and the organizational environment [67]. Person-Organization fit is achieved when (a) one provides what the other needs—either the individual's abilities meet the organization's demands (demand-ability fit) or the organization satisfies the individual's needs (need-supply fit); (b) they share similar values, attitudes, and goals; or (c) both [67].

Psychological Empowerment: Psychological empowerment was formed on the basis of an individual's assessment of a task in terms of competence, meaning, impact, and self-determination [346]. Empowerment reflects personal perceptions of a task and one's ability to control, shape, or influence that task [368, 346]. Individuals are intrinsically motivated when they experience these cognitions (competence, meaning, impact, and self-determination) in relation to a task [99].

Protection Motivation Theory: This theory was originally proposed by Rogers to understand individuals' health behaviors [309]. The theory posits that when an individual is confronted with a threat, they cognitively assess the threat and possible associated remedies [309]. On the basis of their assessment of the threat (threat susceptibility, threat severity, and rewards) and their coping appraisals (response efficacy, self-efficacy, and response cost), the individual decides to act in either an adaptive or maladaptive way [309, 257].

Rational Choice Theory: This theory "offers a theoretical explanation for how individuals make decisions when faced with choices. Rational Choice Theory argues that an individual determines how he will act by balancing the costs and benefits of his options" [53, p.527]. Whereas Awudu and Terzis [25] did not refer to a specific theory in their study design, Rational Choice Theory was emphasized in their source of measurement items.

Reactance Theory: Reactance Theory suggests that individuals desire freedom and that individuals will strive to restore freedoms that they perceive to be threatened by external control [49]. The attempt to restore freedom is referred to as psychological reactance, "a motivational state that drives freedom restoration" [312, p.1]. Reactance is conceptualized as being a stable personality trait as well as a behavioral response [389].

Self-Determination Theory: Self-Determination Theory proposes that "humans have evolved to be inherently curious, physically active and deeply social beings. Individual human development is characterized by proactive engagement, assimilating information and behavioral regulations, and finding integration within social groups" [315, p.4].

Social Bond Theory: Social Bond Theory posits that "when people build upon social bonds, their urge to indulge in anti-social or anti-establishment behaviors is reduced" [191, p.70]. There are four social bonds that promote socialization and conformity: attachment, commitment, involvement, and personal norms [191].

Social Cognitive Theory: Social Cognitive Theory posits that "individuals are actively engaged in their own development and obtain desired results when they believe that their actions are under their own control" [191, p.70]. Social cognitive theory "emphasizes the critical role played by the social environment on motivation, learning, and self-regulation" [329, p.1].

Social Exchange Theory: Social Exchange Theory posits that "individuals interact with one another when expecting beneficial outcomes, such as social rewards. Social rewards comprise reputation, status, respect, and social image" [129, p.4529]. Social exchanges lead to mutually beneficial transactions and relationships over time [83].

Social Influence theory: Kelman [212] distinguished three different processes of influence: compliance, identification, and internalization. "*Compliance* occurs when an individual accepts social influences in an attempt to receive a certain reward or avoid punishment. *Identification* happens when an individual perceives the importance of an issue and then shows a willingness to conform. *Internalization* takes place when an organization's value systems and norms coincide with those of the individual via the admission of social influences" [282, p.4724].

Theory of Planned Behavior: An individual's intentions to engage in certain behaviors are determined by *attitude*, *subjective norm*, and *perceived behavioral control* [5]. Attitude is an individual's positive or negative feelings toward engaging in a specified behavior, and the formation of attitude can be examined through an expectancy-value formulation [5]. Subjective norms describe an individual's perceptions of others' expectations. Perceived behavioral control, conceptually similar to self-efficacy, captures the extent to which an individual has the ability to perform the behavior and how much the behavior is under their control [5].

Theory of Primary Message Systems: This theory provides a taxonomy of behavioral patterns used to interpret and understand culture [364]. E.T. Hall identified 10 primary systems, each representing a distinct stream of cultural communication that interacts with others to produce the complex patterns of behavior [364]. These systems form the underlying structure through which cultural norms and values are conveyed, often nonverbally, within a society [364].

Theory of Workarounds: This theory describes the idea that established work practices may be adhered to or deviated from based on a variety of factors [19]. These include "the quality and practicality of the work practices, obstacles or anomalies that may be encountered by work system participants, and the monitoring and reward systems governing the work system" [202, p.58]. This theory emphasizes the dynamic nature of work processes, where employees often adapt or bypass formal procedures to achieve their goals under varying conditions [19].

Work Design Theory: Work Design Theory explores the relationships between the characteristics of work and the resulting employee outcomes, such as job satisfaction, motivation, and performance [159]. According to this theory, work can be structured in various ways to influence these outcomes, with a particular emphasis on task characteristics that shape the employee's experience [159]. Morgeson and Humphrey [265] expanded on traditional models by incorporating three key types of autonomy in task characteristics: work scheduling autonomy, decision-making autonomy, and work methods autonomy.

Chapter B

Publication Venue of reviewed studies

Table B.1: Publication venue of reviewed studies.

Type	Scope	Venue	Reviewed study
Journals	Information Systems	*Information & Management*	[191, 344, 364, 377]
		Journal of Management Information Systems	[257, 288, 338]
		MIS Quarterly	[53]
		Information Systems Research	[58]
		European Journal of Information Systems	[92]
		Journal of the Association for Information Systems	[99]
		Decision Support Systems	[170]
		Information Systems Journal	[236]
		Information Systems and e-Business Management	[7]
		ACM The Data Base for Advances in Information Systems	[202]
		AIS Transactions on Replication Research	[408]
	Interdisciplinary	*Computers in Human Behavior*	[42, 318]
		Behaviour & Information Technology	[67, 204]
		Information Technology & People	[203]
		Journal of Medical Systems	[332]
		Journal of Psychosocial Research on Cyberspace	[138]
		Journal of Organizational and End User Computing	[186]
		Journal of Computer Information Systems	[278]
		Cognition, Technology & Work	[235]
		International Journal of Advanced Computer Science and Applications	[22]
	Security-focused	*Computer & Security*	[12, 131, 277]
		Information & Computer Security	[136]
		Journal of Information Privacy and Security	[389]
Conferences	Information Systems	*Hawaii International Conference on System Sciences*	[129, 201, 282]
		AIS International Conference on Information Systems	[386, 218, 361]
		IEEE International Conference on Information Management	[21]
	Security and Privacy	*Symposium on Usable Privacy and Security (SOUPS)*	[43]
		Human Aspects of Information Security, Privacy, and Trust International Conference	[185]
		The Dewald Roode Workshop in Information Systems Security	[287]
	Human aspects of technology use	*International Conferences on e-Society and Mobile Learning*	[25]
		IFIP World Conference on Information Security Education	[336]
		ACM SIGMIS Computers and People Research	[164]

Chapter C

The core constructs of Expectancy-Value Theory

The core constructs of Expectancy-Value Theory as described in Eccles and Wigfield's work [111] are as follows:

- **Expectation of Success**: Individuals' beliefs regarding their potential effectiveness in executing tasks or resolving challenges [111].
- **Achievement-Related Choices and Performance**: The outcomes that individuals target when they choose to engage with an activity or perform a task, informed by their interpretation of expectation of success and perceived value of the specific task [111].
- **Subjective Task Value**: Individuals' assessment of a task's significance, utility, emotional resonance, and perceived cost [111].
- **Goal**: Cognitive representation of a future outcome that an individual is striving to achieve [110].
- **Self-schemata**: Cognitive generalizations about oneself, derived from past experiences and focused on self regarded importance [213].
- **Affective Reactions and Memories**: Individuals' emotional responses to specific tasks or scenarios, alongside the emotive memories derived from past experiences [399].
- **Perception of**: Individuals' interpretation and understanding of their previous experiences and socialization influences [399].
- **Interpretation of Experience**: The personal lens through which individuals perceive prior achievement-related events, influenced by a confluence of cultural, social, external feedback, and intrinsic cognitive and emotional factors [399].
- **Cultural Milieu**: A system of social roles, each with its associated responsibilities and obligations [400], this construct has been extended in our study to encompass "organizational culture."
- **Socializer**: Originally pertaining to parents, educators, and extended social circles in EVT [399], this construct has been adapted in our context to also include "colleagues and supervisors."
- **Person Characteristics**: The array of individual variances, encapsulating aspects such as abilities, personality dimensions, gender, age, and cultural origins [399].
- **Previous Achievement-Related Experiences**: Individuals' past experiences in activities or tasks that had a measurable outcome [111].

Chapter D

The templates and focus group protocol

Introduction: Thank you for participating in this focus group discussion. This study is one part of the "AES Anti-phishing" project, funded by the Institute for Advanced Studies. This focus group aims to learn about employees' participation in and opinions on phishing awareness campaigns and reporting suspicious emails.

During the discussion, we will record audio and video and collect the paper materials. The collected data will only be used for this study. You have the right to access, rectify, and erase your data. Your participation in the project is voluntary; you can withdraw at any point without giving reasons. You may skip any task you do not wish to participate in for any reason, at any time, without explanation.

There are no right or wrong answers to the questions we prepared; also, we will not ask you questions about your passwords or whether you have encountered phishing attacks in the past. All your answers will be kept strictly confidential and will be anonymized, encrypted and only reviewed by the researchers of this project. Any data shown externally, for example in publications or presentations, will also be anonymized. Your data will be stored and processed only for the purpose of the study stated above for a period of 63 months on internal, on-premises servers.

The focus group will take approximately 90 minutes. Each participant will be compensated with a 40-euro voucher for participation. Do you have any questions so far? If you agree with the terms, please sign the consent form, and then we can start the recording and begin the focus group discussion. The focus group includes four main parts: warm-up activity, discussion, brainstorming and debriefing. Let's first have the warm-up activity.

Part 1: Warm-up activity (10 minutes):

Icebreaker: Now, you have 2 minutes to observe the items presented in the lab, try to spot one item that can be used to describe you today. We will share our thoughts after 2 minutes.

Explore motivational and discouraging factors for a leisure activity: Great, now we know each other. Let's move on to explore factors that motivate and discourage you from engaging in a leisure activity. You have 5 minutes to answer the questions on Template 1 (see Figure D.1). After you finish, we will collect the paper.

Part 2: Group discussion (60 minutes):

Now, let's move on to the discussion session. Phishing attack is a type of social engineering attack where attackers send spoofed or deceptive messages to trick a person into revealing sensitive information to the attacker or to deploy malicious software on the recipient's devices. Currently at our

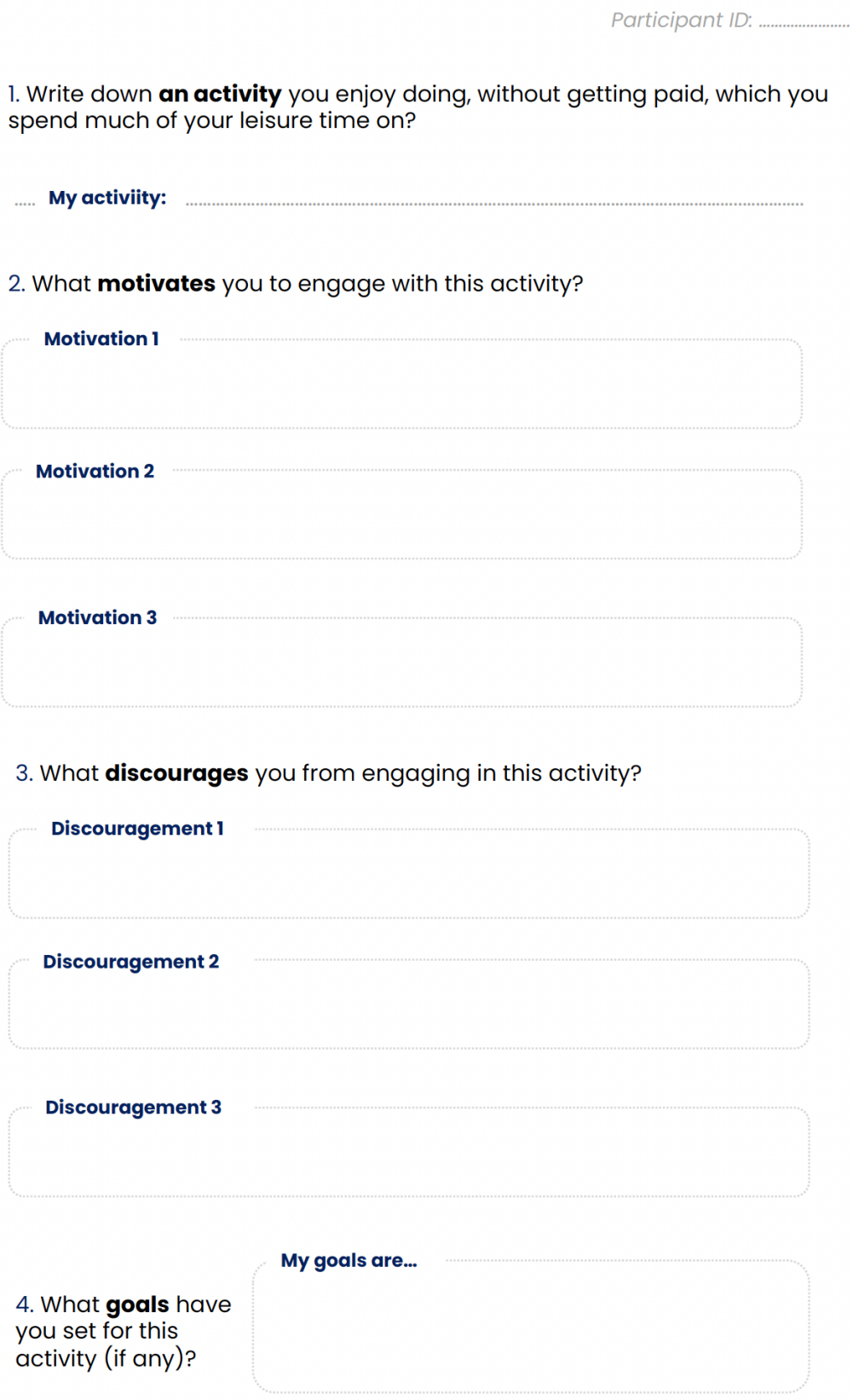

Figure D.1: Template 1, what motivates and discourages you in a leisure activity.

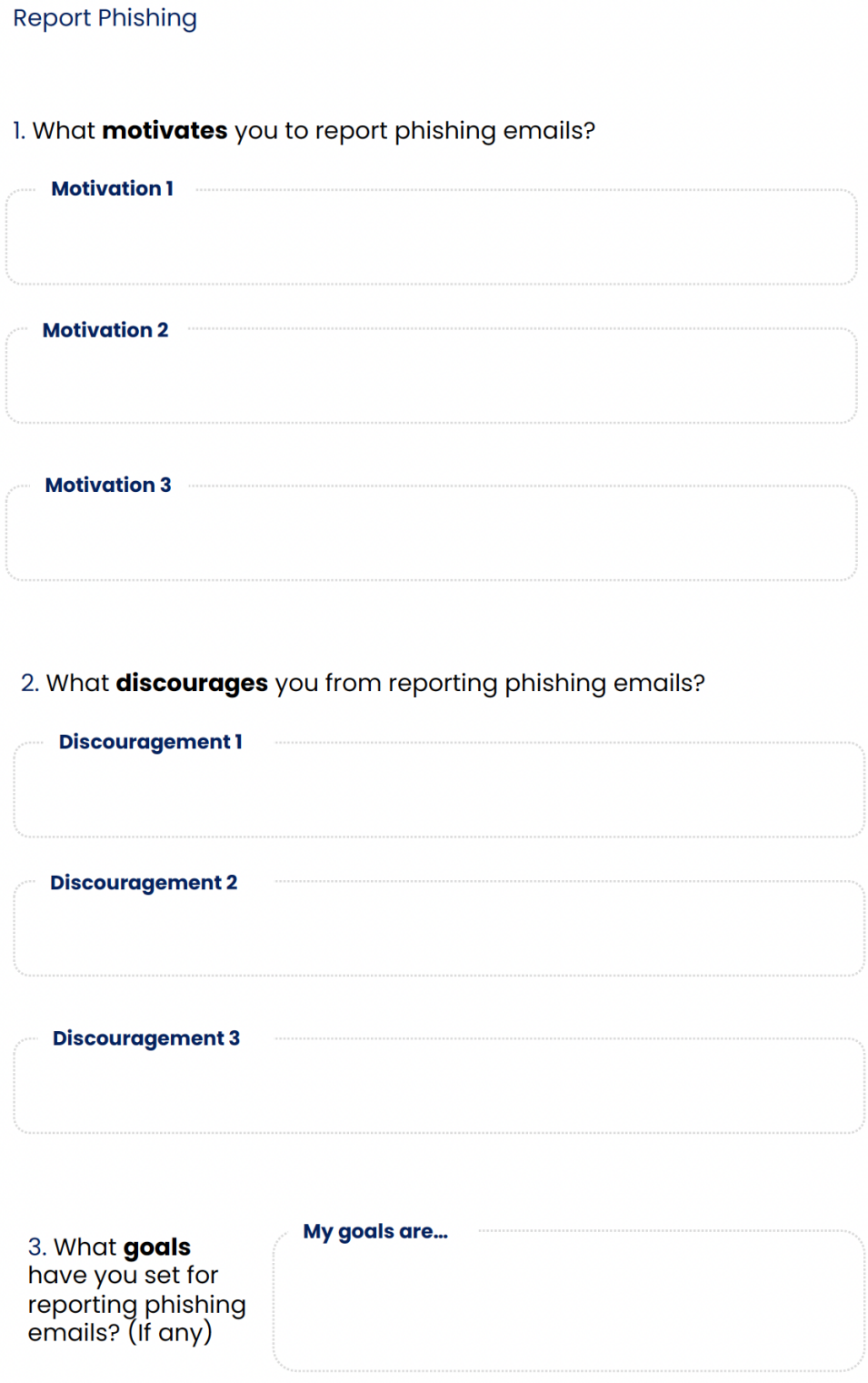

Figure D.2: Template 2, what motivates and discourages you in reporting.

organization, we have several practices to raise employees' awareness of phishing attacks. First, the IT department sends simulated phishing emails to employees to raise awareness of potential phishing attacks. Second, our university has purchased online security courses from a service provider; you can access the learning platform via this link: "Anonymized". Third, the IT department distributes posters and sends emails to inform employees of online security courses. Some of you might have received these emails or saw the posters at the entrance to the administrative buildings.

Discuss phishing awareness campaigns:

1. What do you think of these three phishing campaigns offered by the IT team?
2. What are the benefits of participating in phishing campaigns?
3. What are the costs of participating in phishing campaigns?
4. Assuming that you know how to take the online security courses, what would discourage you from taking these courses?
5. Have you set any goals for yourself in terms of defending yourself from phishing attacks?
6. How confident are you in protecting yourself from phishing attacks?

Thank you for sharing these opinions with us. In our university, the IT department recommends that employees report phishing emails to report-a-phish@uni.lu; the Outlook client now also has a report phishing emails button, so you can report with one click.

Now, you have five minutes to fill-in Template 2 (see Figure D.2), "what motivates and discourages you from reporting suspicious emails"... Thank you and let's move on to **discuss reporting suspicious emails**:

1. The IT department suggests that we report phishing emails, what do you think of this suggestion?
2. How confident are you about identifying and reporting suspicious emails?
3. As a member of the organization, how do you see your role in reporting suspicious emails?
4. What are the benefits of reporting suspicious emails?
5. What are the costs of reporting suspicious emails?
6. What would discourage you from reporting phishing emails?

Part 3: Brainstorming (15 minutes): Assume that you are our university's new chief information security officer (CISO), and you learned that there are increasing phishing emails targeted at our university. What would you do to motivate employees to engage with these counter-phishing practices?

Part 4: Debriefing (5 minutes): Introduce the IT department recommendations of participating in phishing awareness campaigns and reporting suspicious emails.

Chapter E

Coding scheme for focus groups

E.0.1 Factors associated with phishing awareness campaigns

Motivating factors

Gaining phishing knowledge: Participants learned about the techniques and tricks of phishing attacks.

If you were participating in this awareness campaign, maybe get to know some new tricks and what is going on. Maybe there are new types of phishing. (P3)

Acquiring skills: Employees acquired skills in identifying whether the emails, links and website URLs are legitimate or not.

(Phishing campaigns)... train people to recognize what is phishing and prevent them from actually falling into one when it happens. (P22)

Enhancing phishing awareness: The phishing campaign raised employees' awareness of phishing attempts and made them more vigilant against potential attacks.

The good thing is if we make mistakes, they don't cost anything because they're internal mistakes. But they raise our awareness. (P26)

Cyber safety: Participants felt better prepared to protect themselves, their emails, and their workplace from phishing attacks.

It not only benefits you because you will protect your data and your e-mail accounts and so on; will also help the university as an institution to be better protected. (P9)

Personal development: Participants believed that the knowledge gained could benefit their daily life.

It's not only about fear of being attacked, you need to understand what's inside these technology tools... Everything related to cybersecurity is very fundamental now and, in the future, would become even more fundamental, like reading. (P29)

Discouraging factors

Perceived low value: Participants assumed that online phishing courses only provide very basic knowledge or use too complex terms for them to understand.

Don't give me a half hour course for two minutes' value. (P13)

Lack of interest: Negative impressions of the courses, such as "not interesting" and "too easy".

They look like really boring corporate mandated trainings and also the title "Anonymized", look at that and I'd be like oh no... (P17)

Secondary task: Participants mentioned that the phishing campaign was not relevant to their area of expertise or job position.

My role is more task oriented. So, I have to finish my tasks by the end of the day. If I take a course that's one hour long, that means I leave one hour later. (P24)

Lack of incentive: Participants considered lack of incentives, such as course credits, compensation, or praise from the team leader, as discouraging engagement with the awareness campaign.

What is my incentive to do an optional course here? (P24)

Time: Participants mentioned time as a constraint that discourages them from engaging in phishing campaigns.

Sometimes when you are busy, it is very hard to find an hour or so in a day to do them, and so it's quite a big constraint on that. I would say it's mainly time. (P9)

Interrupting workflow: Participating in awareness campaigns required people to switch away from the task at hand to phishing-related content.

The cost is the time spent, but also entering into the actual narrative and that type of discourse. Because you're doing something else and then you're switching to this. And you're like, OK, it's a completely different world, so it takes you away from your attention span. (P25)

Optimism bias: Participants mentioned that they believed they were less likely to fall for phishing than others.

I always had this thinking like, it won't happen to me because this is so stupid. (P14)

Overconfidence: Participants stated that they are very confident in their knowledge of the topic.

I should spend my time doing something else so it's like a prerequisite of this course like ... like 70 to 80% of course material they have already known. (P21)

Procrastination: Participants shared that procrastination resulted in delaying or "forgetting to" take the courses.

If there's no deadline, if there's no shock, I'll do it tomorrow, tomorrow, tomorrow. (P32)

Negative inference: Participants would become more worried about all the potential threats they might receive if they participated in awareness campaigns.

More negative inference ... we become a bit more scared about all these potential threats that we might receive. A little bit of stress in a sense that we need to be careful. (P30)

Fear of failing training: Previous bad experiences with awareness campaigns might evoke fear of failing the training.

The fear or the worry that if I failed the course, it would be tracked. Because I experienced that in the previous job. If you didn't get a certain grade, then you would be forced to retake it and retake

it. (P8)

E.0.2 Factors associated with reporting

Motivating factors

Collaborating with the IT team: Participants considered reporting as a collaboration with the IT team. The IT team assists the employees in verifying the legitimacy of the emails, and employees assist the IT team in detecting the phishing attempts in real-time.

I think this is essential that we can report phishing to IT; and based on that they can have some statistics and see how the attacks are evolving. (P5)

Safeguarding the workplace: Participants regarded reporting as a measure to protect their workplace and colleagues.

Safeguard yourself, your institution, because I'm aware of phishing attacks that cause huge damages in the banking and insurance sector, in research departments overseas, and it's reputational damage that I would not like to be associated with. So protection for the whole institution and for me ultimately. (P13)

Expectation of mitigation: Participants expected that the organization would improve its spam filters and mitigate the attack promptly with the reported emails.

The main benefit of reporting is that the IT team could create more filters for phishing emails if they have more data (from reporting), making us safer (P27)

Recognition: Participants regarded the "congratulations" email they received from the IT team as a kind of recognition and extrinsic reward for their reporting.

And personally, it's always nice to have, like the congratulations, it's a nice accomplishment and you have the impression that you'd be helping the university community, so it's kind of rewarding. (P9)

Fear of consequences: Worries and fears related to not reporting prompt participants to report phishing attempts.

There're serious consequences if a phishing goes through, from a company perspective or on a personal level. (P13)

Sense of belonging: Participants expressed being part of the community prompts them to engage in reporting phishing.

We need to participate. We're all, we're all active users and it's not just IT who has to deal with it. (P32)

We are actors within the community. So, we are together. (P34)

Responsibility: Participants regarded reporting phishing as part of their job and shared the responsibility of reporting.

I see my role as a little more than this reporting, but also trying to reduce all the risk ... we have a duty. And you owe it to your colleagues as well as yourself. (P11)

Peer influence: Participants reported phishing emails because of the influence of their colleagues.

I used to ignore these emails, but then like one of my colleagues told me, it's better to report. So then I started doing it, yeah, but even I don't do it like every time, but most of the time I try to report them. (P21)

Easy to report: Participants mentioned that the positive user experience with the reporting process motivates them to report.

The reporting button is really easy, even if you're in doubt, you tend to click the button. (P13)

Protecting oneself: Participants considered reporting to benefit them in protecting personal accounts, avoiding financial losses, and safeguarding data.

If I never report anything, I can't expect it to just magically get better, so that's why I see a benefit for myself. (P26)

Phishing experience: Participants mentioned their experiences with phishing incidents as a driver for reporting.

I had this scam attack, and I felt bad about myself. I felt bad about trusting the others, so I wouldn't like someone, other people to feel the same way I felt once. (P4)

Empowerment: Participants considered reporting as an initiative against phishing attempts, giving them a sense of control and empowerment.

I had the initiative to defend against the phishing attack. And knowing that I can stop spreading this attack for other people and for my future self. That really helped me, like empowering. (P16)

Satisfaction: Participants expressed their sense of accomplishment/satisfaction for reporting suspicious emails.

I can relate to the sense of satisfaction. Once you've reported it, you feel like you played your role. You did a good job. (P11)

Enjoyment: Participants considered the reporting as a playful game or "nice welcome distraction" from work.

When you click to report phishing attempts, then you receive 'congratulations'. I'm happy and it's like a game. (P28)

Personal Value: Participants reported phishing attempts because it is the right thing to do or the suggestion is good.

It's a very good action to ask us to report suspicious emails. (P6)

Altruism: Participants wanted to help others and vulnerable groups, reducing their chances of being phished.

I want to help others avoid being deceived by phishing. (P15)

Pride: Participants mentioned pride stemming from their ability to consistently identify and avoid being phished.

I don't want to break my streak of always reporting the phishing attacks. I've not clicked on one socially engineered phishing e-mail, I'm quite proud of that. (P8)

Discouraging factors

Perceived low threat: If the participants regarded the incoming phishing emails as too obvious/low threat, they chose not to report.

If I consider the content of phishing emails so apparent, so explicit that everyone can find out that it's phishing, then I don't try to report it. (P16)

Negative outcomes: Assumed negative outcomes from reporting the email discouraged participants.

I feel like there's negative benefits for me reporting them because they don't seem to do anything with it and I just get more emails. So I would get the same amount of spam if I didn't report it. (P17)

Report too much: Participants expressed the concern that they reported too many suspicious emails and burdened the IT team.

It's already the second one I sent this week, so I said, what shall I do? (P28)

Worries of being judged: Participants expressed reservations about reporting suspicious emails due to worries of being judged by the IT team.

If I report Netflix or something as phishing, then they would think 'stupid woman'... This feeling unnerved me and discouraged me from reporting. (P34)

Privacy concerns: Participants expressed they were hesitant about reporting when they felt that it might divulge private information or create a false impression about their personal life.

I worry what they (the IT team) will think of me. So, I try to avoid informing them, because what are they doing with this information? (P28)

Switching between interfaces: Participants mentioned that even they intended to report suspicious emails, they tended to delete or ignore them when checking email on their smartphone.

I use the web client sometimes. I don't know if there is a report phishing on there, and I also don't know if it's on like the iPhone app. (P24)

Unclear procedures: Participants shared that unclear reporting procedures discouraged them from reporting suspicious emails.

I think you should report the suspicious emails, but it needs to be made clearer what suspicious e-mail is and how to properly report it. (P8)

Requiring too much effort: Participants who use Linux and Mac OS expressed that the reporting procedure requires too much effort.

It's too much effort for me, like not much effort, but it's not very easy. (P27)

Lack of feedback: Without follow-up or feedback on their reporting action, participants felt discouraged from reporting.

We don't know what the effectiveness of report-a-phish is. We don't know the numbers, so it would be really good to have a kind of feedback status. What has been done last year? What was the success rate? (P31)

Lack of communication: Participants felt discouraged due to not knowing whether their colleagues reported or not and the organization's status quo for reporting.

I report phishing emails regularly and religiously, but I'm thinking is everyone else doing the same as me, putting in the same effort as I am on reporting? It takes maybe 30 seconds of your time, butt I'm still very careful about it. (P25)

Low response efficacy: When they perceived no impactful results of their reporting, participants felt discouraged and even stop reporting.

If we feel it works, maybe we continue to report, but if it does not work so well, we will not report phishing again. (P1)

Habitual behavior: Participants shared that they often postponed or forgot to report because they reverted back to old habits of simply deleting emails.

Just going back to your old habits because this report phishing button for me is new. And in my other like personal e-mail, Gmail, what I do is delete. So, I might result in just deleting and then other times I might remember. (P11)

Laziness: Participants mention "laziness" as a self-reported reason for not reporting suspicious emails.

I'm able to report them, but sometimes I'm too lazy to report it. (P17)

Low self-efficacy: If they had too high doubts and were not confident about whether it was a phishing attempt or not, participants would not report.

For reporting, I'm not sure because sometimes I am not sure it indeed is a phish or not, so then sometimes, I just prefer to delete it and not to report. (P5)

Simulated or real attack: When simulated phishing tests are overused or not accompanied by a clear protocol, they result in reduced reporting intentions.

For me, every phishing email that I received was a simulated one. So, I didn't see the point of reporting that because I knew that it was from IT. (P27)

Contextual factors: Overload at work, time pressure and stress when they received the email

could discourage them from reporting.

Sometimes when I'm in a rush, I just delete. (P31)

Chapter F

The demographic table

Table F.1: Demographic table of focus groups.

Focus group	**Participant**	**Job title**	**Field**	**Work experience (years)**[a]
FG01	P1	Doctoral researcher	Computer science	1
	P2	Lead software developer	IT	21
	P3	Doctoral researcher	Energy	4
	P4	Doctoral researcher	Robotics	12
	P5	Postdoctoral researcher	Security and cryptography	5
FG02	P6	Doctoral researcher	Psychology	7
	P7	Doctoral researcher	Psychology	2
	P8	Administrative assistant	Administration	2
	P9	Doctoral researcher	Political science and human rights	5
FG03	P10	Doctoral researcher	Neuroscience	5
	P11	Doctoral researcher	Social economics	5
	P12	Postdoctoral researcher	Engineering	8
	P13	Postdoctoral researcher	Digital health	23
	P14	Doctoral researcher	Political sciences	5
	P15	Doctoral researcher	Law	3
	P16	Doctoral researcher	Social sciences	1
FG04	P17	Doctoral researcher	Computer science	5
	P18	Software developer	IT	20
	P19	Doctoral researcher	Computer Science	8
FG05	P20	Administrative assistant	Administration	25
	P21	Doctoral researcher	Supply chain management	2
	P22	Postdoctoral researcher	Security and cryptography	5
	P23	Doctoral researcher	Engineering	3
FG06	P24	Building project manager	Administration	13
	P25	Academic facilitator	Administration	26

	P26	Alumni relations	Administration	34
	P27	Software developer	IT & Admin	23
FG07	P28	Research facilitator	Administration	30
	P29	Data analyst	Administration	7
	P30	Research facilitator	Administration	21
	P31	Project manager	Administration	25
	P32	Research facilitator	Administration	27
	P33	Secretary	Administration	30
	P34	Administrative assistant	Administration	35

[a] Work experience indicates the participants' total years of work experience, including previous jobs. We removed gender, age, and months working at the current organization to avoid re-identification.

Chapter G

Self-efficacy and support-seeking scale

Self-efficacy scale:

Title of scale given to respondents: Self-evaluate confidence

"*Phishing attack is a type of social engineering attack, where attackers send spoofed or deceptive messages to trick a person into revealing sensitive information to the attacker or to deploy malicious software on the recipient's devices.*

Please indicate, to what extent, you agree or disagree with the following statements:"

The following three items are anchored on a 5-point Likert scale, Strongly Disagree (1)/Strongly Agree (5) (from [403]):

1. It would be easy for me to keep up to date with phishing techniques.
2. I am able to keep up to date with phishing techniques.
3. I feel confident in my ability to keep up to date with phishing techniques.

The following four items are anchored on a 7-point Likert scale, Disagree (1)/Agree (7) (from [272]):

1. I am confident I can recognize a suspicious email.
2. I am confident I can recognize suspicious email headers.
3. I am confident I can recognize suspicious email attachment filenames.
4. I can recognize a suspicious email attachment even if there was no one around to help me.

Support-seeking scale (adapted from [147]) :

Title of scale given to respondents: Decision-making in countering phishing

"*Imagine that you have just received a suspicious email in your work account. Please indicate, to what extent, you agree or disagree with the following statements*."

Respondents are presented with four alternatives: "not at all true", "barely true", "somewhat true", and "completely true."

(In scoring responses, 1 is assigned to "not at all true, 2 to "barely true", 3 to "somewhat true" and 4 to "completely true".)

1. When receiving a suspicious email other people's advice can be helpful.

2. I try to talk and explain the suspicious elements of an email in order to get feedback from my colleagues.
3. Information I get from others has often helped me deal with suspicious emails.
4. I can usually identify people who can help me when dealing with suspicious emails.
5. I ask others what they would do when they receive a suspicious email.
6. Talking to others can be really useful because it provides another perspective on properly handling suspicious emails.
7. Before clicking anything within a suspicious email I'll talk with a colleague about it.
8. When I am in doubt of an email I can usually find a solution with the help of others.

Chapter H

Coding scheme for mixed-design experiment

A. Coding system for changes in counter-phishing practices

1. **Check email header**: Participants check email header elements to verify the email.

1.1 Search online/official website: Participants search online for the mentioned organization/stakeholder in suspicious emails to decide whether it's a phishing.

I google the email address that sent the email. (P93)

1.2 Email subject: Participants check the subject/title of the email.

If I'm not familiar with the domain, then I will check the sender name and the email subject to get clarity of the message and purpose. (P11)

1.3 Verify sender: Participants verify the sender's email address, name, and domain to decide whether it is a phish.

I verify the sender's mail-address and electronic signature. (P100)

2. **Evaluate email content**: Participants evaluate the email content to decide the legitimacy of the email.

2.1 Check attachment: Participants check the filename of the attachment.

I verify the extension of any attachment files. (P19)

2.2 Check URL: Participants check the link included in the email.

If I have to click a URL, I check the destination of the link before clicking on it. (P68)

2.3 Analyze the request: Participants check the requests from the incoming email (e.g., content cues).

I read through content and understand what action is required on my end - any personal information request will be a red flag. (P20)

2.4 Theme and content: Participants check the topic/theme/content of the email to decide whether it is legitimate.

My boss has a personal way of writing, so if it's not his style, I will check the email whether it is from him. (P23)

2.5 Expectation and context: Does the email fulfill the participants' expectations and fit their routine/context?

Do I know person? Am I expecting email from that person? Do I expect link or file from that person? (P90)

2.6 Read with caution: Think before clicking and read content carefully before reacting.

Being much more careful about the time I take to read emails and ensuring I check everything. (P55)

2.7 Quality of the text: Grammar, spelling, format, and language of the email.

I check spelling and formatting, graphic design and what is usually done by specific departments. (P15)

3. **Do not respond**: Participants choose not to respond to the email's request.

3.1 Do not click/respond: Participants mention they do not click/respond to the suspicious email.

Don't engage, do not click on any links or images. (P93)

3.2 Delete: The participants indicate that they delete the email.

In general, I delete them immediately. (P85)

4. **Block/Report**: Participants choose to block/report the email.

4.1 Block the sender: Participants mention that they block the sender.

I block the sender. (P21)

4.2 Reporting: Participants only indicated reporting, but did not specify reporting to their organization.

I will report the suspicious emails, then delete them. (P20)

4.3 Report-a-phish: Participants report suspicious emails to the organization's IT team.

I send it as an attachment to report-a-phish to have it checked professionally. (P98)

5. **Interact with Colleagues**: Participants mentioned their colleagues.

5.1 Talk with colleagues: Participants indicate that they talk with their colleagues about the phishing email to make a decision.

I ask colleagues if they received similar emails. (P58)

5.2 Inform my colleagues/friends: Participants mention that they will inform their colleagues/friends of the phishing email.

Not only I will alert my colleagues about any phishing email I may receive, I will also report it to our administrator. (P11)

B. Coding system for usefulness of the training

1. **Phishing knowledge**: Participants mention that they learned about different types of phishing emails and attack techniques during the training.

It was useful to learn about different phishing strategies and experiences from colleagues. (P24)

2. **Skills for safe responses**: Participants mention that they learned how to identify phishing

emails and how to respond to them.

To be more vigilant and looking at various details to identify such attacks and differentiate between which email is legit and which is an attack. (P7)

3. **Enhanced phishing awareness**: Participants mention that they became more aware of phishing attacks and their severity and prevalence after the training.

I realized the threat is serious, and my information could get compromised much easier than I thought. I realized, as an employee of an institute, I'm a target of interest. I used not to take these stuff seriously, always thinking not me, I'm not a celebrity, or rich. I've never realized being an employee here could make me attractive to hackers. Now I know! (P67)

4. **Emphasized reporting**: Participants mention that they will take reporting phishing emails more seriously in the future.

Be more cautious and report phishing emails more consistently. (P37)

5. **Group interaction**: Participants mention that they learned phishing knowledge and skills by discussing with the group.

I found the workshop helpful and informative enough. I mainly enjoyed working in analyzing cases and discussing them with the group. I learned a lot from others' experiences and how they deal with phishing. (P51)

6. **Think like a hacker**: Participants mention that they found the group work of designing phishing emails and thinking like hackers to be useful.

The exercise was very useful for understanding how hackers and scammers use relevant and specific information to attack us - very enjoyable to work with colleagues. (P18)

7. **Interesting/fun**: Participants mention that the training they attended was interesting/fun/enjoyable.

I find this workshop very useful and interesting. (P31)

Chapter I

Chi-square analysis of non-clicking and reporting of each phishing test

Table I.1: Chi-square analysis ($\chi^2(2)$, N = 105) of each phishing test.

	Non-clicking		**Reporting**	
	Value	**Sig.**	**Value**	**Sig.**
Email client upgrade	3.639	.162	6.036	.049
Data breach	.520	.771	15.428	< .001
Security alerts	1.019	.601	10.246	.006

A significantly lower report rate for "Data breach" in the control group ($p < .001$) was observed. Using the right-tailed probability of the chi-squared distribution function in the post hoc analysis, we found:

- A significantly higher report rate for phishing test "Email client upgrade" in group discussion condition (adjusted p = .042).
- A significantly higher report rate for phishing test "Data breach" in role-playing training (adjusted p = .01).
- A significantly higher report rate for phishing test "Security alert" in both group discussion and role-playing training (adjusted p = .003)

We noticed that "Email client upgrade" had the lowest number of reported incidents compared to the other two tests. To ensure data accuracy, we validated the numbers with the security expert responsible for implementing the phishing tests. They confirmed the accuracy of the reporting numbers for "Email client upgrade" and proposed two plausible explanations: firstly, employees may have mistaken the email for spam; secondly, many employees may have been on holiday when they received this email.

Chapter J

Linear Regression Results

Note that for count data, Poisson regression is typically the preferred method of analysis. As a check, we also conducted Zero-inflated Poisson regression (for reporting) and Quasi-Poisson regression (for non-clicking), both of which led to the same conclusions.

Table J.1: Linear regression with non-clicking (sum) as the dependent variable.

Variable	**Estimate**	**Std. Error**	**statistic**	**p.value**
(Intercept)	2.9857	0.2399	12.447	$< 2e-16$
Working month	0.0011	0.0008	1.345	0.182
Gender male	0.0219	0.0794	0.275	0.784
Gender non-binary	0.1759	0.3490	0.504	0.615
Admin	0.1552	0.1384	1.121	0.265
Other faculties	0.0335	0.0903	0.371	0.711
Q3SE	-0.0041	0.0053	-0.775	0.440
Q3SS	-0.0006	0.0066	-0.094	0.926

Table J.2: Linear regression with reporting (sum) as the dependent variable.

Variable	**Estimate**	**Std. Error**	**statistic**	**p.value**
(Intercept)	-0.3275	0.7603	-0.431	0.6677
Working month	0.0037	0.0026	1.432	0.1557
Gender male	-0.0287	0.2516	-0.114	0.9095
Gender non-binary	-0.1548	1.1063	-0.140	0.8890
Admin	0.6438	0.4388	1.467	0.1458
Other faculties	0.1059	0.2861	0.370	0.7121
Q3SE	0.0295	0.0167	1.761	0.0817
Q3SS	-0.0013	0.0210	-0.064	0.9494

Chapter K

Box Plot of Self-Efficacy and Support-seeking

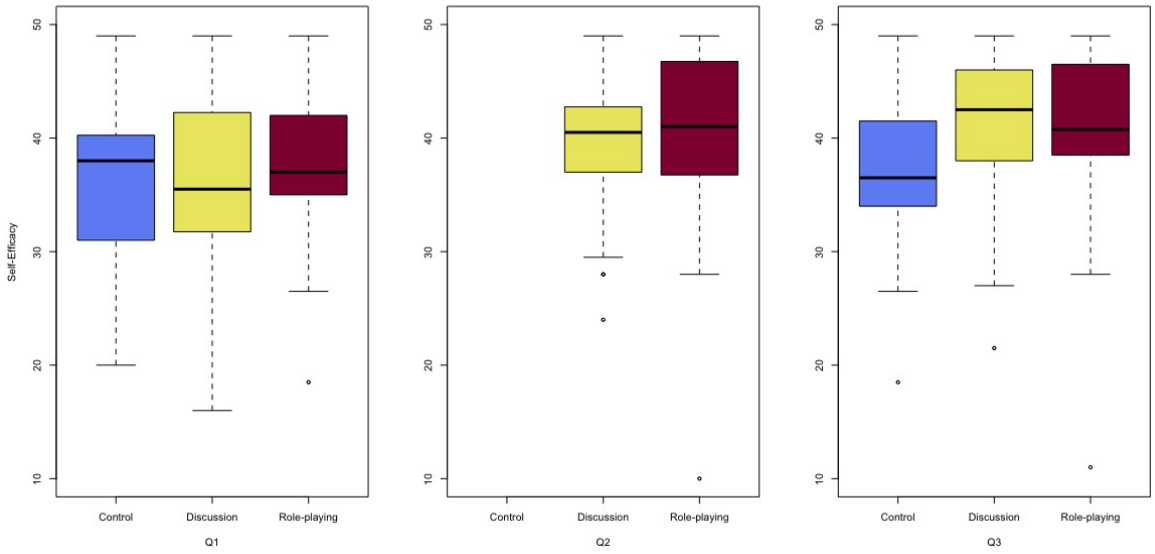

Figure K.1: Box plot of self-efficacy scores.

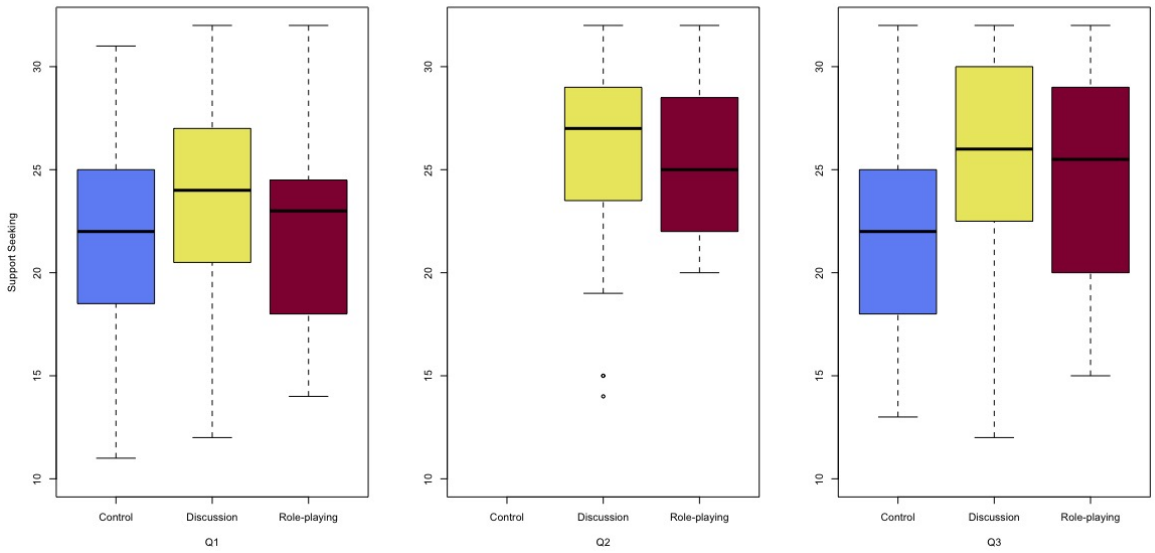

Figure K.2: Box plot of support-seeking scores.

Chapter L

Factor loading for self-efficacy and support-seeking scales

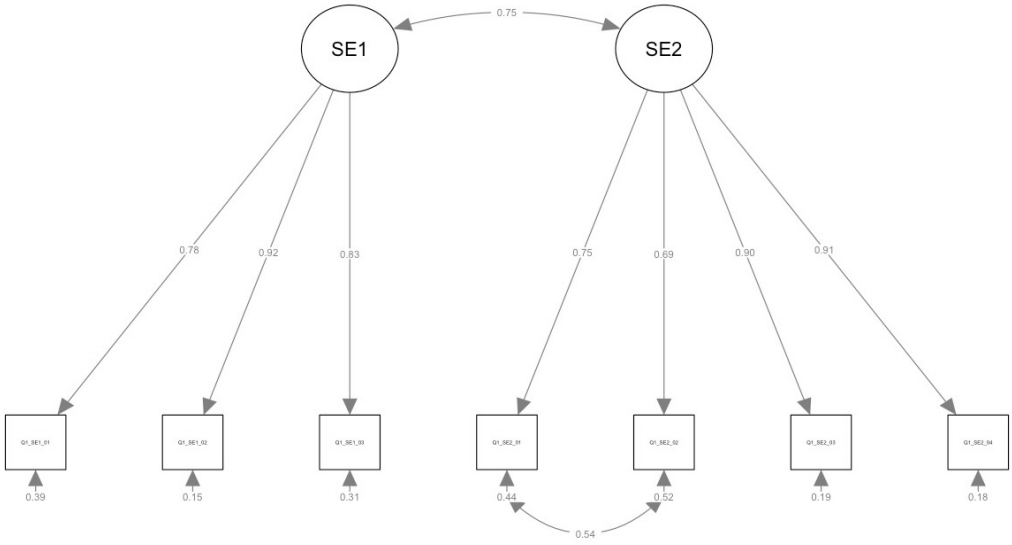

Figure L.1: Factor loading for self-efficacy scale items.

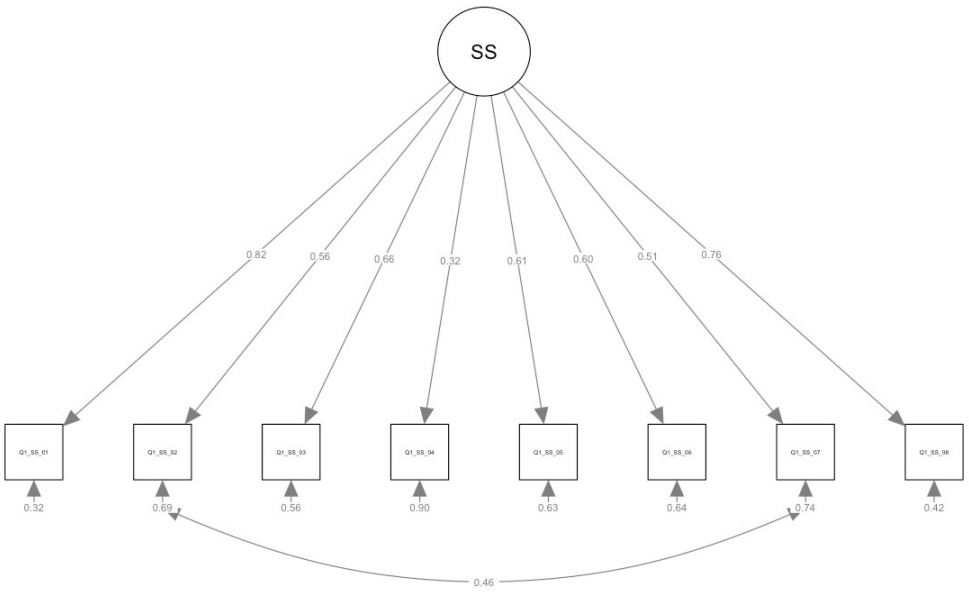

Figure L.2: Factor loading for support-seeking scale items.

Chapter M

Coding scheme and exemplar quotes for formative study

M.0.1 Challenges reported by parents

No answer: when parent left the question empty.

No challenges: when parent indicates they have not encountered challenges or difficulties in teaching security and privacy (S&P) topics.

I didn't encounter any. He is very smart and understands everything, even showing me a few things.

Child - Invincibility/Overconfidence: parent expresses their concern of their child's overconfidence and invincible feeling.

He's young and hasn't really had a bad experience, so he tends to have a sense of invincibility.

Child - Resistance to parent's guidance: parent mentions their child often resist parental advice or rules.

My daughter is always listen what I talk to her but she is never follow the rule.

Child - Lack of awareness/experience: parent indicates their child lack the awareness of online predators and how to handle unwanted contact online.

I think the biggest difficulty is just that he doesn't always understand how dangerous it can be online, so he doesn't always listen carefully.

Parent - Knowledge gap: parent expresses their lack the necessary knowledge to guide their children on S&P topics.

The terms can be technical and abstract. Digital natives are more tech-savvy and confident than previous generations.

Parent - Knowledge update: parent mentions it can be challenging for them to keep update with S&P topics relevant for their child.

I would have to say keeping updated with new and relevant online security and privacy topics especially as it relates to her age groups (teens) may be challenging!

Parent child interaction - Relational related challenge: disconnect between parent and child regarding child's online activities.

I just don't feel close to him in general and it makes discussing things more difficult.

Parent child interaction - Proper way of teaching: parent wants to find a proper way to teach children of the risks, importance, and consequences of their online activities.

Teaching kids about online security is challenging due to complex topics, balancing fear with awareness, understanding privacy settings, resistance to monitoring, and peer pressure to engage in risky behaviors online.

Parent child interaction - Engaging child: parent expresses their difficulty in engaging their child with S&P topics.

Difficult to keep him engaged without him acting like he knew everything already.

Parent child interaction - Parental monitoring: parent faces difficulties in monitoring their children's online behavior, e.g., how to setup parental control and monitoring without being too intrusive.

It's tricky to let kids know that you trust them to make good decisions online but to also let them know - and enforce - that parents have the right and obligation to monitor their online activities for dangers.

Parent child interaction - Communication strategy challenges: parent wants to learn more effective communication strategies, using age-appropriate and relatable language and engaging their child.

When teaching [anonymized] about online security and privacy, I found it challenging to explain complex concepts in a way that's relatable and easy to understand.

Critical reflection: parent reflects on the difficulty in finding a balance between allowing freedom and ensuring safety.

I don't know how far I should go to show them what to watch out for. There is a fine line between teaching them and them tuning you out if you go too far down the rabbit hole.

M.0.2 S&P topics that parents want to learn

Privacy-related topics: when parent refers to privacy protection, including privacy settings, sharing personal info online, and digital footprints.

Security-related topics: when parent mentions topics related to online account or personal device security, including protecting themselves from hacking, phishing, scams, and email/website safety.

Online protective actions: when parent emphasizes online protective actions related to cyberbully, online predators, catfishing, grooming, and identity thief.

Social media platform: when parent indicates social media platforms, such as Discord group chats, social media in general, social media usage.

Teaching strategies: when parent indicates the topics that they would like their child to grasp, such as good online habits, mental mindset, general tips, consequences of oversharing, websites should

avoid, examples of dealing with specific website, online safety rules, creating strong passwords, and how to identify a scam.

Parenting support: when parent mentions that they want more knowledge in parental control and monitoring tools, S&P topics, and how can parents spend online time together with their child.

Emerging technologies: when parent mentions emerging technologies, including deepfake, AI, and AI created content.

Chapter N

Intervention goals for producing the short videos

For Episode 1: a) Explain why security and privacy are relevant to children's online activities in family settings; b) provide easy-to-understand definitions for online security and online privacy, as parents expressed difficulties when trying to explain these concepts to their children; c) provide examples (age-appropriate and relevant) of online security and privacy; and d) emphasize why parents should dedicate more time to updating their knowledge on these topics and supporting their children in family settings.

For Episode 2: We used established educational videos from the Family Online Safety Institute. This episode highlights seven good practices that parents can adopt to support their children (from elementary through middle school) in navigating the digital world.

For Episode 3: a) Introduce various approaches that parents can adopt to monitor their children's online activities, such as using applications, setting rules, and designating spaces; b) provide an overview of parental control functions across different systems, namely iOS, Android, macOS, and Windows; c) discuss the benefits and risks of employing parental controls; and d) emphasize the ultimate goal of parental monitoring—understanding their children's online activities and providing support when needed. Parents should also support their children in developing the skills necessary to manage their digital devices.

For Episode 4: Parents reported difficulty in keeping up with the new applications their children use and the associated risks. This episode aims to: a) provide an overview of the popular applications among American teenagers; b) describe different categories of risks with relevant examples, such as account security, cyberbullying, online privacy invasion, exposure to harmful content, online safety (including risks from predators and online grooming), and digital wellbeing; and c) emphasize that parents should stay informed, set boundaries, and engage in open discussions with their children about these risks.

For Episode 5: Parents indicated that they wanted to learn more about parenting strategies. Accordingly, we introduced various conversation strategies that parents can adopt to discuss security and privacy topics, as informed by Alghythee et al. [10]. In previous episodes, we introduced rule-based and example-based conversations. In this episode, we focus on sharing decision-making thought processes, consequence-based conversations, and contextual conversations.

For Episode 6: We describe the current trends and emerging risks that parents are interested in learning about. We focused on the three topics that parents specifically mentioned in the qualitative survey: the risks associated with generative AI usage (as informed by [410]), deepfakes, and sextortion

targeting teenagers. Additionally, we provided a recap of the key topics discussed in the previous five episodes.

Chapter O

Measurement items

Parental security awareness:

Description provided to respondents: How much do you agree with the following statements?

The following five items are anchored on a 5-point Likert scale, 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree:

- CSA1 I am aware of the potential of online threats.
- CSA2 I am aware of what {Nickname} is accessing.
- CSA3 I know that {Nickname} knows how to use the Internet the right way.
- CSA4 I'm a role model for {Nickname} in helping them use mobile devices and the Internet positively.
- CSA5 I exercise self-discipline when using mobile devices and the Internet.

Parental concerns about S&P of their children: Description provided to respondents: How concerned are you about the following potential online risks for your child?

The following 12 items are anchored on a 5-point Likert scale, 1 = Not at all, 2 = Slightly, 3 = Moderately, 4 = Quite a bit, 5 = Very much:

- PC1 Being called insulting names.
- PC2 Being contacted by strangers.
- PC3 Being socially excluded.
- PC4 Receiving repeated unwanted messages.
- PC5 Having lies or rumours spread about them.
- PC6 Having pornography shown/sent to them.
- PC7 Having their accounts accessed without their permission.
- PC8 Being shown or sent violent or racist content.
- PC9 Having someone pretend to be them online.
- PC10 Receiving threats.

- PC11 Having their personal information used in ways they don't like.
- PC12 Having inappropriate photos of them posted without their consent.
- PC13 Having distressing information about them disclosed to others.

Internet-specific parenting self-efficacy:

Description provided to respondents:How confident do you feel in your ability to prevent your child from ...

The following eight items are anchored on a 5-point Likert scale, 1 = extremely unconfident, 2 = not confident, 3 = neutral, 4 = confident, 5 = extremely confident:

- SEFF1 coming in contact with dangerous persons?
- SEFF2 being bullied?
- SEFF3 coming in contact with inaccurate information?
- SEFF4 coming in contact with material that will make him/her upset?
- SEFF5 ending up on a website with pornographic content?
- SEFF6 ending up on a website with violent/gory pictures?
- SEFF7 ending up on a website that has hateful content against individuals or groups?
- SEFF8 giving out or posting personal information that could be problematic for safety reasons?

Internet-specific parental mediation:

Description provided to respondents: How often do you ...

The following 12 items are anchored on a 5-point Likert scale, 1 = never, 2 = seldom, 3 = sometimes, 4 = often, 5 = always:

Subscale: Restrictive monitoring:

- RM1 filter software installed on {Nickname}'s devices?
- RM2 monitor the time {Nickname} spends on the Internet?
- RM3 check what {Nickname} has done on the Internet?
- RM4 track the websites that {Nickname} has visited?
- RM5 check the {Nickname}'s messages (e.g., email, Facebook, texts)?

Subscale: Demand disclosure:

- DD1 demand to know which websites {Nickname} has visited?
- DD2 demand to know whom {Nickname} chats with?

Subscale: Active mediation:

- AM1 talk to {Nickname} about what he/she is doing on the Internet?
- AM2 talk to {Nickname} about potential risks that he/she can encounter on the Internet?

Subscale: Proximity

- P1 sit with {Nickname} when {Nickname} is online?
- P2 stay nearby when {Nickname} is online?
- P3 watch the screen when {Nickname} is online?

Parental conversation approaches on S&P:

Description provided to respondents: How much do you agree with the following statements?

The following 12 items are anchored on a 5-point Likert scale, 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree:

Subscale: Rule-based conversations

- RBC1 I regularly discuss online safety rules with my child.
- RBC2 The rules I set for my child about online behavior are specific.
- RBC3 I stress the importance of applying online safety rules across all different online platforms and situations.
- RBC4 I regularly reinforce online safety rules to my child.

Subscale: Example-based conversations

- EBC1 I show my child phishing emails I receive to help them recognize online scams.
- EBC2 I show my child how to do things like making strong passwords to teach them about online safety.
- EBC3 I show my child examples of fake emails or scams to illustrate how online deception works.

- EBC4 I believe showing examples is more effective than just telling my child about online safety and privacy.

Subscale: Decision-making thought process:

- DMTP1 I explain to my child how I make choices about keeping our online information safe.
- DMTP2 I walk my child through how I check if a website is secure before entering personal information.
- DMTP3 I discuss my reasoning with my child when deciding whether to share information online.
- DMTP4 I talk to my child about my steps of addressing online concerns.

Subscale: Consequence-based conversations:

- CBC1 I actively discuss with my child the potential risks and consequences of their online actions.
- CBC2 I emphasize to my child that using a weak password can lead to account hacked.
- CBC3 I tell my child the consequences of over-sharing information online.
- CBC4 I explain to my child the dangers of sharing personal information even in private online groups.

Subscale: Contextual conversations:

- CC1 I discuss privacy issues with specific apps and platforms my child uses.
- CC2 I set clear rules for my child's use of specific platforms and apps.
- CC3 I explain to my child the different privacy settings of different apps.
- CC4 I talk with my child about how specific app or website collects data.

Online privacy concerns: Description provided to respondents: For this part of the survey, we are interested in any concerns you might have when online. Please answer every question using the full scale provided.

The following five items are anchored on a 5-point Likert scale, 1 = Not at all, 2 = Slightly, 3 = Moderately, 4 = Quite a bit, 5 = Very much:

- PRIV1 Are you concerned about online organisations not being who they claim they are?

- PRIV2 Are you concerned about online identity theft?
- PRIV3 Are you concerned about people online not being who they say they are?
- PRIV4 Are you concerned that if you use your credit card to buy something on the internet your credit card number will obtained by someone else?
- PRIV5 In general, how concerned are you about your privacy while you are using the internet?

Chapter P

Dropout rate of the intervention group

Table P.1: Dropout rate between consecutive contact points.

From → To	Participants (n)	Dropout (n)	Dropout Rate (%)
Invitation → Episode 1	185 → 152	33	17.84%
Episode 1 → Episode 2	152 → 134	18	11.84%
Episode 2 → Episode 3	134 → 114	20	14.93%
Episode 3 → Episode 4	114 → 101	13	11.40%
Episode 4 → Episode 5	101 → 93	8	7.92%
Episode 5 → Episode 6	93 → 88	5	5.38%

Chapter Q

General linear regression analysis of each target outcome

Table Q.1: General linear regression analysis of parental awareness at T_2 as a function of included predictors. Note: *B* for unstandardized coefficient; *SE* for standard error.

Predictor	***B***	***SE***	***t***
(Intercept)	1.38	0.29	4.8***
Awareness_T_1	0.48	0.06	7.62***
Security attitude_T_1	0.21	0.04	4.73***
Privacy concerns_T_1	0.01	0.03	0.29
IG	1.38	0.42	3.26**
Applied	0.02	0.09	0.21
Sex	0.04	0.06	0.65
Age	-0.002	0.10	-3.11**
Awareness_T_1:IG	-0.31	0.10	-3.11**

Table Q.2: General linear regression analysis of parenting self-efficacy at T_2 as a function of included predictors.

Predictor	***B***	***SE***	***t***
(Intercept)	1.24	0.43	2.89**
Self-efficacy_T_1	0.59	0.08	7.12***
Security attitude_T_1	0.13	0.08	1.75
Privacy concerns_T_1	0.00	0.06	-0.07
IG	1.18	0.46	2.53*
Applied	-0.14	0.16	-0.88
Sex	-0.01	0.10	-0.07
Age	-0.01	0.01	-1.63
Self-efficacy_T_1:IG	-0.25	0.13	-1.94

Table Q.3: General linear regression analysis of consequence-based conversations (CBC) at T_2 as a function of included predictors.

Predictor	***B***	***SE***	***t***
(Intercept)	1.09	0.30	3.66***
CBC_T_1	0.55	0.06	9.34***
Security attitude_T_1	0.22	0.06	3.67***
Privacy concerns_T_1	0.13	0.04	0.73
IG	0.82	0.37	2.22*
Applied	0.07	0.11	0.64
Sex	0.04	0.07	0.53
Age	0.00	0.00	-0.92
CBC_T_1:IG	-0.19	0.09	-2.22*

Table Q.4: General linear regression analysis of decision-making thought processes (DMTP) at T_2 as a function of included predictors.

Predictor	***B***	***SE***	***t***
(Intercept)	0.97	0.32	3.00**
DMTP_T_1	0.57	0.07	8.69***
Security attitude_T_1	0.21	0.07	2.89**
Privacy concerns_T_1	0.04	0.04	0.95
IG	0.87	0.35	2.53*
Applied	0.14	0.12	1.17
Sex	-0.02	0.08	-0.30
Age	-0.01	0.00	-1.26
DMTP_T_1:IG	-0.20	0.08	-2.41*