International Symposium on Technikpsychologie (TecPsy) 2023

DOI: 10.2478/9788366675896-fm

Proceedings TecPsy 2023

International Symposium on Technikpsychologie – celebrating 100 years of interdisciplinary research at TU Darmstadt

Joachim Vogt (Symposium Chair)

N. Gereber & V. Zimmermann (Eds.)

15.02.2023
Darmstadt, Germany

DOI: 10.2478/9788366675896-toc

sciendo

Proceedings of the International Symposium on Technikpsychologie

Contents

DOI: 10.2478/9788366675896-001 • Technikpsychologie • 1-13

Proceedings of the International Symposium on Technikpsychologie

How do you Feel about Cybersecurity? – A Literature Review on Emotions in Cybersecurity

Alexandra von Preuschen[1,†], Verena Zimmermann[2], Monika C. Schuhmacher[1]

[1]*ETH Zürich Professorship for Security, Privacy & Society D-GESS, Haldeneggsteig 48006 Zürich, Schweiz*
[2]*Justus-Liebig-Universität Giessen Professorship BWL X Economics and Business Studies, Licher Straße 6235394, Gießen*

Abstract

Despite the relevance of emotions in explaining human behavior, emotions have traditionally been neglected in models explaining cybersecurity behavior. In recent years, however, the consideration of emotions has gained increasing interest. To enrich our understanding of emotions in the field of cybersecurity, this systematic review analyses 23 papers in terms of their emotion-related focus based on a categorization system derived from cybersecurity literature. The findings indicate that emotions are fundamental in a wide variety of cybersecurity processes and require further research. The analysis furthermore reveals five essential challenges to emotion research in cybersecurity, e.g., with regards to the conceptualization of emotions and the tools to measure them within the cybersecurity context. Finally, the articles provide recommendations for possible research avenues to address these posed challenges.

Keywords

cybersecurity • emotion • affect • literature review

1. Introduction

In today's digital world, protecting the cyberspace from malicious attacks is more important than ever. According to a risk assessment, cyber incidents are considered the greatest business risk, well before natural disasters and the outbreak of a pandemic (Allianz, 2022). This assessment is not unfounded: caused by a cyber incident, an average cost of USD 4.35 million will be incurred in 2022, an increase of 12.3% compared to 2020 (IBM, 2022). In up to 95 % of these incidents, human error is named as a major contributor ((ISC)², 2020; ENISA, 2019; IBM Global Technology Services, 2014). Consequently, humans are often referred to as the 'weakest link' (Schneier, 2004) and, in response, are largely excluded from the system (e.g., severe restriction of the assignment of permissions or intensified focus on technical or physical controls within the defense-in-depth strategy), or restricted by extensive security policies (e.g., strengthen administrative controls without considering the influence of human factors Cisco, 2018; Siponen et al., 2010). While most cybersecurity attacks use well known methods and approaches, users often do not take protective measures, even if they are aware of the threat (IBM, 2022; Kok et al., 2020).

In order to meet security objectives, employees are obliged to comply with security processes and to regularly learn about security (McIlwraith, 2021). Cybersecurity education and behavior, however, often proves to be time-consuming and eventually results in an overload through additional security tasks or employees finding workarounds for existing security-usability trade-offs (Beautement et al., 2008; Kirlappos et al., 2014), such as circumventing regular password change through adding an easy-to-guess ascending number to the same password rather than creating new ones.

According to the current state of research, security behavior is largely explained on the basis of cognitive models such as protection motivation theory, theory of planned behavior or the General Deterrence Theory (Bulgurcu et al., 2010; D'Arcy & Hovav, 2009; Johnston & Warkentin, 2010). Yet, this often leads to inconsistent results that cannot explain the gap between knowledge and behavior.

However, this approach considers only half of the picture. In the view of Dual Process Theories, two systems of thinking are prevalent. System 1 is described by a fast and emotional process whereas system 2 is characterized by rationality

[†]*Corresponding author*: *Alexandra von Preuschen*
E-mail: Alexandra.vonPreuschen@ wirtschaft.uni-giessen.de

and a high cognitive load. Surprisingly, only System 2 is considered dominantly in previous research: a rationally guided process of high persistence. However, cybersecurity behavior is a secondary task and concerns numerous, quick decisions every day, such as decisions to click on links in e-mails or to provide sensitive information on websites. Therefore, focusing on the rational mode of thinking might not be sufficient to explain users' cybersecurity behaviors. System 1 in contrast deals with quick, automatic decisions based on heuristics, biases or emotional tendencies (Kahneman, 2012), that match the context of cybersecurity. Thus, System 1 information processing including the influence of emotions should be stronger considered in cybersecurity research.

The concept of emotions

The concept of emotions is often oversimplified. Terms such as affect, mood, feelings and emotions are used interchangeably (Ekman & Davidson, 1994; Frijda, 1993; Schwarz & Clore, 2007). In order to understand the meaning of emotions in the field of cybersecurity, however, a clear differentiation is essential, as the different terms take on significant importance both on the causes and consequences (Martin et al., 1993). In general, affect is often used as an umbrella term for moods and emotions, which takes on either positive or negative valence and varies in its degree of arousal (Clore et al., 1994; George, 1996). More specifically, affect can be defined as the presubjective and nonconscious encounter of continuous varying intensities, describing the body's readiness for interaction (Massumi, 1995).

Emotions refer to short-lived and relatively intense emotional experiences, which can either describe a general tendency of a person to an emotional sensation, similar to personality (preexisting propensity within a person; 'trait') or are felt over short periods ('state'; Isen, 1984). Feelings, in contrast, are exclusively mental and describe sensations (including e.g., tactile sensations such as coldness) that are reviewed against previous experiences (Lutz, 1988; Shouse, 2005). Emotions display those feelings and are put eventually into a social context (Ekman, 1971; Shouse, 2005). Moreover, emotions can be classified at the level of object-relatedness. Emotions that have a direct relation to a stimulus, decision, or perception are considered *integral* (e.g., anticipated regret when deciding whether to ask for a promotion), while emotions that are perceived regardless of the stimulus at the time of the decision are termed *incidental* (e.g., the feeling caused by a frustrating work event on the decision which movie to watch in the evening; Lerner & Keltner, 2000). A mood could result from this short-lived emotional experience, if it is maintained over a longer period of time. Hence, moods are characterized by prolonged persistence and mild emotional intensity (Isen, 1984). As a result, emotions, in contrast to moods or the global term affect, are characterized by a high informative content for the actor and significantly influence perception, decisions, and behavior (Martin et al., 1993).

Emotions in cybersecurity research

Emotions and other affective states are, regardless of the definition, generally often neglected in the field of cybersecurity research. Although in other domains the relevance of emotions in perception, learning, and decision-making behavior has already been fundamentally established (Estrada et al., 1994; Han et al., 2007; McConnell & Eva, 2012; Pekrun et al., 2011). For instance, in medical education, it was demonstrated that emotions influence the perception of information and, therefore, the availability of the information when required (McConnell & Eva, 2012). In general, various studies show that human cognitive processes such as attention, learning, and memory, judgment, or problem-solving are consciously or unconsciously influenced by emotions (Brosch et al., 2013). The affect-as-information hypothesis postulates that affective feelings, including emotions, are used as a source of information and that these are incorporated into the overall assessment of the object of judgment (Clore et al., 2001; Schwarz, 1990). According to affect heuristics, it is argued that the unconscious perception of affect offers a more important role than concrete information in decision-making behavior in order to be able to make quick decisions (Slovic et al., 2007). This heuristic can be regarded as the basis of System 1 of dual process theory according to Kahneman (2011). Therefore, information processing takes place either quickly, effortlessly, and emotion-based (System 1) or slows down, based on a deliberate and logical approach (System 2; Kahneman, 2012). Further studies generally argue that emotions, since they give a motivational impetus, have a direct influence on behavior (Lazarus, 1991; Lerner & Keltner, 2000).

Apart from the central influence of emotions on behavior, the peripheral effect of emotions also becomes apparent when looking at common theories that look at behaviors within cybersecurity. Theories such as the Theory of Planned Behavior, the COM-B model, the Protection Motivation Theory, Fogg Behavior Model, or the Knowledge Attitude Behavior Model are characterized by variables such as self-efficacy (Bandura, 1977), motivation (Reeve, 2018), or attitude (Ostrom, 1969), which in themselves are centrally influenced by the effect of emotions. This becomes evident, for example, when looking at the relationship between self-efficacy and emotions: stimuli that induce fear lead to a lower self-efficacy expectation and thus to fear of the fear-inducing situation and avoidance behavior (Bandura, 1977).

The consideration of these two routes of influence is essential since emotions accordingly not only direct behaviors on the central route but also determine learning and retrieval,

simple approach and avoidance processes, as they influence important variables to explain secure behaviors, for instance attitudes. At the same time, humans are faced with new challenges in the digital context. Known triggers and behavioral tendencies have not yet been established.

The focus of this paper is of a psychological nature and aims to apply findings from emotion research to the context of cybersecurity and thereby provide insight into how these findings influence cybersecurity aspects. Therefore, this paper aims to highlight the importance of considering emotions in the context of cybersecurity based on an overview of the current state of research by first, identifying researched emotions, and providing used research methodologies. Second, reviewing the focus of the examined emotions and, third, analyze how emotions influence cybersecurity behavior. Furthermore, this work describes the challenges of emotion research in the field of cybersecurity and gives a brief outlook on possible research avenues.

2. Literature Review on Emotions in the Field of Cybersecurity

Emotions and emotion-based mechanisms play a powerful and central role in dealing with IT systems (P. Zhang, 2013). On various occasions, they are attributed the highest importance to the explanation of behavior. The following section looks at the current state of emotion research in the field of cybersecurity. In particular the role of emotions for explaining cybersecurity behavior is considered. A systematic literature search was conducted that included the use of multiple databases to capture the interdisciplinary character of the research.

Procedure

First, EBSCOhost including the databases APA PsycArticles, APA PsycInfo, Business Source Premier, PSYNDEX Literature with PSYN-DEX Tests, ACM Digital Library and IEEE Xplore Digital Library were searched. In order to be able to include various definitions in the search, the search terms have been diversified to a great extent. The search terms applied on titles and abstracts were („emotion*“ or „affect“ or „mood“) and („cybersecurity“ or „cyber security“ or „internet security“ or „cyber-security“ or „IT security“). This resulted in a total of 42 papers. Second, abstracts of these papers were reviewed and those were excluded that either considered emotions only incidentally (e.g., as a casual remark in an interview, but not elaborated upon) or as a trait (i.e., personality), or focused on related constructs (e.g., stress), or did not differentiate clearly between stress and emotions. Additionally, a forward and backward analysis was carried out for articles that matched the given criteria.

Finally, due to the exclusion criteria, this approach resulted in a final sample of 23 papers which differ in their focus regarding the researched emotions and measurements, the focus of the examined emotions and the effects of the emotions on cybersecurity behavior. See Table 1 for an overview of the papers and Figure 1 for the article selection process based on the PRISMA statement (Page et al., 2021).

Researched Emotions and Measurements

The studies can generally be divided into approaches that explore the existence of affect (e.g., Conrad et al., 2020; van Schaik et al., 2020) or emotions (e.g. Buck et al., 2018; Fagan et al., 2017), which consider explicit emotions based on selected frameworks and theories (e.g., Burns et al., 2019; Cheung-Blunden et al., 2019) or which illuminate complex constructs based on emotions (e.g., Cram et al., 2021; Pham et al., 2019). In the former, the exploration of the meaning of affect, valence and arousal (e.g., Conrad et al., 2020; van Schaik et al., 2020) or exclusively valence (e.g., Beris et al., 2015; Gulenko, 2014) are usually considered. The measurement proves to be diverse. For example, Beris et al. (2015) use interview data as a basis, Conrad et al. (2020) use a combination of EEG data and behavioral data based on the International Affective Picture System, while Kok et al. (2020) make use of questionnaire data.

Within these studies, it is demonstrated that affect plays a central role in the field of cybersecurity. For example, van Schaik et al. (2020) demonstrates an effect of affect heuristics on risk perception and Conrad et al. (2020) that notifications cause negative affect during Internet browsing, regardless of their communication style.

On the other hand, as a measurement methodology in the exploration of the existence of emotions, perceptions are usually selected from a list of given emotion terms (e.g., Buck et al., 2018; Fagan et al., 2017). Three studies, on the other hand, chose a qualitative-exploratory approach and thus did not specify a selection of emotions (Budimir et al., 2021; Menges et al., 2022; Renaud, Zimmermann, et al., 2021). However, it becomes clear that a pure consideration of affect based on arousal and valence cannot sufficiently cover the diversity of emotions. For example, Fagan et al. (2017) identified 45 relevant emotions in dealing with password managers or Budimir et al. (2021) 75 emotions triggered by a cybersecurity incident. Regardless of the methodology, however, emotions of positive and negative valence are discovered.

Within research on explicit emotions, nonetheless, there is a clear trend towards negative valenced emotions. In particular, fear (e.g., Abroshan et al., 2021; Cheung-Blunden et al., 2019; Johnston & Warkentin, 2010; X. A. Zhang & Borden, 2020), anxiety (e.g., Abroshan et al., 2021; Bachura et al., 2022; Burns et al., 2019; Cheung-Blunden et al., 2019) and sadness (e.g., Bachura et al., 2022; Beris et al., 2015; X. A.

Table 1: Overview of selected papers

Phase	Context	Subcontext	Author	Emotion(s)	Emotion Focus
Pre-Incident	Precaution-taking	General	Burns et al., 2019	Happiness, Interest, Sadness, Anxiety	Integral
		General	Kok et al., 2020	Affect	integral
		General	Liang et al., 2019	Inward/ outward emotion-focused coping	integral
		General	van Schaik et al., 2020	Affect (valence & arousal)	integral
		Internet browsing	Conrad et al., 2020	Affect (Valence & Arousal)	Elicited, integral
		Password managers	Fagan et al., 2017	Emotional state	integral
		Policies	Beris et al., 2015	Affect (Valence)	integral
		Pop-up Warnings	Buck et al., 2018	Discrete emotions	Anticipated, integral
	Awareness, education & communication	Anti-spyware use	Johnston & Warkentin, 2010	Fear	Induced integral (Fear appeals)
		Password	Dupuis et al., 2021	Fear	Induced integral (Fear appeals)
		Password	Dupuis et al., 2022	Fear	Induced integral (Fear appeals)
		Password	Gulenko, 2014	Valence	Induced integral
		Various contexts	Renaud & Dupuis, 2019	Fear	Induced integral (Fear appeals)
		Not specified	X. A. Zhang & Borden, 2020	Fear & Sadness	Induced integral (Fear & anxiety appeals)
Incident	Emotion as a target	COVID-19 Phishing	Abroshan et al., 2021	Fear & Anxiety	Integral (phishing content) & incidental
	Emotions caused by breaches	Office of Personnel Management (OPM) data breach of 2015	Bachura et al., 2022	Anxiety, Anger, Sadness	integral
		General	Budimir et al., 2021	Emotions	integral
Post-Incident	Emotional coping	Self-caused incidents	Renaud, Searle, & Dupuis, 2021	Shame and guilt	integral
Global	Global term „Cybersecurity“	-	Cheung-Blunden et al., 2019	State anxiety & fear	incidental
		-	Cram et al., 2021	Fatigue	Complex construct
		-	Pham et al., 2019	Burnout	Complex construct
		-	Renaud, Zimmermann, et al., 2021	Emotions	integral
		Security experts	Menges et al., 2022	Dysfunctional relationships (negative emotions)	integral

Zhang & Borden, 2020) are considered, whereby it should be emphasized that only a few studies clearly differentiate between fear and anxiety (e.g., Abroshan et al., 2021; Cheung-Blunden et al., 2019). Furthermore, only one study examines the explicit mode of action of positive emotions (happiness and interest). Ultimately, this focus on specific emotions, as well as individual selected emotions, risks neglecting the complexity as well as diversity of emotions.

Despite complexity being taken into account within the consideration of complex structures, diversity of emotional valence is still neglected. Cram et al. (2021) consider, for example, cybersecurity fatigue, a socio-emotional state felt by employees who are tired of their company's security policies and thus show the result of an intensive interplay of different emotions, but do not consider emotions explicitly. Interview data (Cram et al., 2021) as well as questionnaires are used as measurement methods of these complex constructs (Pham et al., 2019).

Focus of the Examined Emotions

The present studies, both, can be divided into their respective focuses, emotions, and cybersecurity. At the level of emotions, it can first be considered whether emotions are incidental (emotions, regardless of the given decision/perception) or integral (emotions, relevant to the given decision/perception). Only one study considers incidental emotions (Liang et al., 2019), all other studies consider emotions that refer to a

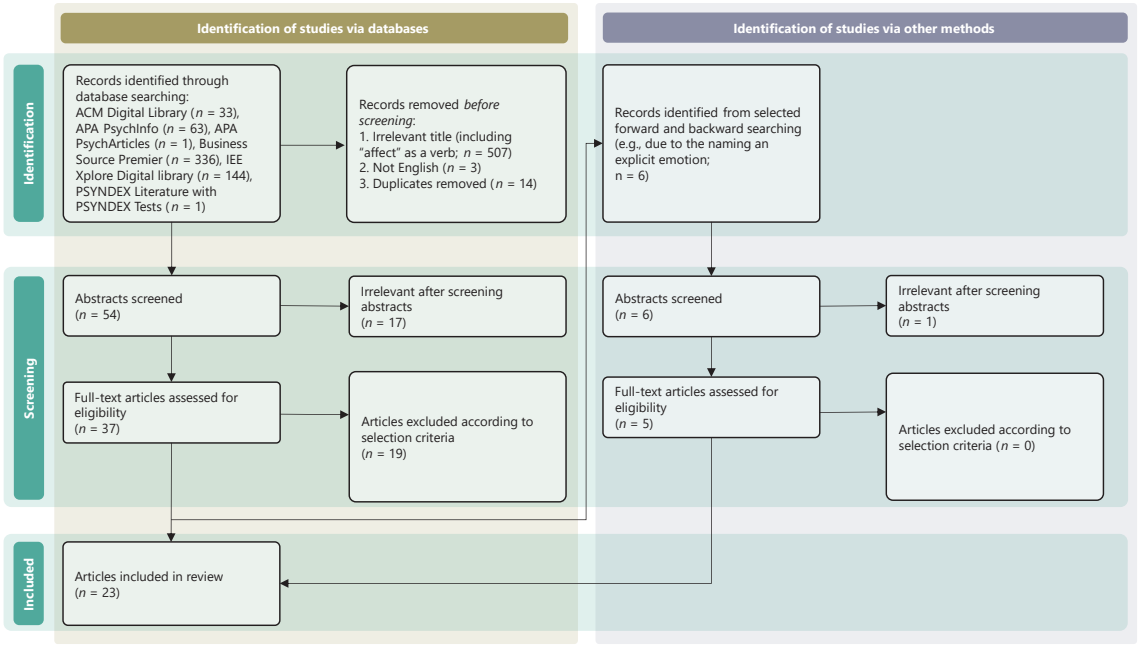


Figure 1. Flow diagram of the article selection process.

specific object/subject (cybersecurity sub-areas (e.g., Burns et al., 2019; Fagan et al., 2017; Kok et al., 2020), cybersecurity global (e.g., Renaud, Zimmermann, et al., 2021), oneself (e.g. Buck et al., 2018; Renaud, Searle, & Dupuis, 2021), others (e.g., Bachura et al., 2022; Buck et al., 2018). Furthermore, it is important to distinguish between, on the one hand, the study of emotions that are induced and, on the other hand, existing emotional connections. For example, studies on awareness, education & communication only consider induced emotions, for example via fear appeals (e.g., Dupuis et al., 2021; Dupuis et al., 2022; Johnston & Warkentin, 2010; Renaud & Dupuis, 2019).

In particular, a further subdivision of the processes of cybersecurity proves to be necessary in order to grasp the complexity of the term. Here, cybersecurity is defined as a process of protecting the cyberspace by preventing, detecting, and responding to attacks. A security incident in this context is defined as an undesired violation of a canonical security objective (confidentiality, integrity, availability Böhme et al., 2018). Therefore, cybersecurity can be divided in three essential phases (Barrett, 2018):

1. *Pre-incident*: Actions, including education, to prevent an incident
2. *Incident*: Detecting an attack and taking action to bring it under control
3. *Post-incident*: Actions to minimize the impact of an incident, restore or modify the normal state and general handling of an incident within a company
4. *Global*: An additional fourth category „global" was added for processes which occur along all phases of cybersecurity, such as general communication or emotions toward the general term of cybersecurity.

In many cases, this distinction on the level of cybersecurity and emotion focus allows conclusions to be drawn about the antecedents of emotions and thus gives indications of how processes can be improved. Although interestingly similar emotions are investigated in many subdivisions (fear, anxiety, sadness, anger, etc.), it becomes clear, when looking at the studies, that the reference of emotions has a significant influence on which emotion is felt and what tendency to act results from it. Renaud, Zimmermann, et al. (2021) discovered that looking at emotions related to cybersecurity is non-trivial. In addition to positive emotions such as feeling secure, negative emotions such as uncertainty, anxiety, anger or overwhelm are primarily mentioned. However, when looking at the approach and avoidance behavior (by means of a push-pull task) towards selected terms of cybersecurity, inconsistent results are found, which suggests that the emotions vary across different areas (Renaud, Zimmermann, et al., 2021). Based on this result, the next sections will describe the results of the literature analysis with regards to the defined classification. Table 1 also provides a summary of the articles assigned to each phase and the emotional aspects analyzed in each article. In addition, the table gives an indication of the reference of the studied emotions to cybersecurity as described above.

Effects of Emotions on Cybersecurity Behavior

Pre-Incident

Precaution taking. When considering general affect, a significant effect on cybersecurity behavior is apparent, which, however, cannot be explained exclusively by valence. For example, Kok et al. (2020) demonstrate that the affective component of attitudes have a positive relationship to behavioral intention of preventive behavioral measures, even higher than the cognitive component or knowledge. Beris et al. (2015), on the other hand, show that the effect of valence cannot be generalized. The authors identify 16 types of behavior based on risk understanding and valence to security policies. Positive emotions, regardless of the understanding of risk, lead to ambivalent types of behavior, while negative emotions result in undesirable behaviors (e.g., avoidance, conscious misbehavior or shadow security; Beris et al., 2015). Looking at emotions in a more nuanced form gives a more accurate indication of these inconsistent outcomes. In line with Beris et al. (2015), Burns et al. (2019) discover opposite behavioral tendencies for positive valenced emotions. For example, interest, classified as a high-activation emotion, results in an increase of psychological capabilities and ultimately in desirable behavior, whereas happiness, classified as a low-activation emotion, results in psychological distancing, indicating that employees who are satisfied with the current situation might feel like no further intervention is required. As with previous findings on affect, negative emotions generally lead to undesirable behavior such as psychological distancing or avoidance (Burns et al., 2019; Fagan et al., 2017) and decreased problem-focused coping (Liang et al., 2019).

Awareness, Education & Communication. Studies in this section deal primarily with research that considers measures or communication that intend to modify behavior through emotion-inducing incentives. The studies demonstrate inconsistent results for the use of fear appeals. Considering contextual conditions (e.g., incentives to increase self-efficacy or intensity of threat) and individual characteristics (e.g., personality), fear appeals can indeed provide desirable behavioral tendencies (e.g., Dupuis et al., 2021; Johnston & Warkentin, 2010; Renaud & Dupuis, 2019; X. A. Zhang & Borden, 2020). Nonetheless, it should be considered that triggering another emotion, for example, sadness instead of fear can result in avoidance behavior instead of compliance (X. A. Zhang & Borden, 2020). In addition, cybersecurity fear appeals are likely to be presented in everyday work along with various other fear appeals (e.g., job loss as a result of poor job performance), leading to emotional overload. This overload, i.e., fear fatigue, can also lead to undesirable behavioral tendencies (Renaud & Dupuis, 2019). Negative emotions can also cause negative side effects such as reduced well-being or job satisfaction, which can have a general negative spillover effect on cybersecurity in general (Dupuis et al., 2022; Renaud & Dupuis, 2019). Positive emotions, however, can lead to a greater extent to security behaviors as opposed to measures that serve only as deterrents (Gulenko, 2014).

Incident

Emotion as a Target. In the field of social engineering, the importance of emotions is becoming increasingly important in the current time. Abroshan et al. (2021) conducted a study in which they examined the susceptibility to phishing emails based on emotions to Covid-19. Participants with a high fear of Covid-19 were more likely to click on Covid-19-specific phishing emails, but not on common phishing emails. A general high level of anxiety (triggered by Covid-19), in contrast, led to a higher phishing susceptibility to both Covid-19 and common phishing emails. This study shows the importance of dealing with emotions in specific areas in order to be able to address them in a targeted manner and to provide employees with strategies for emotion regulation (Abroshan et al., 2021).

Emotions caused by Breaches. If an event is emotionally arousing regardless of its valence, the likelihood that it will be remembered is higher. Depending on the valence or the explicit emotions, however, it is how the event is processed and integrated into learning processes (Williams et al., 2022). Studies demonstrate that the valence of an emotion caused by a cybersecurity incident provides information about the processing (solution-oriented strategies opposed to attack and withdrawal), and the behavioral response (e.g., blaming, information seeking, coping strategies; Bachura et al., 2022; Budimir et al., 2021). Moreover, feeling a strong negative emotion can turn into a physical reaction toward cybersecurity breaches and can lead to long-term consequences (Budimir et al., 2021; e.g., an increased heartbeat as a reaction to a breach could occur in later stages when interacting with cybersecurity on a daily basis). These emotions can vary over time after an incident. For instance, the perception of a breach might result in high anxiety and turn into sadness over time (Bachura et al., 2022).

Post-Incident

Emotional Coping. When it comes to emotional coping, the following section examines how incidents, attacks or security policy requirements are emotionally processed, either individually or at corporate level.

At the individual level, albeit in terms of precaution taking, outward emotion-focused-coping (e.g., venting, reaching out to colleagues) takes beneficial effects on security behavior.

The inward-looking processing, in contrast, leads to denial and distancing (Liang et al., 2019). At the company level, the employer's handling of an incident triggered by an employee can affect whether shame or guilt is felt. Shame as a result of blaming can lead to self-protective actions and ultimately result in a downward spiral, while a response of understanding, resulting in guilt, from the employer can lead to self-acceptance, action, and learning behaviors (Renaud, Searle, & Dupuis, 2021). Both studies emphasize the importance of conveying explicit emotion-coping strategies, as well as the relevance of social networks and the handling of peers or companies with the incident.

Global

Studies show that negative emotions generally dominate towards cybersecurity (e.g., Cram et al., 2021; Pham et al., 2019; Renaud, Zimmermann, et al., 2021) and security experts (Menges et al., 2022), which are not only characterized by their complex nature, but can also be ambivalent as demonstrated by exploratory approaches (e.g., Renaud, Zimmermann, et al., 2021). This emphasizes once again that emotions from various sub-areas take an emotional spillover to the concept of cybersecurity as well as well-being or job satisfaction. This effect is illustrated by the consideration of complex constructs such as cybersecurity fatigue or burnout, which causes employees to ignore policies, develop workarounds or reduce security efforts to a minimum (Cram et al., 2021; Pham et al., 2019). However, they do not consider emotions specifically, which can prove fatal. For example, fear can lead to avoidance behaviors, while anxiety can lead to surveillance and vigilance (Cheung-Blunden et al., 2019).

3. Challenges to the current approach to emotion research in the field of cybersecurity

Reviewing the literature reveals five, albeit partially interrelated, key challenges to the study of emotions in cybersecurity:

1. Conceptualization of Emotions & Cybersecurity

As pointed out in the field of conceptualization, the terms "affect", "mood" and "emotions" are used interchangeably in many studies. However, this approach can involve high risk, as the terminologies, even if conditionally similar, have key differences (Martin et al., 1993). The definition of the terminology itself also entails different approaches. For example, some studies consider affect purely in terms of valence and arousal (e.g., Beris et al., 2015; Conrad et al., 2020), while other studies refer to 'affect' as valenced emotions in general (e.g., Gulenko, 2014). Nevertheless, the studies presented show that in the context of cybersecurity, a mere examination of valence is not sufficient. For example, fear and anxiety, although identical in valence, result in fundamentally different behavior. While fear ends in avoidance or flight behavior (Frijda et al., 1989; Roseman et al., 1994), anxiety, as an emotion triggered by the absence of obvious signals of danger in the case of fear, can lead to information search as well as precaution-taking to the resolution of ambiguity (Bachura et al., 2022; Cheung-Blunden et al., 2019; Woody & Szechtman, 2011). Furthermore, it is considered central to clearly explain which approach is chosen for the definition of emotions, e.g., the functionalist approach views emotions as a reaction to a trigger, which is followed by clearly defined behaviors, depending on the definition (Campos et al., 1994). However, the connection between antecedent-behavior proves to be difficult in the digital context. Triggers considered so far are abundant on the internet and if recognized, there is the possibility that previously defined behavior patterns cannot be applied due to a lack of technical know-how (Cheung-Blunden et al., 2019).

Conceptualization of cybersecurity also presents researchers with a fundamental challenge. On the one hand, it is essential to define the role that humans play within the security system. Thus, they can either be seen as responsible and therefore the "weakest link", who is moved, for example, by fear to conform behavior (e.g., Johnston & Warkentin, 2010; Renaud, Searle, & Dupuis, 2021). On the other hand, humans can also be seen as an opportunity and, on the basis of targeted measures, become part of the solution, e.g. by stimulating interest in promoting psychological capabilities (Burns et al., 2019) or modifying the design of an interface to improve password strength (Gulenko, 2014). Additionally, the definition of the phase of the cybersecurity process proves to be fundamental, as an initial indication is given that emotions vary across the phases and also cause different behaviors. For example, for precaution taking anxiety results in avoidance behavior (Burns et al., 2019), while anxiety after a data breach might motivate information seeking (Bachura et al., 2022).

2. Measurement of emotion

The multidimensional nature of emotions, for example, how emotions are perceived subjectively, and how individual arousal expresses itself or with which physical reaction a person responds, poses the challenge to researchers as to how to accurately measure emotions (Frijda, 1988).

On the one hand, quantitative measurement methods often risk not being able to adequately grasp the complex character of emotions. Even though the popularity of word counts of emotional term, or the use of face recognition is increasing (Maithri et al., 2022), these methodologies prove challenging. For example, emotions can be verbalized differently by individuals or the declaration of an emotion

can evoke different associations and be recorded incorrectly accordingly (Hoemann et al., 2020). Furthermore, these analyses often include only basic emotions and neglect more complex ones (such as helplessness or uncertainty). There is also evidence that even if only several defined emotions are considered in face recognition due to the lack of authenticity, even with premium devices, 20% of misinterpretations take place (Skiendziel et al., 2019), which can possibly have fatal consequences in the interpretation. This, of course, only under the premise that people also show the appropriate facial expression for their emotion and do not mimic the facial expressions of others (e.g. the interviewer; Landowska & Miler, 2016). Furthermore, it was shown that there are no universal features of facial expressions that identify an emotion. This assumption is due to artifacts based on a closed data collection design (e.g., the selection of a represented emotion from a given emotion list Gendron et al., 2018). Therefore, a multimodal measurement approach is suggested according to theory of constructed emotions (e.g., vocalization and facial movements). However, if only one modality is to be assessed, a self-report should be adopted since it can reveal the subjective experience of an emotion. “Objective” measurement procedures do not exist according to the theory of constructed emotions (Feldman-Barrett & Westlin, 2021).

On the other hand, qualitative approaches such as self-reports also pose challenges. For example, during an interview, the interviewee must be able to retrospectively fall back on the emotional experience, have the appropriate reflection and verbalization skills and overcome fears of social desirability (Quigley et al., 2014).

However, regardless of the methodical procedure, it is particularly essential to check for the manipulation in the case of induction of emotions. As an example, when using fear appeals, it should be checked whether fear has been successfully induced and no other emotions such as anger about the type of communication. For instance, for the difference between anger and fear, Lerner and Keltner (2001) found that anger results in more optimistic and risk-seeking, while fear made pessimistic and risk-avoidant behavior. Otherwise, research will be led by a fundamental misinterpretation.

The time frame of the survey can also have a significant impact on the outcome of the study. Cross-sectional studies risk not only neglecting developmental effects, but also failing to adequately identify effects (Anstey & Hofer, 2004). At the same time, longitudinal studies also have their downsides (Schaie & Hofer, 2001). In particular, in the example of emotions, drop-out may occur, in which only positively tuned subjects are retained. This could be misinterpreted as a positive development.

3. Measurement of cybersecurity behavior

Measuring cybersecurity behavior or behavioral intent using a survey brings numerous efficiency benefits but carries both the risk of low external validity as well as the risk of overlooking significant behavioral patterns. In particular, employees with strong negative affect tend not only to avoid policies but also to develop alternative behavior patterns or workarounds (‚shadow security‘; Beris et al., 2015), which are not captured or distorted by records of behavioral intention (Kirlappos et al., 2014). Other side effects of subparts of cybersecurity, such as for awareness, education and communication: The risk that threat appeals will have an impact on well-being or job satisfaction due to negative affect, can remain undetected here if no further measurements are implemented. Fear appeals, for example, can lead to a desirable outcome, while possibly leading to more complaints (Janis & Feshbach, 1953). Consequently, evoked negative emotions can result in spillover effects, e.g., on the relationship between employees and security experts.

In general, the action dimension of the behavior to be measured should be considered. Here, behavior can be characterized on three levels. First, in terms of behavior frequency (one-time or repeated), then based on behavior direction (omission, commission or inhibit) and finally in terms of behavior novelty (new behavior or behavior change; Renaud & Dupuis, 2019). Further, it is important to explicitly distinguish between behavioral intention and actual behavior. An employee may have the intention to actually carry out preventive behavior, but for example, the emotion in the retrieval situation can ultimately lead to emotional behaviors being shown as for instance demonstrated by studies examining emotion as a target such as Abroshan et al. (2021).

For the challenges of the data collection time frame, the same is applicable for Cybersecurity behavior as for the measurement of emotions. Especially in the field of information systems research, its use is strongly recommended to look at contextual conditions and biases. For this purpose, researchers are encouraged to integrate a qualitative part that uncovers these effects (Anstey & Hofer, 2004; Venkatesh & Vitalari, 1991).

4. Context delimitation

In emotion research, the context of how an emotion is perceived, expressed, and regulated is central (Greenaway et al., 2018). When looking at emotions, apart from cultural differences (Butler et al., 2007), it can make a fundamental difference whether the perception of emotions is considered, for example, in a private or professional context. It is also important for the measurement that emotions can be interpreted differently in a different context and retrospectively. Accordingly, emotions should ideally be measured where they occur (Kouamé & Liu, 2021).

In a professional context, the definition of the concept of security (cybersecurity, information security, or similar) of a company is also central to the understanding of the context, as this understanding indicates the role attributed to humans in the system. While definitions of information security often entail a focus on the protection of data and view the human being as a responsible instance and vulnerability, definitions of cybersecurity often includes humans as a part of the security system worth protecting (X. A. Zhang & Borden, 2020). Despite this inclusion in the system, humans are currently often blamed for errors and seen as a weak point (Renaud, Searle, & Dupuis, 2021; Schneier, 2004). Furthermore, socio-technical systems that involve humans as part of the solution to the cybersecurity strategy can trigger other emotions towards cybersecurity (assuming that security hygiene precautions have been taken (Beris et al., 2015; Zimmermann & Renaud, 2019). For example, in the case of supportive behavior of the supervisor, a mistake becomes the basis of learning behavior (Guilt), while accusations can result in a downward spiral (Shame; Renaud, Searle, & Dupuis, 2021). Thus, researchers need to clearly delimitate the context of their study, in order to make correct interpretations of the effect of emotions on cybersecurity behavior.

5. Complexity of cybersecurity and emotions

Emotions are characterized by complexity. The perception of one emotion does not exclude the perception of another one. Even in their valence, emotions can be contradictory. Furthermore, cybersecurity represents a novel context: previously given triggers and behavioral patterns are not given, and the simple consideration of basic emotions, let alone affect, is not sufficient to grasp this complex core. Renaud et al. (2021) show a clearly negative tone towards cybersecurity but at the same time inconsistent results to approach behavior. Emotions not only seem to vary across areas but are also perceived as ambivalent towards cybersecurity itself. These areas must also be identified to be able to detect which area has the greatest influence on behavioral patterns. For cybersecurity, the impact areas need to be identified in a similar way to account for complexity. For instance, Renaud, Zimmermann, et al.'s identification of different approach tendencies toward different terms of cybersecurity in 2021 suggests that different sub contexts of cybersecurity are also invested with different emotions. More specifically, with regard to the contradiction of emotions, the following observations can be made: Initial studies show that a negative attitude or negative emotions towards cybersecurity generally exist. Humans feel overwhelmed, scared, helpless, or confused. At the same time, employees feel responsible and feel the need to feel safe (Renaud, Zimmermann, et al., 2021). It becomes apparent that a negative tone towards cybersecurity is prevalent, however, at the same time, it emerges that these emotions are complex in nature and ambivalent stances are possible, nonetheless. This ambivalence can lead to cognitive dissonance and as a result negative feelings (Festinger, 1957). However, humans strive for a state of balance, which is why it would be assumed that in order to compensate for this dissonance and to protect the self-image (e.g., that insufficient knowledge/security behavior could be disclosed, even though one considers oneself a compliant employee). Cognitive dissonance, in turn, causes human to „bury their heads in the sand" and avoid cybersecurity in general, as postulated by the Ostrich effect (Carlson, 2013). This cognitive dissonance is an exemplary case of what could be examined in more detail to illuminate the gap between knowledge and behavior. Thus, emotion research in the field of cybersecurity is challenged by the complexity of both concepts, cybersecurity as well as emotions.

4. Conclusion and Research Outlook

The literature review demonstrates that emotions in the field of cybersecurity have a fundamental influence. Thus, the mere mapping of security behavior by cognitive models (e.g., Herath & Rao, 2009; Ifinedo, 2012) appears insufficient. Therefore, emotions should be fundamentally included in research work. In fact, emotions prove to be significant not only in the global concept of cybersecurity but also in all related process stages (e.g., precaution-taking, coping, ...) as well as in the emotion-related processes (e.g., perception, learning, decision-making behavior, ...).

Despite this conceptual contribution, this literature review also makes a significant practical contribution. Among other things, it indicates that conventional Security Education Training and Awareness ("SETA") is not a sufficient answer to „the human vulnerability" and that the type of communication influencing emotion is highly relevant. Communication, however, needs to be carefully designed as, for example, threat appeals can be inconsistent and can result in disadvantageous side effects, while pure positivity / „it's a fun thing" risks not bringing about a change in behavior, as negative feelings such as fear, or uncertainty are not sufficiently addressed (‚toxic positivity'; Gross & Levenson, 1997; Sokal et al., 2020). In addition, a study of incidental emotions on security behavior can provide information about the necessary emotional workplace design for desired security behavior.

To meet these challenges and objectives, it is necessary to investigate in an explorative and unbiased way which emotions are generally perceived towards cybersecurity. A qualitative approach is recommended as the review of recent

studies indicates that the mode of action of emotions on cybersecurity is still unclear and that previous results are too heterogeneous. In this approach, it is important to consider the complexity of both cybersecurity and emotions to capture all significant sub-areas of cybersecurity as well as the potential ambivalence of emotions. In addition, antecedents, as well as consequences of emotions, should be considered in order to create a basis for further research. A quantitative approach at this point, in turn, risks neglecting the complexity and thereby overlooking central mechanisms.

Based on these results, the means of emotion integration can be considered, for example, in the design of cybersecurity or awareness measures, as well as how a workplace should be emotionally designed to promote security behavior. This research can emphasize the importance of a suitable design of security measures (e.g., in terms of usable security) and the important role of humans as an opportunity.

References

(ISC)². (2020). *Cybersecurity Workforce Study: Cybersecurity professionals stand up to a pandemic.* International Information System Security Certification Consortium (ISC)².

Abroshan, H., Devos, J., Poels, G., & Laermans, E. (2021). COVID-19 and Phishing: Effects of Human Emotions, Behavior, and Demographics on the Success of Phishing Attempts During the Pandemic. *IEEE Access*, *9*, 121916–121929. https://doi.org/10.1109/ACCESS.2021.3109091

Allianz. (2022). *Allianz Risk Barometer 2022*. https://www.agcs.allianz.com/news-and-insights/reports/allianz-risk-barometer.html

Anstey, K. J., & Hofer, S. M. (2004). Longitudinal designs, methods and analysis in psychiatric research. *The Australian and New Zealand Journal of Psychiatry*, *38*(3), 93–104. https://doi.org/10.1080/j.1440-1614.2004.01343.x

Bachura, E., Valecha, R., Rui Chen, & Rao, H. R. (2022). The OPM Data Breach: An investigation of shared emotional reactions on Twitter. *MIS Quarterly*, *46*(2), 881–910. https://doi.org/10.25300/MISQ/2022/15596

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Barrett, M. P. (2018). *Framework for improving critical infrastructure cybersecurity.* https://doi.org/10.6028/NIST.CSWP.04162018

Beris, O., Beautement, A., & Sasse, M. A. (2015). Employee Rule Breakers, Excuse Makers and Security Champions: Mapping the risk perceptions and emotions that drive security behaviors. In A. Somayaji, R. Böhme, P. van Oorschot, & M. Mannan (Eds.), *Proceedings of the 2015 New Security Paradigms Workshop* (pp. 73–84). ACM. https://doi.org/10.1145/2841113.2841119

Böhme, R., Laube, S., & Riek, M. (2018). A fundamental approach to cyber risk analysis. *Variance*, *11*(2). https://informationsecurity.uibk.ac.at/pdfs/blr2019_fundamentalapproachcyberriskinsurance_variance.pdf

Brosch, T., Scherer, K. R., Grandjean, D., & Sander, D. (2013). The impact of emotion on perception, attention, memory, and decision-making. *Swiss Medical Weekly*, *143*, w13786. https://doi.org/10.4414/smw.2013.13786

Buck, R., Khan, M., Fagan, M., & Coman, E. (2018). The User Affective Experience Scale: A Measure of Emotions Anticipated in Response to Pop-Up Computer Warnings. *International Journal of Human–Computer Interaction*, *34*(1), 25–34. https://doi.org/10.1080/10447318.2017.1314612

Budimir, S., Fontaine, J. R. J., & Roesch, E. B. (2021). Emotional Experiences of Cybersecurity Breach Victims. *Cyberpsychology, Behavior and Social Networking*, *24*(9), 612–616. https://doi.org/10.1089/cyber.2020.0525

Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2010). Information Security Policy Compliance: An Empirical Study of Rationality-Based Beliefs and Information Security Awareness. *MIS Quarterly*, *34*(3), 523. https://doi.org/10.2307/25750690

Burns, A. J., Roberts, T. L., Posey, C., & Lowry, P. B. (2019). The Adaptive Roles of Positive and Negative Emotions in Organizational Insiders' Security-Based Precaution Taking. *Information Systems Research*, *30*(4), 1228–1247. https://doi.org/10.1287/isre.2019.0860

Butler, E. A., Lee, T. L., & Gross, J. J. (2007). Emotion regulation and culture: Are the social consequences of emotion suppression culture-specific? *Emotion (Washington, D.C.)*, *7*(1), 30–48. https://doi.org/10.1037/1528-3542.7.1.30

Campos, J. J., Mumme, D., Kermoian, R., & Campos, R. G. (1994). A Functionalist Perspective on the Nature of Emotion. *JAPANESE Journal Of Research On Emotions*, *2*(1), 1–20. https://doi.org/10.4092/jsre.2.1

Carlson, E. N. (2013). Overcoming the Barriers to Self-Knowledge: Mindfulness as a Path to Seeing Yourself as You Really Are. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *8*(2), 173–186. https://doi.org/10.1177/1745691612462584

Cheung-Blunden, V., Cropper, K., Panis, A., & Davis, K. (2019). Functional divergence of two threat-induced emotions: Fear-based versus anxiety-based cybersecurity preferences. *Emotion*, *19*(8), 1353–1365. https://doi.org/10.1037/emo0000508

Cisco. (2018). *Cisco 2018 Annual Cybersecurity Report*. https://www.cisco.com/c/en/us/products/security/security-reports.html

Clore, G. L., Gasper, K., & Garvin, E. (2001). Affect as information. In *Handbook of affect and social cognition* (pp. 121–144). Lawrence Erlbaum Associates Publishers.

Clore, G. L., Schwarz, N., & Conway, M. (1994). Affective causes and consequences of social information processing. *Handbook of Social Cognition*, *1*, 323–417.

Conrad, C., Aziz, J., Smith, N., & Newman, A. (2020). What Do Users Feel? Towards Affective EEG Correlates of Cybersecurity Notifications. In F. D. Davis (Ed.), *Lecture notes in information systems and organisation (Print): Vol. 43. Information systems and*

neuroscience: NeuroIS Retreat 2020 (1st ed., Vol. 43, pp. 153–162). Springer. https://doi.org/10.1007/978-3-030-60073-0_17

Cram, W. A., Proudfoot, J. G., & D'Arcy, J. (2021). When enough is enough: Investigating the antecedents and consequences of information security fatigue. *Information Systems Journal*, *31*(4), 521–549. https://doi.org/10.1111/isj.12319

D'Arcy, J., & Hovav, A. (2009). Does One Size Fit All? Examining the Differential Effects of IS Security Countermeasures. *Journal of Business Ethics*, *89*(S1), 59–71. https://doi.org/10.1007/s10551-008-9909-7

Dupuis, M., Jennings, A., & Renaud, K. (2021). Scaring people is not enough: an examination of fear appeals within the context of promoting good password hygiene. *Proceedings of the 22st Annual Conference on Information Technology Education*, 35–40. https://doi.org/10.1145/3450329.3476862

Dupuis, M., Renaud, K., & Jennings, A. (2022). *Fear might motivate secure password choices in the short term, but at what cost?* https://doi.org/10.24251/HICSS.2022.585

Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. *Nebraska Symposium on Motivation*, *19*, 207–283.

Ekman, P., & Davidson, R. J. (1994). *The nature of emotion: Fundamental questions*. *Series in affective science*. Oxford University Press.

ENISA (2019). Cybersecurity culture guidelines: behavioural aspects of cybersecurity. *European Union Agency for Network and Information Security*.

Estrada, C. A., Isen, A. M., & Young, M. J. (1994). Positive affect improves creative problem solving and influences reported source of practice satisfaction in physicians. *Motivation and Emotion*, *18*(4), 285–299. https://doi.org/10.1007/BF02856470

Fagan, M., Albayram, Y., Khan, M. M. H., & Buck, R. (2017). An investigation into users' considerations towards using password managers. *Human-Centric Computing and Information Sciences*, *7*(1), 1–20. https://doi.org/10.1186/s13673-017-0093-6

Feldman-Barrett, L., & Westlin, C. (2021). Navigating the science of emotion. *Emotion Measurement*, 39–84. https://doi.org/10.1016/B978-0-12-821124-3.00002-8

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Frijda, N. H. (1988). The laws of emotion. *American Psychologist*, *43*(5), 349–358. https://doi.org/10.1037/0003-066X.43.5.349

Frijda, N. H. (1993). Moods, emotion episodes, and emotions. In *Handbook of emotions* (pp. 381–403). The Guilford Press.

Frijda, N. H., Kuipers, P., & Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*, *57*(2), 212–228. https://doi.org/10.1037/0022-3514.57.2.212

Gendron, M., Crivelli, C., & Barrett, L. F. (2018). Universality Reconsidered: Diversity in Making Meaning of Facial Expressions. *Current Directions in Psychological Science*, *27*(4), 211–219. https://doi.org/10.1177/0963721417746794

George, J. M. (1996). Trait and state affect. *Individual Differences and Behavior in Organizations*, *1*, 145–171.

Greenaway, K. H., Kalokerinos, E. K., & Williams, L. A. (2018). Context is Everything (in Emotion Research). *Social and Personality Psychology Compass*, *12*(6), 1-18. https://doi.org/10.1111/spc3.12393

Gross, J. J., & Levenson, R. W. (1997). Hiding feelings: The acute effects of inhibiting negative and positive emotion. *Journal of Abnormal Psychology*, *106*(1), 95–103. https://doi.org/10.1037/0021-843X.106.1.95

Gulenko, I. (2014). Improving passwords: Influence of emotions on security behaviour. *Information Management & Computer Security*, *22*(2), 167–178. https://doi.org/10.1108/IMCS-09-2013-0068

Han, S., Lerner, J. S., & Keltner, D. (2007). Feelings and Consumer Decision Making: The Appraisal-Tendency Framework. *Journal of Consumer Psychology*, *17*(3), 158–168. https://doi.org/10.1016/S1057-7408(07)70023-2

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

IBM. (2022). *Cost of a Data Breach Report*. https://www.ibm.com/reports/data-breach

IBM Global Technology Services. (2014). *IBM security services 2014 cyber security intelligence index*. https://omnipush.com/docs/IBM_Cyber_Security_Intelligence_20450.pdf

Ifinedo, P. (2012). Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory. *Computers & Security*, *31*(1), 83–95. https://doi.org/10.1016/j.cose.2011.10.007

Isen, A. M. (1984). Toward understanding the role of affect in cognition. In *Handbook of social cognition, Vol 3* (pp. 179–236). Lawrence Erlbaum Associates Publishers.

Janis, I. L., & Feshbach, S. (1953). Effect of fear-arousing communications. *Journal of Abnormal Psychology*, *48*(1), 78–92. https://doi.org/10.1037/h0060732

Johnston, A. C., & Warkentin, M. (2010). Fear Appeals and Information Security Behaviors: An Empirical Study. *MIS Quarterly*, *34*(3), 549–566. https://doi.org/10.2307/25750691

Kahneman, D. (2012). *Thinking, fast and slow*. Penguin Books.

Kirlappos, I., Parkin, S., & Sasse, M. A. (2014, February 23). Learning from "Shadow Security:" Why Understanding Non-Compliant Behaviors Provides the Basis for Effective Security. In M. Smith & D. Wagner (Eds.), *Proceedings 2014 Workshop on Usable Security.* Internet Society. https://doi.org/10.14722/usec.2014.23007

Kok, L. C. de, Oosting, D., & Spruit, M. (2020). The Influence of Knowledge and Attitude on Intention to Adopt Cybersecure Behaviour. *Information & Security: An International Journal*, *46*(3), 251–266. https://doi.org/10.11610/isij.4618

Kouamé, S., & Liu, F. (2021). Capturing emotions in qualitative strategic organization research. *Strategic Organization*, *19*(1), 97–112. https://doi.org/10.1177/1476127020935449

Lazarus, R. S. (1991). *Emotion and adaptation*. *Emotion and adaptation*. Oxford University Press.

Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition and Emotion*, *14*(4), 473–493. https://doi.org/10.1080/026999300402763

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology*, *81*(1), 146–159. https://doi.org/10.1037/0022-3514.81.1.146

Liang, H., Xue, Y., Pinsonneault, A., & Wu, Y. (2019). What Users Do Besides Problem-Focused Coping When Facing IT Security Threats: An Emotion-Focused Coping Perspective. *MIS Quarterly*, *43*(2), 373–394. https://doi.org/10.25300/MISQ/2019/14360

Lutz, C. A. (1988). *Unnatural Emotions: Everyday Sentiments on a Micronesian Atoll and Their Challenge to Western Theory*. University of Chicago Press.

Martin, L. L., Ward, D. W., Achee, J. W., & Wyer, R. S. (1993). Mood as input: People have to interpret the motivational implications of their moods. *Journal of Personality and Social Psychology*, *64*(3), 317–326. https://doi.org/10.1037/0022-3514.64.3.317

Massumi, B. (1995). The Autonomy of Affect. *Cultural Critique*(31), 83. https://doi.org/10.2307/1354446

McConnell, M. M., & Eva, K. W. (2012). The role of emotion in the learning and transfer of clinical skills and knowledge. *Academic Medicine : Journal of the Association of American Medical Colleges*, *87*(10), 1316–1322. https://doi.org/10.1097/ACM.0b013e3182675af2

Menges, U., Hielscher, J., Buckmann, A., Kluge, A., Sasse, M. A., & Verret, I. (2022). Why IT Security Needs Therapy. In S. Katsikas, C. Lambrinoudakis, N. Cuppens, J. Mylopoulos, C. Kalloniatis, W. Meng, S. Furnell, F. Pallas, J. Pohle, M. A. Sasse, H. Abie, S. Ranise, L. Verderame, E. Cambiaso, J. Maestre Vidal, & M. A. Sotelo Monge (Eds.), *Lecture Notes in Computer Science. Computer Security. ESORICS 2021 International Workshops* (Vol. 13106, pp. 335–356). Springer International Publishing. https://doi.org/10.1007/978-3-030-95484-0_20

Ostrom, T. M. (1969). The relationship between the affective, behavioral, and cognitive components of attitude. *Journal of Experimental Social Psychology*, *5*(1), 12–30. https://doi.org/10.1016/0022-1031(69)90003-1

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, *36*(1), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Pham, H. C., Brennan, L., & Furnell, S. (2019). Information security burnout: Identification of sources and mitigating factors from security demands and resources. *Journal of Information Security and Applications*, *46*, 96–107. https://doi.org/10.1016/j.jisa.2019.03.012

Reeve, J. (2018). *Understanding motivation and emotion* (Seventh edition). John Wiley & Sons Inc.

Renaud, K., & Dupuis, M. (2019). Cyber security fear appeals: Unexpectedly complicated. In M. Carvalho (Ed.), *ACM Digital Library, Proceedings of the New Security Paradigms Workshop* (pp. 42–56). Association for Computing Machinery. https://doi.org/10.1145/3368860.3368864

Renaud, K., Searle, R., & Dupuis, M. (2021). Shame in Cyber Security: Effective Behavior Modification Tool or Counterproductive Foil? In *ACM Digital Library, New Security Paradigms Workshop* (pp. 70–87). Association for Computing Machinery. https://doi.org/10.1145/3498891.3498896

Renaud, K., Zimmermann, V., Schürmann, T., & Böhm, C. (2021). Exploring cybersecurity-related emotions and finding that they are challenging to measure. *Humanities and Social Sciences Communications*, *8*(1), 75. https://doi.org/10.1057/s41599-021-00746-5

Roseman, I. J., Wiest, C., & Swartz, T. S. (1994). Phenomenology, behaviors, and goals differentiate discrete emotions. *Journal of Personality and Social Psychology*, *67*(2), 206–221. https://doi.org/10.1037/0022-3514.67.2.206

Schaie, K. W., & Hofer, S. M. (2001). Longitudinal studies in aging research. In *Handbook of the psychology of aging, 5th ed* (pp. 53–77). Academic Press.

Schneier, B. (2004). *Secrets and lies: Digital security in a networked world* (1st edition). John Wiley & Sons Inc.

Schwarz, N. (1990). Feelings as information: Informational and motivational functions of affective states. In *Handbook of motivation and cognition: Foundations of social behavior, Vol. 2* (pp. 527–561). The Guilford Press.

Schwarz, N., & Clore, G. L. (2007). Feelings and phenomenal experiences. In *Social psychology: Handbook of basic principles, 2nd ed* (pp. 385–407). The Guilford Press.

Shouse, E. (2005). Feeling, Emotion, Affect. *M/C Journal*, *8*(6). https://doi.org/10.5204/mcj.2443

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

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research*, *177*(3), 1333–1352. https://doi.org/10.1016/j.ejor.2005.04.006

Sokal, L., Trudel, L. E., & Babb, J. (2020). It's okay to be okay too. Why Calling Out Teachers "Toxic Positivity" may Backfire. *Canadian Education Network (EdCan)*, *60*(3).

van Schaik, P., Renaud, K., Wilson, C., Jansen, J., & Onibokun, J. (2020). Risk as affect: The affect heuristic in cybersecurity. *Computers & Security*, *90*, 101651. https://doi.org/10.1016/j.cose.2019.101651

Venkatesh, A., & Vitalari, N. P. (1991). Longitudinal surveys in information systems research: An examination of issues, methods, and applications. *The Information Systems Challenge: Survey Research Methods*, 115–144.

Williams, S., Ford, J., & Kensinger, E. (2022). The power of negative and positive episodic memories. *Cognitive, Affective, & Behavioral Neuroscience*. Advance online publication. https://doi.org/10.3758/s13415-022-01013-z

Woody, E. Z., & Szechtman, H. (2011). Adaptation to potential threat: The evolution, neurobiology, and psychopathology of the security motivation system. *Neuroscience and Biobehavioral Reviews*, *35*(4), 1019–1033. https://doi.org/10.1016/j.neubiorev.2010.08.003

Zhang, P. (2013). The Affective Response Model: A Theoretical Framework of Affective Concepts and Their Relationships in the ICT Context. *Management Information Systems Quarterly (MISQ)*, *37*, 247–274. https://doi.org/10.25300/MISQ/2013/37.1.11

Zhang, X. A., & Borden, J. (2020). How to communicate cyber-risk? An examination of behavioral recommendations in cybersecurity crises. *Journal of Risk Research*, *23*(10), 1336–1352. https://doi.org/10.1080/13669877.2019.1646315

Zimmermann, V., & Renaud, K. (2019). Moving from a 'human-as-problem" to a 'human-as-solution" cybersecurity mindset. *International Journal of Human-Computer Studies*, *131*, 169–187. https://doi.org/10.1016/j.ijhcs.2019.05.005

DOI: 10.2478/9788366675896-002 • Technikpsychologie • 14-30

Proceedings of the International Symposium on Technikpsychologie

Clean up my Phone: Field Trial Results of a Privacy Tool Increasing Transparency about App Behavior

Susen Döbelt†, Josephine Halama

Chemnitz University of Technology, Professorship of Cognitive Psychology and Human Factors, Wilhelm-Raabe-Straße, 4309120 Chemnitz, Germany

Abstract

Smartphones and mobile apps have become everyday companions and offer various services. However, it is often not transparent how apps deal with sensitive user data and how to keep control over it. To increase transparency and enable users to make qualified privacy decisions, we investigated our developed tool – the AndProtect app - that combines dynamic and static analysis during a six-week field trial with N = 26 participants. We were interested in whether our tool nudges users to privacy-preserving behavior and if it is suitable for everyday use. Results revealed that it provoked a short increase of uninstallations – a "clean up", but lacks to promote knowledge or use of the permission function. In contrast to the constant user experience, the longitudinal usability ratings decreased, but both positively differ from benchmarks over time. Participants were interested in the content, considered to be well presented, and provoked behavioral implications, e.g., increased attention to and reasoning of app behavior. Thus, the AndProtect app could contribute to strengthening privacy-related awareness when dealing with mobile apps in everyday life. To increase behavioral effectiveness, tools like ours could expand information, offer more options for action, and should be tailored to the user.

Keywords

privacy • transparency • smartphone app • user behaviour

1. Introduction

Smartphones have become companions in our everyday lives. A large number of applications (apps) make a wide variety of services available everywhere and at any time. Currently and in the future, the operating system Android is widely used [IDC, 2021] and about three million apps are available for download [AppBrain, 2021]. Even though they offer many advantages, it is often not transparent to users how applications deal with sensitive user data and how to control it [Nissenbaum, 2009], and - as well as with other digital services - it is difficult to preserve one's privacy. Existing methods are often not accessible or insufficient usability prevents users from effectively use of the privacy functionality offered [Hansen et al., 2004].

For instance, the listing of permissions allows the user to judge which (groups of) data are accessed by an app and to exercise control over some (not all) by granting or rejecting permissions. Nevertheless, it is often not clear which data precisely is affected, at what level of granularity it is processed, to whom and how it is forwarded. App analyses [Enck et al., 2014], [Fraunhofer AISEC, 2014] showed that the majority of apps revealed suspicious behavior in terms of potential privacy invasion with the implicit or explicit consent of the user. In addition, user studies showed that permissions are often little-noticed or understood by users [Kelley et al., 2013], [Felt et al., 2012]. Only a few infer the scope of permissions correctly [Shen et al., 2021] and rather accept presented defaults instead of adapting them to their privacy preferences [Joeckel and Dorgruel, 2020]. Therefore, permissions do not provide an adequate and usable basis for qualified privacy decisions. Hence, an asymmetry of informational transparency [Friedewald, 2018] or "*informational inequality*" [Nissenbaum, 2009] between app users and app providers evolves.

Transparency about data access and processing is an essential condition for enabling personal data protection [Cavoukian, 2009]. However, users are concerned [Trepte et al., 2018] and criticize the fact [Döbelt et al., 2020] that they have no insight into what is happening with their data. Conversely, if inappropriate data collection from apps becomes transparent, the feeling of privacy invasion is connected [Friedewald, 2018], [Shklovski et al., 2014]. To increase transparency and enable users to make qualified

†*Corresponding author*: *Susen Döbelt*
E-mail: susen.doebelt@psychologie.tu-chemnitz.de

privacy decisions, we have developed a tool that analyzes app behavior. Our AndProtect app combines dynamic and static analysis methods [Feizollah et al., 2015], which provide insights about potential and actual data flows by analyzing an app's source code (static analysis), during usage (dynamic analysis), and preparing the results in a user-centered way. The presented field trial aimed to investigate empirically, whether the AndProtect app nudges users to change data protection-relevant app usage behavior and how user-friendly it is perceived in everyday life.

2. Related Work

The concept of nudging aims to improve individual well-being without limiting the freedom of choice [Acquisti et al., 2020]. Digital nudges enable people to change behavior in a specific way without prohibiting options [Cunha and Aguiar, 2020]. They can encourage users for example towards more beneficial privacy and security choices by accounting for hurdles in decision-making [Thaler and Sunstein, 2008]. One hurdle is incomplete or asymmetric information [Acquisti et al., 2020]. Furthermore, overconfidence can lead to the underestimation of risks as well as availability heuristics can lead to a misestimating of probabilities [Acquisti et al., 2020]. In particular, these hurdles apply to mobile app usage when heuristic decision processes are typical. As described above, users' accessibility to detailed information on the data processing of apps is limited. Furthermore, in case of uncertainty, users rather rely (overconfidently) on that apps only incorporate the personal information required by their functionality [King, 2012]. Additionally, users fail to remember permissions in long term [King, 2012], which diminishes the salience of potential risks.

Previous authors suggested that "*dedicated mobile apps can assist users with nudges in making beneficial privacy decisions*" [Acquisti et al., 2020], p. 30], or "*a nudge may take the form of an alert that informs the user of the risk*" [Balebako et al., 2011]. User research on privacy nudges in the field of mobile apps often focused on the usage and design of permission screens. Kelley et al. [Kelley et al., 2013] were able to show that a modified privacy facts sheet highlighting data access, increases the decision for downloading a privacy-protecting app. However, the lab participants wanted a better understanding of why certain apps requested permissions. Therefore, the authors emphasized that an advanced display should include frequencies and purposes of permission utilization [Kelley et al., 2013]. In addition, Gerber and colleagues [Gerber et al., 2017] could show that participants of an online study were able to understand complex permission information and concluded that merely a simplification could even lead to an impairment of the privacy-related decision. Therefore, the right composition of detailed information and usable presentation seems crucial for nudging users toward privacy.

In a study by Almuhimedi and colleagues [Almuhimedi et al., 2015] the exploitation of runtime permissions during the usage of apps was investigated. They implemented a nudge that reports the frequency of accessed permissions and examined behavioral consequences of a weak (one message within 7 days) and after that a strong nudge (daily overlays within 8 days) within a field trial (N = 23). Results showed that the weak nudge leads to privacy-protective behavior by reviewing and adjusting apps' permissions. However, the strong privacy nudges could reinforce this effect. Furthermore, the authors highlighted the right composition of salience as crucial for effective nudge design.

In another study [Gerber et al., 2018] the effects of *FoxIT*, an Android app incorporating education modules with gamification elements and a static permission analysis method, were investigated. The app could provide information about the risk of installed apps using the Google permissions classification, the total number of requested permissions per app, and explanations for certain permissions. Results of a two-week field trial showed that the *FoxIT* app leads to increased privacy awareness and knowledge of the participants (N = 31). Furthermore, the participants indicated to have made changes to smartphone settings. Beyond that, the use of security measures or the setting of social network privacy did not change.

In addition to the analysis of permissions, some approaches analyze the behavior of apps in detail. In a laboratory experiment on a provided smartphone, Bal and colleagues [Bal et al., 2014] were able to show that their approach *Styx* (dynamic monitoring of information flows via *TaintDroid*) was perceived as user-friendly. Furthermore, the usage of *Styx* contributed to increased user confidence due to the transparent information.

Van Kleek et al. [Van Kleek et al., 2017] suggested using *Data Controller Indicators* to access app information flows and disclose the transfer to third parties. They compared different versions of user interfaces in a lab study (N = 21). Results revealed that more transparent information about data leaks and flows leads to decisions for apps with fewer organizations receiving app data. The authors concluded that transparency about app information flow could support users to make confident and adequate decisions about their preferences.

Based on the literature, our research project aimed to take up analysis approaches about application behavior and make them available to mobile app users in a usable way. In particular, the conducted field trial aimed to investigate the potential of a transparency-increasing tool to change smartphone app behavior in everyday life. Besides behavioral effects, we were interested in usability [ISO 9241-210,

2010] and user experience [ISO 9241-11, 1998], as privacy functionality with insufficient usability prevents usage [Hansen et al., 2004]. Typically, short-term experience with digital tools is evaluated and related to available benchmarks. But different qualities seem to be of different importance [Karapanos et al., 2009], [Kujala et al., 2011] and therefore usability and user experience (UUX) are likely to change over time. Qualitative product studies of mobile phones have shown that prolonged use is tied to how meaningful a product is for personal life [Karapanos et al., 2009]. Furthermore, usability aspects seem to increase over time since user experience decreases [Kujala et al., 2011]. Since our tool provides information to users and behavioral effects may manifest later, the longitudinal course of UUX was interesting for us besides benchmark comparisons.

3. Research Questions and Hypotheses

The research questions of our field trial were: 1.) How does our tool, the AndProtect app, affect users' app usage behavior? and 2.) How suitable is our tool for everyday life? We were interested in whether our tool can nudge app users and could lead to an increased number of uninstallations, engagement with the permission function, and withdrawal of permissions. Therefore, we hypothesize (H):

H1.1: The total amount of installed apps will decrease over time and drop after the installation of our tool.

This effect is probably caused by an altered ratio of uninstallations and installations.

H1.2: The number of uninstallations of apps will increase over time and rise after the installation of our tool.

On the one hand, the installations can increase selectively because of replacements for uninstalled apps. On the other hand, fewer apps could be installed because users behave more conservatively. Furthermore, our tool delivers post-installation information. Taken together, we hypothesize:

H1.3: The number of installations will not differ over time.

As described in section 1, mobile app users could still be granted or withdrawn some permissions after the installation of an app. However, several studies revealed that users do rarely engage with this function. We assume, that our tool could probably change this, as it addresses data access of apps.

H2.1: The knowledge about the permission function will be higher after the trial compared to before.

H2.2: The usage of the permission function will be higher after the trial compared to the baseline.

H2.3: The frequency of withdrawal of permissions will increase over time and will be higher after the installation of our tool.

As described above, we want to gather feedback on the design of our tool and gather quantitative UUX evaluations. As we followed a user-centered design approach during the development, we hypothesize:

H3.1: The UUX evaluations will differ positively from neutral evaluations during the whole field trial.

Concerning the temporal development of UUX, we assume a change of evaluations. As previous and similar research on longitudinal development is rare, our hypothesis is undirected:

H3.2: UUX evaluations will change over time.

Additionally, qualitative impressions of our participants are gathered to enrich quantitative UUX ratings and to answer research question 2 about the usage in everyday life. Therefore, we asked exploratory and open-ended to describe the usage context, advantages as well as disadvantages of our tool.

4. Method

Study Design and Data Analysis

Our study was conducted using a within-subjects design. The independent variable presents either the multiple times of measurement or the presence of our tool. For the dependent variables, e.g. the number of installed and uninstalled apps, the usage of the permissions function, the frequency of withdrawals of permissions, and the UUX, we used mixed methods to gather the participants' behavior and evaluations (for an overview see Appendix A.1).

Quantitative Data

We gathered quantitative data to monitor changes in app usage behavior and UUX evaluations. For H1.1 - 1.3, we collected the app list weekly to identify in- and uninstallations on an individual level. To test H2.1 and 2.2, we asked closed-ended questions concerning the knowledge and usage of the permission function. Furthermore, a self-assessment of behavioral change gathered participants´ estimation of apps affected by permission management at the beginning (T0) and the end of the trial (T6). To answer H2.3, we asked for granting and withdrawal of permissions starting from T1. Usability and user experience (H3.1 and H3.2), were assessed at T2.2, T4, and T6 during the trial period.

For the continuously gathered variables, we ran repeated-measurements ANOVAs to identify changes over time. We tested the assumption of sphericity beforehand. In case of violation, we applied a Greenhouse-Geisser (GG) correction for *F*-ratios. We used a partial eta-square (η_p^2) to quantify effect sizes and Cohen's conventions [Cohen, 1988] for classification. For variables assessed only once during the baseline and trial period, pre-post comparisons were applied after the assumption of normal distribution was tested. Either parametric dependent t-tests or nonparametric Wilcoxon signed-rank tests were used. We tested one-tailed, whenever our hypotheses implicate an increase or decrease of variable values. Here, we used Cohen's *d* according to conventions [Cohen, 1988] as effect size measurement.

Descriptive statistics, such as relative frequencies (in %), means and medians (*M*; *Mdn*), standard deviations (*SD*), and ranges of values (*min*; *max*) are used to describe characteristics of the sample or illustrate common answers.

Qualitative Data

To gain a deeper understanding of everyday suitability, we gathered qualitative data about the usage context, advantages, and disadvantages of the AndProtect app as well as changes caused by the tool. We used an inductive category formation [Mayring, 2014], where categories were built bottom-up from the participants' suggestions (multiple answers possible). Maximum two levels of categories were formed to meet the requirements of exclusiveness and comparable degrees of abstraction. For an efficient presentation of results, answers were summarized across repeated times of measurement. A second, independent coder was included to ensure reliability. Intercoder-reliability amount to κ = .76 for usage context, κ = .78 for advantages, κ = 1.00 for disadvantages, and κ = .90 for perceived changes. Therefore, we met the criteria [Landis and Koch, 1977] of κ = .75 indicating an "*excellent*" agreement. The remaining discrepancies have been eliminated by a third coder. To identify the most common answers, relative response frequency per category of this final solution is reported. We elaborate on the most frequently named category by describing second-order categories and/or an example citation.

5. Sample

Recruitment and Selection of Participants

We used multiple channels to disseminate the invitation to the field trial and released it to the test participant panel of our professorship, on the university and research project website. To diminish self-selection, the purpose of the field trial was described as an "app usability test" and a remuneration of 50€ (10€ for the completion baseline phase; 40€ for the trial period) was offered. The invitation contained a registration link, which assessed demographics (age, gender), smartphone, and app usage data (operating system, amount of installed, and used apps). It further contained mandatory requirements: the usage of Android as an operating system and the willingness to participate. We received 55 completed registrations and selected those for the field trial, who used at least Android 6.0 since a selective permission granting was implemented from this version on. Furthermore, we considered gender balancing. Afterwards, we sent a confirmation to the selected applicants and a code to pseudonymize survey data.

Demographics

The baseline phase started with 27 participants, and *N* = 26 (*n* = 12 female) completed it. Those, who started the trial phase, had a mean age of *M* = 34.35 (*SD* = 7.79; *Mdn* = 31.5; min = 24.00; max = 61.00), which was younger than the German population (*Mdn* = 46.2; [United Nations, 2015]). The majority of our participants (65%) could be assigned to the age group ranging from 30 to 49 years. In the German population, almost all individuals in this age group (97%) use smartphones [Bitkom, 2017]. Most frequently, they hold a completed apprenticeship (31%), which was above national statistics ([Statistisches Bundesamt Deutschland, 2015]; 24.5%) and the majority were employed full-time (69% slightly above the German population, 61%, [Statistisches Bundesamt Deutschland, 2022]). At least, 22 persons completed the whole field trial (dropout rate 15%).

Smartphone-App Usage

Our participants reported having *M* = 68.62 (*SD* = 33.95; min = 22; max = 149) apps installed, with a high individual variation. Compared to German smartphone users ([YouGov, 2017], 8%), persons with more than 50 apps were clearly overrepresented in our sample (69%). In accordance with national usage data ([Von Rauchhaupt, 2017]; 112.5 min), our participants estimated to use smartphone apps more than two hours per day (*M* = 137.65 min; *SD* = 112.26; min = 4; max = 480), again with a high individual deviation.

We asked our participants to rate their knowledge [Karrer et al., 2009] about the use of smartphone apps ranging from 1 = "*does not apply at all*" to 5 = "*fully applies*". They rated it (*M* = 4.14; *SD* = 0.60) significantly higher ($t(25) = 5.76$, $p < .001$, $d = 1.13$) as the comparison sample (*N* = 460; *M* = 3.47; *SD* = 0.85; [Karrer et al., 2009]). On a 7-point scale (1 = "*hourly*" to 7 = "*never*"), we asked to rate the usage intensity of different app groups. Messenger (*Mdn* = 1.00), followed by entertainment (*Mdn* = 2.00), social network (*Mdn* = 2.00), and weather (*Mdn* = 2.00) apps were used most intensively. This is comparable available to U.S. statistics [comScore, 2017], indicating that messengers have the highest share of mobile app minutes (96%).

Privacy Concerns
We used the Internet Users' Information Privacy Concerns (IUIPC) questionnaire [Malhotra et al., 2004] to capture our participants' concerns about information privacy. The comparison to available reference values (N = 449; M = 6.06; SD = 0.95; [Malhotra et al., 2004]) revealed that their ratings (M = 6.55; SD = 0.62) only differed significantly ($t(25) = 4.01$; $p < .001$; d = 0.79) on the scale improper access. Thus, our participants indicated high concerns about data access by unauthorized third parties. Furthermore, we asked our participants to specifically rate their mobile privacy concerns (MUIPC; [Xu et al., 2012]). On a 7-point agreement scale ranging from 1 = "*not at all*" to 7 = "*completely*" they on average "*somewhat agreed*" on the perceived surveillance (M_{PS} = 5.35; SD_{PS} = 1.24), perceived intrusion scale items (M_{PI} = 4.80; SD_{PI} = 1.32); and fearing the secondary use of data arising by the usage of mobile apps (M_{SU} = 5.17; SD_{SU} = 1.53). Due to our previous experiences with this questionnaire, these values appear to be within an average range.

Summarizing, our N = 26 participants were younger, hold a higher proportion in apprenticeship occupation, and are comparably employed to the German population. Concerning smartphone app usage, we have more users with many apps installed and they consider themselves very competent. Their daily app usage time was representative as well as their concerns online or about mobile privacy.

6. Procedure

Our field trial was conducted in Germany and lasted six weeks, comprising a two-week baseline and a four-week trial period. The entire procedure was reviewed and approved by the faculty's ethics committee in advance. The participants were asked to complete seven weekly online surveys (T0 – T6, implemented with *LimeSurvey 2.72*). Most surveys took less than half an hour and included different dependent variables (Appendix A.1). The transition from the baseline to the trial phase was realized face-to-face at our laboratory to ensure the correct installation of our tool, and the payment of the first remuneration. This individual appointment affected the timing of the surveys.

The aim of the initial baseline phase (T0 – T2.1) was to assess "unaffected" app usage behavior (un- /installations, permission function usage) and compare it with the trial period (T2.2 – T6) to identify the effects of our tool. The app-list-reporting was demanded one day ahead of the weekly survey. This reporting and questions about permission granting or withdrawal stayed the same during baseline and trial. The trial period ended with a comprehensive survey (T6), which took up weekly variables and those that were first assessed in the baseline as well as a self-assessment of behavioral changes. Finally, the purpose of the study was clarified, and the second remuneration was paid.

7. Material

Independent Variable: The AndProtect App
The independent variable of our field trial represents absence or presence of our tool - the "AndProtect app", described below. It aims to overcome hurdles in informed privacy decisions by increasing transparency about app behavior. The overall design followed a user-centered design process [Döbelt & Halama, 2018], [Döbelt et al., 2020].

User Interface
Our app had a three-level content structure (Appendix A.2, Figure 9) and the content was presented in German. On the initial start page (Figure 9 a), all installed apps of the user are listed. The top pie chart provides information about the amount of installed and analyzed apps as well as the overall risk assessment, represented by a number (between 0 = no and 100 = maximum risk). In addition, the outer circle entails the proportion of apps with a certain risk category: green = low, yellow = medium, red = high-risk potential, black = malicious and grey = no result available yet. The list below can be sorted by tapping on a certain color. The user can select the grey bar below, to initiate an upload and analysis of an app (Figure 9 b). After an analysis is completed, a push notification is sent. At the lower part of the start page, app-specific information is provided: app name, group (entertainment-, map/navigation-, messenger, weather, other app), risk value, app icon, and a color assignment.

By clicking on a row, a user retrieves detailed information in an app report (see Figure 9 c). Each contained analysis information about which data the app accessed. Here, 11 types of data are considered (data types not accessed are grayed out). By clicking on a colored icon, the analysis results are described in text form (Figure 9 d), e.g., whether the access took place while the app was actively used or in background mode or URLs that communicated with the app and classified as dangerous. The encryption section contains, whether unencrypted data packets were identified during data transfer. At the end of the report page, a button led to the uninstallation of the analyzed app. Furthermore, users had the opportunity to individualize risk valuation (Figure 9 e) by shifting a slider. We derived the default settings for these risks from a survey conducted in preparation for the tool design [Döbelt et al., 2020].

Risk Score Design
During the user-centered design process of our tool, users demanded a simple and quick valuation of the risk, displayed

in traffic light colors or a numerical form [Döbelt et al., 2020]. To provide such a valuation, multiple variables from the app analyses results were included and additively linked:

- The mode of use of a particular type of data: when interacting with the app and/or in the background
- Appropriateness of using a data type for a specific app group: not, partially, or appropriate
- Frequency: timing of data transfer in min and sec
- Recipient of the data: servers classified as dangerous or not dangerous
- Sending of the data: encrypted or unencrypted transport

Once these variables and their composition had been defined, a weighting of the variables was carried out based on expert- and user assessments. For example, the variable "type of data used in the background" was weighted fourfold, since it was rated as to be particularly risky in the preliminary investigation [Döbelt et al., 2020]. The additively and weighted variables resulted in a risk score between 0 and 100 for the respective app and averaged for all installed apps. This numerical value X was transferred into colors ($0 < X < 25$ = green; $25 \leq X < 65$ = yellow; $65 \leq X < 95$ = red; $X \geq 95$ = black) used within the start page pie chart, app list, and report page.

Technical Implementation

Our tool uses a two-step analysis approach, combining static and dynamic analysis. The static analysis performs an inter-procedural, field- and inheritance-sensitive data flow analysis. Thereby, control flow graphs are created, which enable determination points of data entry and outflow. Furthermore, information defined in the code can be extracted about an app before it is installed. For our tool the static analysis delivered information on suspicious URIs, confidential information sources (e.g., IMEI, location, and current Wi-Fi information), the app version, as well as signatures indicating malware. These results are supplemented by a subsequent dynamic analysis in which further investigations are carried out in a runtime environment. Here, for example established communication, and the content of the communication can be examined. We used the dynamic analysis to determine the type of data, its transmission interval and encryption, as well as the type of access to the data in the foreground or background. *TaintDroid* [Enck et al., 2014] was used to record data flows of sensitive data types, such as location, calendar, contact, or camera. An app to be analyzed was installed and operated by a tester, who identified, executed, and logged an app function. The network traffic was recorded in parallel to evaluate the encryption usage. Results of both static and dynamic analysis are processed and logged in a central database. From the server side, these results were pushed and the risk score calculation as well as the classification into the colored risk categories was carried out on the client (app) side.

Analyzed server data revealed that 136 completed reports about different app versions were sent to the participants during the field trial phase. Whereby participants only received a report, if they had installed the analyzed app. The mean risk score of these reports amounts to $M = 27.46$ ($SD = 16.11$). After the most frequent category "yellow" (49.3%), reports second frequently were categorized as "green" (42.5%). The majority of reports analyzed apps were from the category "other" (59.6%), followed by "entertainment" (17.6%). A pushed report had an average recipient range of four people with strong fluctuations ($M = 4.01$, $SD = 5.37$, min = 1.00, max = 25.00). At the beginning of the field test (T2.2), 41.9% of all conducted app analyses were available to the participants. At subsequent times of measurement, reports increased by 23.4% (T3), 21.0% (T4), 8.1% (T5), and 5.6% (T6). The absolute report frequency during the field trial varied from min = 6.00 to max = 33.00 per participant.

Assessments of Dependent Variables

From the weekly monitoring of the number of installed apps, we could identify weekly uninstallations (H1.2) and installations (H1.3) from the second time of measurement on (T1 –T6). To realize the app reporting, we provided an automated or written sending option.

We further asked closed-ended, single-choice questions to get insights into reasons for installations and uninstallations. Participants could choose between 1.) "*I do not need the app anymore.*"/"*I need the app.*"; 2.) "*I have installed another app that serves the same purpose.*"/"*I needed an alternative to another app.*"; 3.) "*New information about the app caused me to uninstall it.*"/"*New information about the app made me install it.*"; 4.) "*Other*". As 3.) was of special interest to us, we presented a follow-up, single-choice question for specification: 1.) "*Recommendation from friends or acquaintances*", 2.) "*Description in technical media*", 3.) "*Description in Google Play Store*", 4.) "*Rating in Google Play Store by other users*", 5.) "*Download numbers of an app in Google Play Store*", 6.) "*App analytics that provide information about the risk of an app*", 7.) "*Other*".

To test H2.1, we asked: "*Since Android version 6.0, users have the possibility to switch individual app permissions (e.g., camera, location) on and off. Are you familiar with this function and have you already used it to withdraw permissions from an app?*" to assess participants' knowledge about the permission function at the beginning (T0) and the end of the field trial (T6). Participants could rate their knowledge ranging from 1 = "*Yes, I use the function regularly*", 2 = "*Yes, I know the function but rarely use it.*", 3 = "*Yes, I know the function, but I don't use it",* and 4 = *"No, I don't know and I don't use the function".* Starting with the second survey (T1-T6) and testing

H2.2 and H2.3, we asked our participants to indicate how many apps they granted or withdrawal permissions during the last week. To enrich behavioral data, we asked the participants open-ended to evaluate, whether they noticed any single or multiple changes at the end of the field trial (T6): "*Did using the AndProtect app during the field trial lead to any changes in your app usage behavior? Please briefly justify your answer.*"

Furthermore, we investigated UUX evaluation of our tool (H3.1, H3.2) with established questionnaires. We repeatedly (T2.2, T4, T6) used the System Usability Scale (SUS; [Brooke, 1996]). The participants were asked to indicate their agreement (from 1 = "*strongly disagree*" to 5 = "*strongly agree*") to 10 statements resulting in the overall SUS score (0 - 100), which allows results to be classified, using grades from "A+" to "F" [Lewis and Sauro, 2018]. We chose a score of 68 (grade "C") as a benchmark, which is at the center of the curved grading scale. Additionally, we used (T2.1, T4, T6) the User Experience Questionnaire (UEQ; [Laugwitz et al., 2008]). The 26 items contribute to six scales: attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. Answers are captured via a 7-stage semantic differential (ranging from -3 = fully agree to negative to +3 = fully agree to positive term). We used the minimum for *"below average"* evaluations [Schrepp et al., 2017] as our tool was developed within a research project and its true purpose was not disclosed beforehand.

To answer our second research question (suitability for everyday use), we gathered information about the usage context of the AndProtect app. Therefore, we repeatedly (T3, T4, T5, T6) asked: "*In which situations have you used the AndProtect app? Briefly describe the usage situations in your own words.*" The open-ended format allowed for multiple answers likewise, the questions about dis- and advantages (T2.2, T5, T6; "*What are the [dis-/advantages] of the AndProtect app for you personally?*").

8. Results

Number of Installed Apps, Installations, and Uninstallations (H1)

We were able to monitor the number of installed apps from $N = 21$ participants continuously (T0 – T6). Averaged over the entire study, participants had $M = 65.42$ apps installed ($SD = 31.09$; *min* = 21.00, *max* = 126.71), slightly more during the baseline ($M = 66.33$; $SD = 31.55$) compared to the field trial phase ($M = 64.73$; $SD = 30.85$). To identify statistical differences between the single times of measurement (H1.1; Figure 1), we run a repeated-measurements ANOVA.

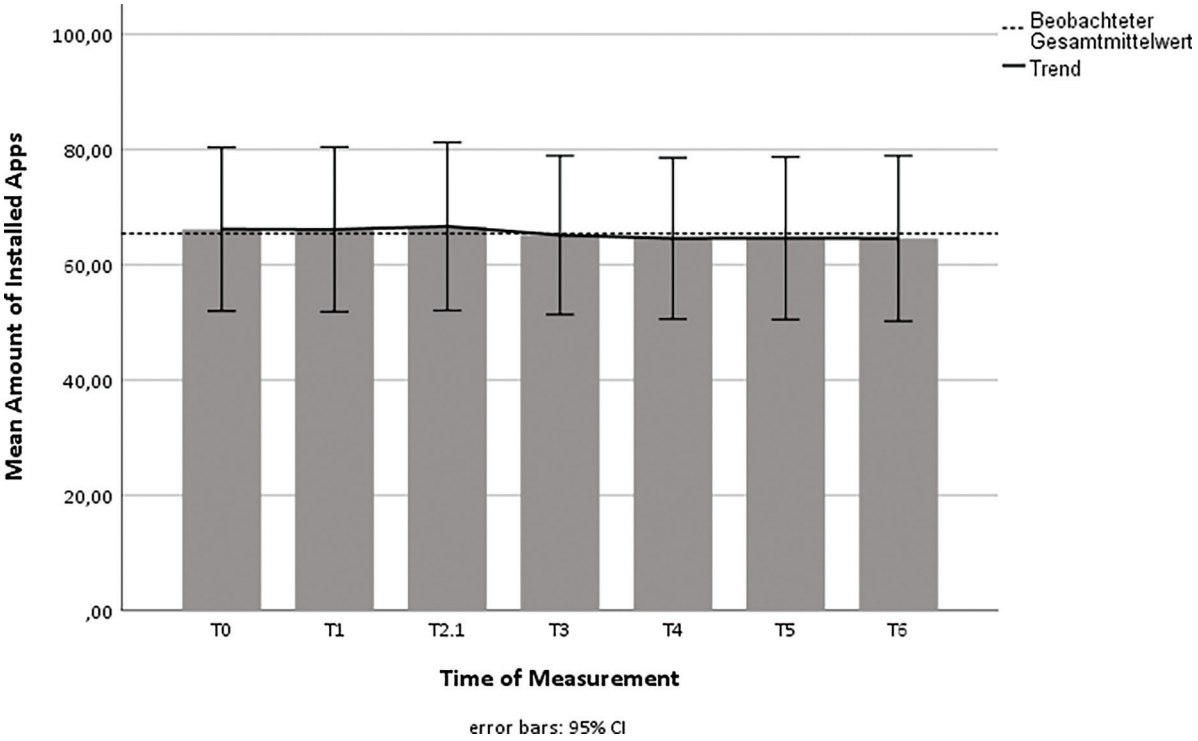


Figure 1. Number of installed apps from $N = 21$ participants for all times of measurements, * highlights a post-hoc significant difference ($p < .05$).

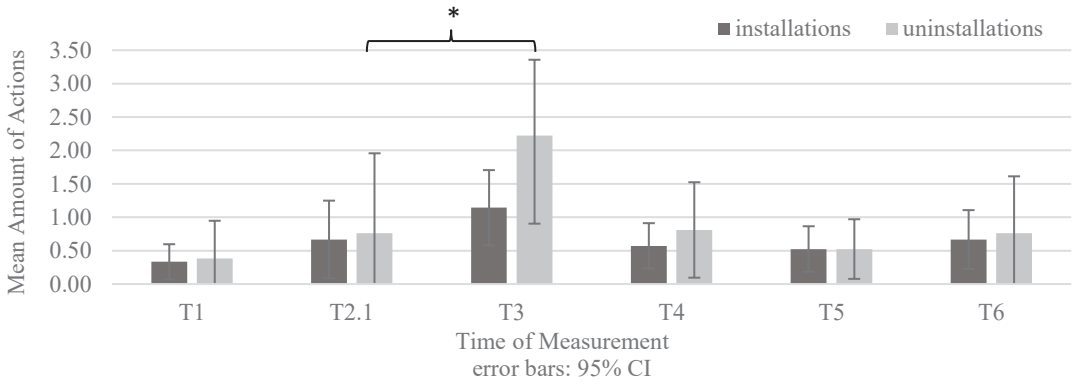


Figure 2. Mean number of installations and uninstallations of $N = 21$ participants; *marks a significant difference ($p < .05$).

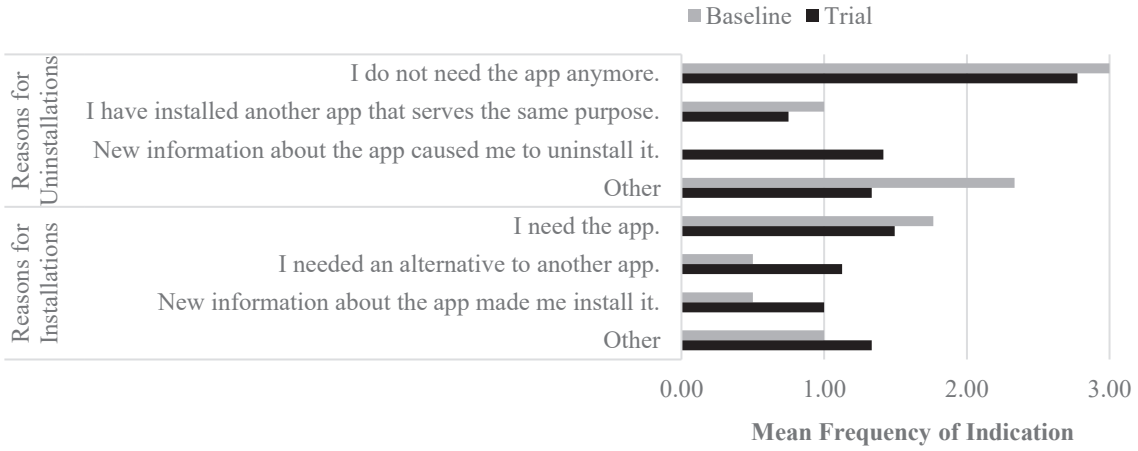


Figure 3. Mean frequency of indicated reasons for uninstallations and installations during baseline (T1-T2.1, $n = 29$ answers) and trial phase ($n = 85$ answers; T3-T6).

The assumption of sphericity was violated ($\chi^2(20) = 141.14$, $p < .001$, $\varepsilon = .28$), therefore a GG correction was applied.

We could not clearly identify a significant main effect for the factor time ($F(1.57, 31.34) = 2.99$, $p = .076$, $\eta_p^2 = .13$). Repeated contrasts revealed that one difference between T2.1 (end of baseline) and T3 (start of trial period) was statistically significant ($F(1, 20) = 3.68$, $p = .034$; one-tailed; $\eta_p^2 = .15$). Despite the marginality of significance, the effect sizes, we found were rather large [Cohen, 1988].

To test our hypotheses that uninstallations will rise (H1.2) and installations will be constant (H1.3), we run the same procedure. The figure below (Figure 2) depicts the temporal course of in- and uninstallations. For the uninstallations the assumption of sphericity was barely met ($\chi^2(14) = 21.94$, $p = .082$, $\varepsilon = .87$). Again, we could hardly identify a statistically significant main effect for times of measurement ($F(5, 100) = 2.29$, $p = .051$, $\eta_p^2 = .10$). Using repeated contrasts, we could identify a significant effect for T2.1 vs. T3 ($F(1,20) = 3.27$, $p = .047$, one-tailed, $\eta_p^2 = .14$) again, with a large effect size [Cohen, 1988]. The assumption of sphericity was violated ($\chi^2(14) = 33.29$, $p = .003$, $\varepsilon = .77$) regarding installations. The GG corrected results of the repeated measurements ANOVA showed, that we could not identify a significant main effect for the factor time ($F(3.18, 63.61) = 2.04$, $p = .114$).

We asked further for respective reasons and aggregated data for the baseline and the field trial phase (Figure 3). The most frequent reason for both was "*I [do not] need the app [anymore]*", which underlines the service-oriented motives for these actions. However, "*new information about the app*" was the second most frequently named reason for an uninstallation during the trial phase.

The follow-up question revealed that "*app analytics*" of our tool has been the most frequently named source of information during the trial period causing uninstallations, followed by "*Other*" reasons such as: "*reflection about the usage frequency of specific apps*", which was also prominent

in the baseline phase. However, the number of participants (n = 3) who indicated the AndProtect app as an information source for uninstallation was rather small.

Knowledge and Usage of Permission Function (H2)

The participants (N = 22) most frequently indicated to "*know the function but rarely use it*" (T0: n = 10, 46%; T6: n = 11, 50%), followed by "*know the function, but do not use it*" (T0: n = 6, 27%; T6: n = 4, 18%), "*use the function regularly*" (T0: n = 4, 18%; T6: n = 4, 18%). Only a few participants stated that they "*do not know nor use the function*" (T0: n = 2, 9%; T6: n = 3, 14%). We tested the statistical difference between the two times of measurement (T0 and T6; H2.1) using the Wilcoxon-signed-rank test. Results revealed that answers did not change (Mdn_{T0} = 2.00, Mdn_{T6}: 2.00, z = 0.00, p = 1.00).

In Figure 4 the number of apps affected weekly by granting and withdrawal is depicted. Normal distribution was also violated for every time of measurement for the permission usage (granting + withdrawal): $D_{T1}(21) = 0.38$, $p < .001$; $D_{T2.1}(21) = 0.45$, $p < .001$; $D_{T3}(21) = 0.50$, $p < .001$; $D_{T4}(21) = 0.36$, $p < .001$; $D_{T5}(21) = 0.47$, $p < .001$; $D_{T6}(21) = 0.44$, $p < .001$. Therefore, we used Friedman's ANOVA to test H2.2. Results revealed no significant effect ($\chi^2(5) = 5.88$; p = .318) for the time of measurement.

In order to test H2.3, we run the same procedure for the withdrawal of permissions. Again, normal distribution was violated for every time of measurement: $D_{T1}(21) = 0.50$, $p < .001$; $D_{T2.1}(21) = 0.52$, $p < .001$; $D_{T3}(21) = 0.54$, $p < .001$; $D_{T3}(21) = .048$, $p < .001$; $D_{T5}(21) = 0.51$, $p < .001$; $D_{T6}(21) = 0.52$, $p < .001$. Again, results revealed no significant effect ($\chi^2(5) = 4.58$; p = .470) for the factor time.

Self-Assessment Behavioral Change

At the end of the field trial (T6), we further asked the participants, whether they noticed behavioral changes in the usage of apps. We received 26 statements from 20 participants that included single descriptions. Relative frequencies indicate that in total about to thirds (69%) of statements described any change compared to one-third (31%) of statements that reported "*no change*". The changes described could again be distinguished into five different categories (Figure 5): 1. "*More precise inspection of permissions that apps*

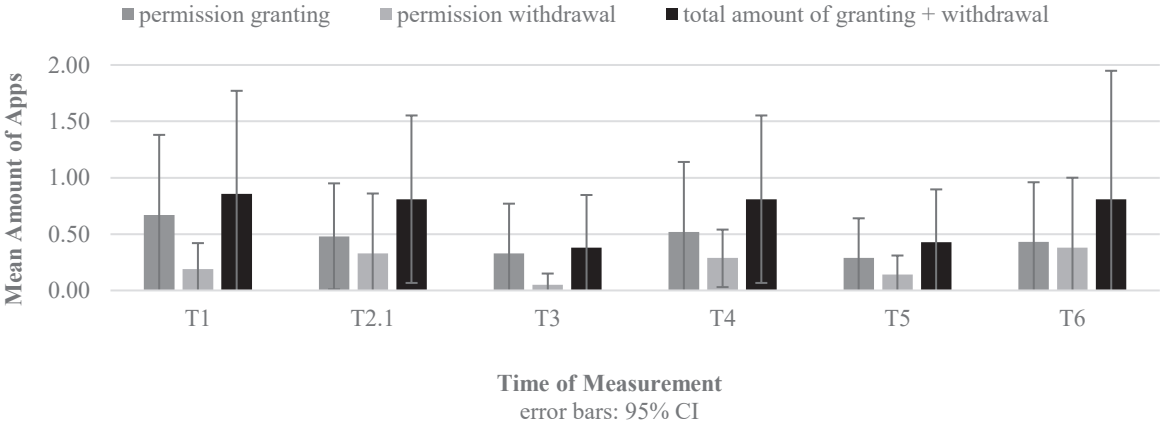


Figure 4. Mean number of apps affected by granting, withdrawal, and mean total amount of actions (granting + withdrawal) of N = 21 participants; T2.1 marks transition from the baseline to field trial phase.

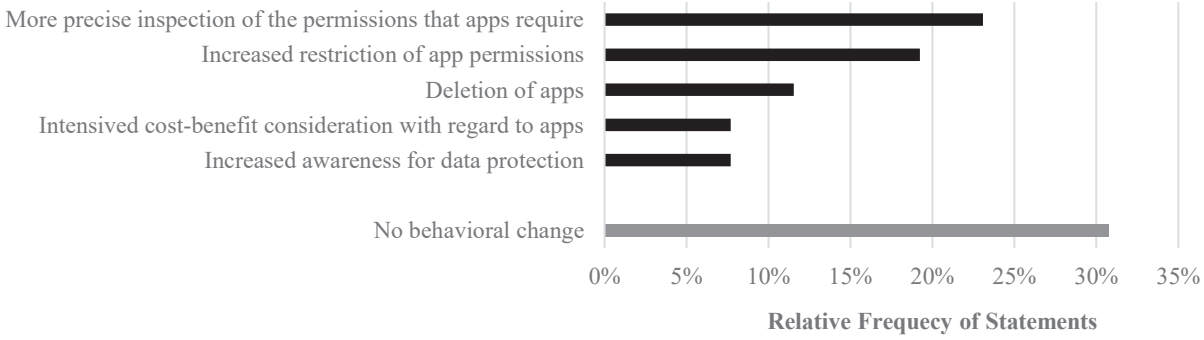


Figure 5. Categorized n = 26 statements describing behavioral changes during the field trial.

require" (23% of all statements), 2. "*Increased restriction of app permissions*" (19%), 3.) "*Deletion of apps*" (12%), 4.) "*Intensified cost-benefit consideration with regard to apps*" (8%), and 5.) "*Increased awareness for data protection*" (8%). One example statement for the most frequently named change (precise inspection app permissions) was: "*Yes. I went through many apps and checked what permissions they have* [...]".

UUX Evaluation (H3)

Usability

Descriptive statistics revealed that the SUS results can be classified [Brooke, 1996] ranging from "A" to "B" for the three measurements ($M_{T2.2}$ = 82.84, $SD_{T2.2}$ = 10.75; M_{T4} = 79.20, SD_{T4} = 12.50; M_{T6} = 76.81, SD_{T6} = 16.48). As the score values met the requirement of normal distribution, we tested (one-tailed) differences between the benchmark of 68.00 and our values (H3.1). For all times of measurements, we discovered significant differences with large to mean effect sizes (T2.1: $t(21) = 6.47$, $p < .001$; $d = -1.38$; T4: $t(21) = 4.21$, $p < .001$, $d = -0.88$; T6: $t(21) = 2.51$, $p = .010$, $d = -0.53$), indicating that our AndProtect app was evaluated above average during the field trial phase (Figure 6).

To investigate the temporal development (H3.2) of the usability evaluation, we ran a repeated-measurement ANOVA. As Mauchly's test of sphericity was violated ($\chi^2(2) = 10.23$, $p = .006$, $\varepsilon = .71$), we applied a GG correction. We found a significant difference with a large effect for the factor time ($F(1.43, 29.99) = 4.23$, $p = .036$, $\eta_p^2 = .17$). Using simple contrasts, the SUS score at the beginning (T2.2) differed from the value at the end of the trial (T6: $F(1, 21) = 5.44$, $p = .030$, two-tailed, $\eta_p^2 = .21$) with a large effect. Figure 7 illustrates decreasing usability but still, the rating was on an above-average level.

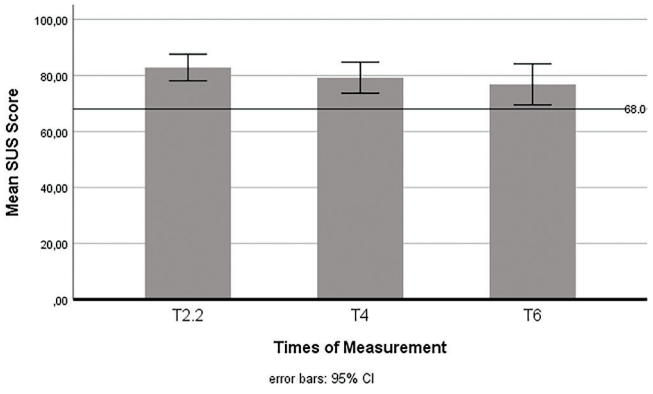


Figure 6. Repeated SUS Scores evaluating mean usability of the AndProtect app during the field trial phase (T2.2; T4; T6); * highlights statistical differences ($p < .05$) of post-hoc tests; $N = 22$

User Experience

For the field trial phase (T2.2, T4, T6), the majority of UEQ scales met the requirement of normal distribution. Descriptive data varied between T2.2: M_{persp} = 2.34, SD_{persp} = 0.61 and M_{nov} = 0.90, SD_{nov} = 1.19; T4: M_{persp} = 2.03, SD_{persp} = 0.88 and M_{nov} = 1.01, SD_{nov} = 1.01; T6: M_{persp} = 2.00, SD_{persp} = 0.99 and M_{nov} = 0.97, SD_{nov} = 1.31; Table 1). To test if our participants' rating was above average (H2.1), we tested (one-tailed), if UEQ scale values differed from minimum scale values for "*below average*" evaluations [Schrepp et al., 2017]. Results (Table 1) revealed that the user experience evaluations at any time of measurement except stimulation at T4 differed (Bonferroni corrected) significantly from the benchmark, indicating at least an average evaluation.

To inspect temporal development (H3.2, Figure 7), we ran six repeated-measurements ANOVAs for each UEQ scale. Mauchly tests of sphericity were violated for the scales attractiveness ($\chi^2(2) = 17.10$, $p < .001$, $\varepsilon = .66$) and dependability ($\chi^2(2) = 7.40$, $p = .025$, $\varepsilon = .81$). Therefore, a GG correction was applied. For most of the scales, we could not find a significant time effect ($F_{attr}(1.27, 26.67) = 0.52$, $p = .518$; $F_{nov}(2, 42) = 1.19$, $p = .829$; $F_{stim}(2, 42) = 1.40$, $p = .259$; $F_{dep}(1.53, 32.08) = 0.68$, $p = .475$, $F_{eff}(2, 42) = 2.10$, $p = .136$). Only for the scale perspicuity time was hardly significant ($F_{persp}(2, 42) = 2.83$, $p = .071$, $\eta_p^2 = .12$) with a medium effect size, but Bonferroni corrected post-hoc comparisons did not revealed significant differences between the times of measurement. Thus, we could not find temporal changes regarding aspects of user experience.

Suitability for Everyday Use

To answer our second research question, we asked multiple open-ended questions concerning the usage context, advantages, and disadvantages of our tool.

Usage Context

Across all times of measurements (T3, T4, T5, T6), we received descriptions of 74 usage situations from 23 participants. Our participants used the AndProtect app most frequently (47% of answers) because of "*interest in content and functions*", which means without any specific trigger (e.g., trying a function or usage during pastime). For instance, one participant stated: "[I used it] *in the evening while watching TV.*" In contrast, 27% of answers indicated that the participants used the tool to "*test a specific app*" (27%) and when a "*push notification about a new report*" arrived (20%).

Advantages and Disadvantages

We further asked for the dis-/advantages of the tool (T2.2, T4, T6) and received 113 statements from 26 participants that included advantages. The majority (96%) of field trial participants´ statements included specific advantages

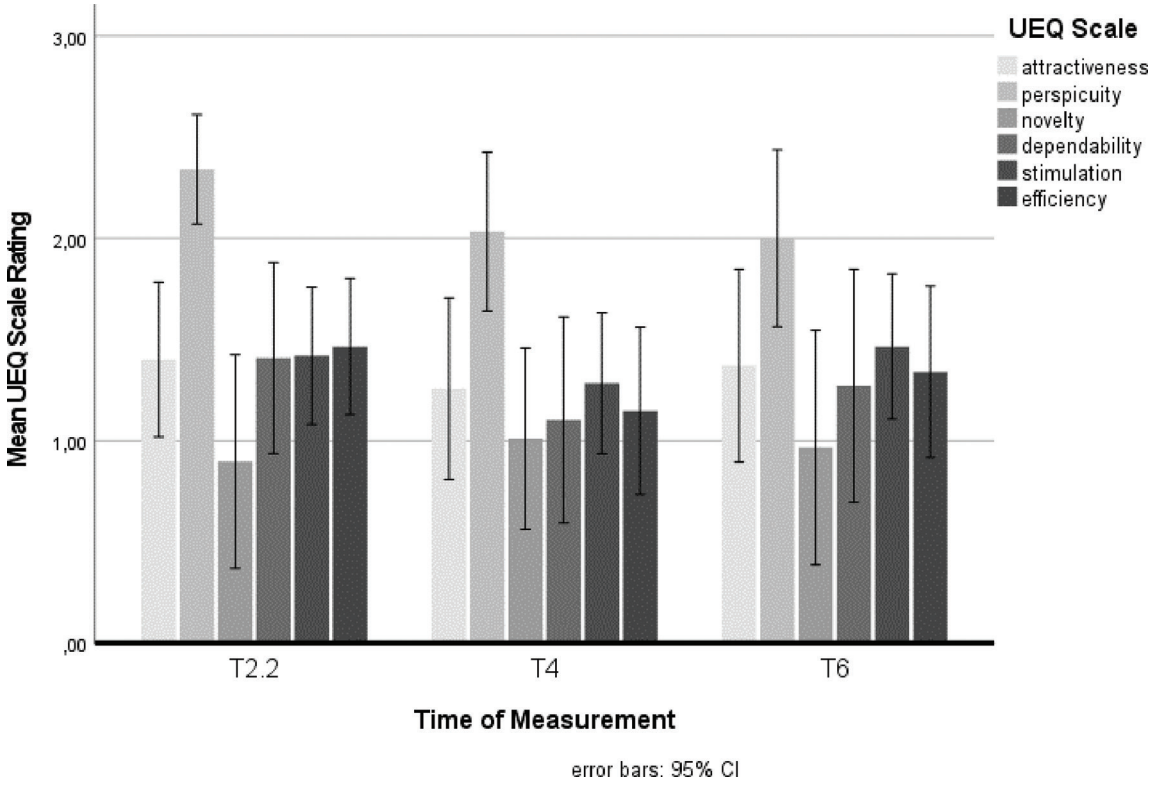


Figure 7. Repeated UEQ scales measurements evaluating aspects of user experience of the AndProtect app during the field trial phase (T2.2; T4; T6); *N* = 22.

Table 1. Results of one-tailed t-tests, test value "below average" [Schrepp et al., 2017], significance level Bonferroni corrected: *p* = .017 Attr = Attactivenss; Persp = Perspicuity; Nov = Novelty; Stim = Stimulation; Dep = Dependability; Eff = Efficiency; *N* = 22.

UEQ Scale	Test-Value	Differences								
		T2			T4			T6		
		M; SD	*p*	*d*	*M; SD*	*p*	*d*	*M; SD*	*p*	*d*
Attr	0.70	1.40; 0.86	<.001	0.81	1.26; 1.01	.009	0.55	1.37; 1.07	.004	0.63
Persp	0.64	2.34; 0.61	<.001	2.79	2.03; 0.88	<.001	1.58	2.00; 0.99	<.001	1.38
Nov	0.30	0.90; 1.19	.014	0.50	1.01; 1.01	.002	0.70	0.97; 1.30	.013	0.51
Stim	0.50	1.41; 1.06	<.001	0.86	1.10; 1.15	.023	0.52	1.27; 1.30	.011	0.60
Dep	0.78	1.42; 0.77	<.001	0.28	1.28; 0.78	.003	0.10	1.47; 0.81	<.001	0.20
Eff	0.54	1.46; 0.76	<.001	1.22	1.15; 0.93	.003	0.65	1.34; 0.96	<.001	0.84

and only very few (4%) included "*no advantages*". Our tool was specifically appreciated for the "*provided information and functionalities*" (39% of all statements). This category is composed of statements referring to e.g., the provided information about specific apps and the risk assessment. For instance, one participant stated: "*It informs about what data each app collects.*" Additionally, the field trial participants appreciated "*behavioral suggestions and implications*" (29%), followed by the "*presentation and preparation of information*" (28%; Figure 8).

In contrast, we received 83 statements from 26 participants describing disadvantages. The category forming resulted in 5 second-order categories. Two-thirds (69%) of the statements included specific disadvantages, while one-third (31%) could identify "*no disadvantages*". Criticism was primarily related to functions and demanded their "*improvement*" (33 % of all statements). This category is composed of statements referring to, e.g., too few explained options for action and the long wait for the report. One participant stated: "*No concrete options* […] *other than uninstallation.*" Further disadvantages

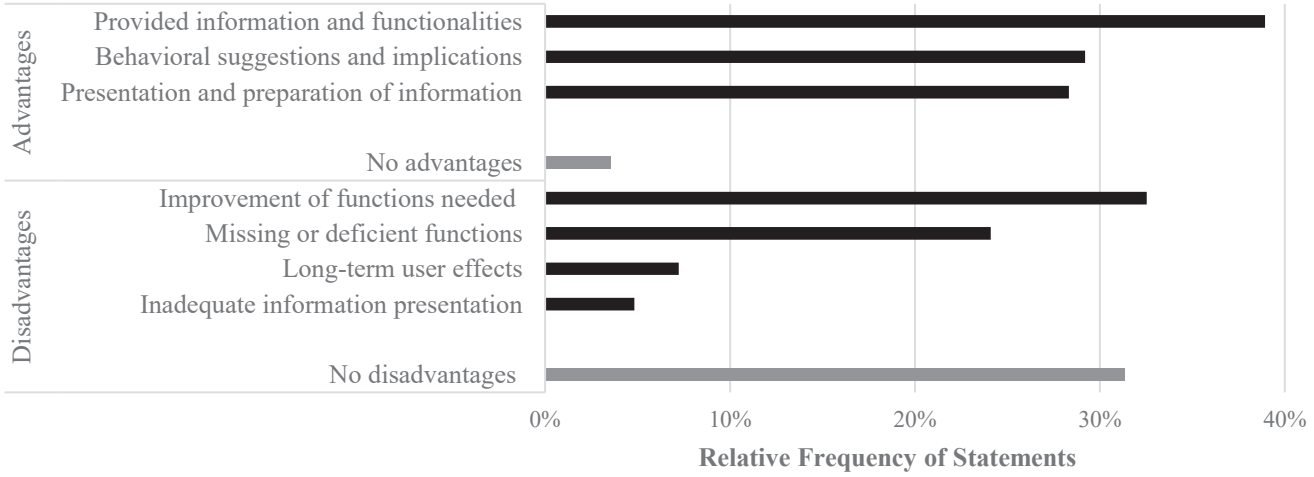


Figure 8. Categorized statements for advantages (upper part, *n* = 113) and disadvantages (lower part, *n* = 83) of the AndProtect app; summarized for T2.2, T4, and T6.

were "*missing or deficient functions*" (24%), followed by negative "*long-term user effects*" (7%), and a few statements (5%) referred to "*inadequate information presentation*" (Figure 8).

9. Summary and Discussion

The research questions of our six weeks lasting field trial were: How does our tool affect users' app usage behavior and how suitable is it for everyday life? For the usage behavior, we hypothesized a decrease in installed apps (H1.1), caused by increased uninstallations (H1.2) and a constant number of installations (H1.3). Furthermore, we assumed that knowledge (H2.1) and usage (H2.2) of the permission function will increase as well as the frequency of withdrawal of permissions (H2.3).

For the installed apps, we could not identify a significant but medium-sized main effect for times of measurement. Post-hoc tests revealed a weakly significant drop from baseline to a trial period with a large effect size. We found similar results for the increase in uninstallations. Due to the large effects and despite the marginality of significance, we thus could support H1.1 and H1.2. The reasons for uninstallations revealed that firstly service-oriented motives and secondly new information provided by our tool caused the uninstallations. Furthermore, the hypothesis H1.3 is valid, as we could not discover any temporal effects concerning installations. In summary, our tool caused a "clean-up"-effect.

However, with a larger and more homogeneous sample, these effects would have become significant. As we were not able to identify comparable effect sizes in the literature beforehand, we composed our sample within the scope of the organizational capacities. Our sample was younger, more competent about apps, and users with many apps were overrepresented. However, our sample is representative regarding daily usage time of apps, favorite app groups, and privacy concerns. Thus, it is a target group for our tool. Although we provided a remuneration, the sample is certainly not free of self-selection and drop-outs may have increased this bias. Attrition rates of 20% are known to pose threats to the validity of randomized, controlled trials [Norvell et al., 2016]. For field studies using survey instruments comparable to ours, attrition rates tend to be even higher [Garcia, 2013]. Therefore, we consider our drop-out rate (15%) as not particularly alarming in terms of biasing the results.

Additionally, it must be noted that the decrease in installed apps and the increase in uninstalls is only short-term in nature and does not last beyond T3. Thus, the long-term affordance of our tool seems to be limited. This could be because either the "cleaning up" for a short time already exhausts the participants' behavioral scope, the mean risk score of the analyzed applications was in the medium range, and/or the presented information was in line with the participants' privacy needs. Furthermore, our tool neither provokes an increase in knowledge nor the usage of the permission function. Descriptive data showed that most frequently our participants stated to "*know the function but…rarely use it*" followed by "…*do not use it*" at the beginning and the end of the trial. This underlines previous findings [Kelley et al., 2013], [Felt et al., 2012], [Joeckel and Dorgruel, 2020], that users rarely engage with the permission function and our tool could not provoke any significant change. Likewise, the withdrawal of permissions did not differ between the baseline and trial. Therefore,

we must reject H2.1, H2.2, and H2.3. The data varied considerably across participants, but the mean amount of affected apps was rather small. We attribute this lack of impact to the fact that (in contrast to the uninstallation possibility), there was no explicit behavioral advice or function regarding using or restricting permissions.

In contrast, the self-assessment of behavioral changes revealed that two-thirds of statements described a change most often about the attention and reasoning level e.g., permissions apps require, costs/benefits of apps, or data protection in general. According to Van Kleek [Van Kleek et al., 2017], these considerations form a sound basis for confident data protection and privacy-related decisions. But also, restrictions of permissions and deletion of apps have been mentioned. For these self-assessments, an influence of the study situation cannot be excluded. At the end of the trial, its purpose may have become clear to the participants, and they may have answered in a socially desirable [DeMaio, 1984] way.

Concerning the suitability of the AndProtect app, we hypothesized that UUX evaluations will be above average (H3.1) and change over time (H3.2). Both measurement comparisons with respective benchmarks led to a positive evaluation: for usability above grade “C” [Lewis and Sauro, 2018], for user experience above a “*below average*” benchmarks [Schrepp et al., 2017]. Therefore, we can support H3.1. Since our tool was developed within a research project, the applied benchmarks are moderate. In a market-ready application, the benchmarks would have to be more demanding.

Regarding the temporal development of UUX, our results revealed the usability evaluations declined. In contrast, the user experience did not change. Therefore, H3.2 could only be partially accepted. This result is in contrast to earlier research, which found that the usability of smartphone evaluations in- and the user experience decreases over time [Kujala et al., 2011]. Reasons could be differences in the chosen methodology (qualitative vs. quantitative) or the subject of the study (product smartphone vs. app prototype). Furthermore, our user experience ratings could be biased by the framing of the study. Our participants did not receive any explanation of the apps’ aim in advance to avoid self-selection. The flip side of this could be that they could not build up user experience-relevant expectations.

The descriptions of the usage situations revealed that the participants used our tool due to motivation and interest. This result could be overestimated, caused by the study procedure with its weekly surveys. Participants stated more advantages than disadvantages of the AndProtect app, especially regarding its information content, presentation, and behavioral implications. Thus, our tool seems relevant and suitable for everyday use. Nevertheless, many possibilities for improvement regarding extended information, functionalities, and behavioral suggestions were mentioned.

10. Limitations and Future Work

Our field trial provided valuable insights into the use and behavioral implications of a tool that increased the transparency of app behavior. However, there is room for improvement, especially concerning the design of the tool. We tested a ‘one-size-fits-all’ solution and the risk assessment was mainly based on expert opinions. Although the individualization function allowed participants to adjust the apps’ settings, a more comprehensive alignment may further increase the effectiveness regarding behavioral change [Liu et al., 2016]. For this purpose, a practicable assessment of the privacy needs of users is desirable to tailor privacy tools [Knijnenburg, 2015], especially concerning the behavioral level. For example, tailored interventions like those used for changing pro-environmental behavior [Bamberg, 2013] could be perceivable [Döbelt and Günther, 2021].

From the criticized disadvantages, we could derive a lot of potential for improvement for our tool. From the participants’ perspective, the shortcoming was in the area of options for action. Our tool focused primarily on the enhancement of transparency about app behavior, but the participants’ feedback revealed, that this is too limited, and they would like to have further guidance on how to take action. As the protection of privacy will demand additional effort several authors demand automated recommendations [Smullen et al., 2020], [Gao et al., 2019]. Furthermore, the delay in waiting for the analysis reports was too long, because it was realized only in a semi-automated way. Automated testing, which delivers information immediately, will meet the information needs of users much better.

Our risk assessment included selected variables identified through static and dynamic analysis. However, there is potential for expansion, e.g., with the variable recipient. The app providers’ server location and integrated third parties to which app data is forwarded [Van Kleek et al., 2018] are also important for a holistic risk assessment. The integration of such extended information would further require a careful user-centered design of the interface.

11. Conclusion

Our field trial investigated behavioral effects and UUX of our tool – the AndProtect app – in everyday life. The tool combines static and dynamic app analysis results to increase transparency about app behavior and nudge users towards privacy-protective behavior. The results

revealed that it provoked a short time drop in installed apps, composed by increased uninstallations and constant installations – a "clean up". However, our tool did not promote knowledge or usage of the permission function. We attribute this to the fact that no respective option for action was offered. Especially, the named disadvantages indicated that more options for action are requested. This leads to the conclusion that merely creating transparency is not sufficient to nudge users towards privacy-protective behavior. Tools like ours should therefore accompany the whole process: provide information and offer options for action. Furthermore, a tailored approach [Knijnenburg, 2015] could further foster effectiveness.

Nevertheless, the self-assessment of our participants revealed that the tool also provokes increased attention to and reasoning of app behavior, such as a check of requested permissions and cost-benefit considerations, which is a sound foundation for responsible privacy decisions [Van Kleek et al., 2017]. The participants indicated that they used our tool because of internal motivation and interest in the content, which they considered to be well presented and had behavioral implications. Therefore, it can be stated that our tool could contribute to strengthening privacy-related awareness when dealing with mobile applications in everyday life.

Acknowledgments

This study was conducted as a part of the research project "*AndProtect: Personal data privacy by means of static and dynamic analysis for Android app validation*" (funding code: 16KIS0349), funded by the German Federal Ministry of Education and Research. We thank our project partners DAI-Labor and secuvera GmbH for their technical support during the field trial. Specifically, we want to thank Benjamin Aissa, Sebastian Fritsch and Ferenc Rózsa, who contributed the technical details in this paper. Furthermore, we want to thank Anna Graue for her contributions and proofreading of the paper.

References

Acquisti, A., Adjerid, I., Balebako, R., Brandimarte, L., Cranor, L. F., Komanduri, S., Leon, P. G., Sadeh, N., Schaub, F., Sleeper, M., Wang, Y., and Wilson, S. (2016). Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. Available at SSRN: https://ssrn.com/ abstract=2859227

Almuhimedi, H., Schaub, F., Sadeh, N., Adjerid, I., Acquisti, A., Gluck, J., Cranor, L. F., and Agarwal, Y. (2015). Your location has been shared 5,398 times: A field study on mobile app privacy nudging. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 787-796. New York: ACM. http://dx.doi.org/10.1145/2702123.2702210

AppBrain (2021). Anzahl der verfügbaren Apps im Google Play Store von April 2018 bis Februar 2021 (in 1.000). In Statista - Das Statistik-Portal. https://de.statista.com/statistik/daten/studie/74368/umfrage/anzahl-der-verfuegbaren-apps-im-google-play-store/ (Accessed 28th of October 2022)

Bal, G., Rannenberg, K., and Hong, J. (2014). Styx: Design and Evaluation of a New Privacy Risk Communication Method for Smartphones. In: Cuppens-Boulahia, N., Cuppens, F., Jajodia, F, El Kalam, A. A., Thierry Sans (Eds.) ICT Systems Security and Privacy Protection. SEC 2014. IFIP Advances in Information and Communication Technology, 428. Berlin, Heidelberg: Springer. http://dx.doi.org/10.1007/978-3-642-55415-5_10

Balebako, R., Leon, P. G., Almuhimedi, H., Kelley, P. G., Mugan, J., Acquisti, A., Cranor, L. F., and Sadeh, N. (2011). Nudging users towards privacy on mobile devices. In Proc. CHI 2011 Workshop on Persuasion, Nudge, Influence and Coercion.

Bamberg, S. (2013). Changing environmentally harmful behaviors: A stage model of self-regulated behavioral change. Journal of Environmental Psychology, 34, 151-159.

Bitkom. (2017.) Anteil der Smartphone-Nutzer in Deutschland nach Altersgruppe im Jahr 2017. In *Statista*. https://de.statista.com/statistik/daten/studie/459963/umfrage/anteil-der-smartphone-nutzer-in-deutschland-nach-altersgruppe/ (Accessed 28th of February 2019)

Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(194), 4-7.

Cavoukian, A. (2009). Privacy by design: The 7 foundational principles. Information and Privacy Commissioner of Ontario, Canada, 5.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hoboken: Taylor and Francis.

comScore. (2017). Mobile app share of total digital time spent on selected content categories in the United States in June 2017. In *Statista*. Retrieved October 28th, 2022 from https://www.statista.com/statistics/466874/us-mobile-content-categories-browser/

Cunha, J. A., and Aguiar, Y. P. C. (2020). Reflections on the role of nudges in human-computer interaction for behavior change: software designers as choice architects. In Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems (pp. 1-6). https://doi.org/10.1145/3424953.3426652

DeMaio T. J. (1984). Social desirability and survey. Surveying subjective phenomena, 2, 257.

Döbelt, S., and Halama, J. (2018). Mobiler Datenschutz: Nutzerzentrierte Gestaltung der AndProtect-App. Mensch und Computer 2018-Workshopband. https://doi.org/10.18420/muc2018-ws09-0547

Döbelt, S., Halama, S., Fritsch, S., Nguyen, M., and Bocklisch, F. (2020). Clearing the Hurdles: How to Design Privacy Nudges for Mobile Application Users. In A. Moallem (Ed.). Second International Conference on HCI for Cybersecurity, Privacy and Trust, HCI-CPT

2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19-24, 2020, Proceedings, 326-353, Springer, Cham. https://doi.org/10.1007/978-3-030-50309

Döbelt, S., and Günther, M. (2021). Two Values Work Alike: Linking Proenvironmental and Privacy Preserving Behavior. In Huckauf, A., Baumann, M., Ernst, M., Herbert, C., Kiefer, M., and Sauter, M. (Eds.), TeaP 2021 - Abstracts of the 63th Conference of Experimental Psychologists. Lengerich: Pabst Science Publishers.

Enck, W., Gilbert, P., Han, S., Tendulkar, V., Chun, B., Cox, L. P., Jung, S., McDaniel, P., and Sheth, A. N. (2014). TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones. ACM Transactions on Computer Systems (TOCS), 32(2). http://dx.doi.org/10.1145/2619091

Feizollah, A., Anuar, N. B., Salleh, R., and Wahab, A. W. A. (2015). A review on feature selection in mobile malware detection. Digital investigation, 13, 22-37. https://doi.org/10.1016/j.diin.2015.02.001

Felt, A. P., Ha, E., Egelman, S., Haney, A., Chin, A., and Wagner, D. (2012). Android permissions: User attention, comprehension, and behavior. In L. F. Cranor (Ed.), Proceedings of the Eighth Symposium on Usable Privacy and Security, 3. http://dx.doi.org/10.1145/2335356.2335360

Fraunhofer AISEC (2014). 10.000 Apps und eine Menge Sorgen. Retrieved from https://www.aisec.fraunhofer.de/de/presse-und-veranstaltungen/presse/pressemitteilungen/2014/20140403_10000_apps.html (Accessed 28th of October 2022)

Friedewald, M. (Ed.) (2018). Privatheit und selbstbestimmtes Leben in der digitalen Welt: Interdisziplinäre Perspektiven auf aktuelle Herausforderungen des Datenschutzes, 303.

García, F. M. (2013). When and Why is Attrition a Problem in Randomized Controlled Experiments and How to Diagnose it. http://dx.doi.org/10.2139/ssrn.2267120

Gao, H., Guo, C., Wu, Y., Dong, Y., Hou, X., Xu, S., and Xu, J. (2019). AutoPer: automatic recommender for runtime-permission in android applications. In 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), Vol. 1, 107-116. IEEE.

Gerber, N., Gerber, P., Drews, H., Kirchner, E., Schlegel, N., Schmidt, T., and Scholz, L. (2018). FoxIT: enhancing mobile users' privacy behavior by increasing knowledge and awareness. In Proceedings of the 7th Workshop on Socio-Technical Aspects in Security and Trust, 53-63. ACM.

Gerber, P., Volkamer, M., and Renaud, K. (2017). The simpler, the better: Presenting the COPING Android permission-granting interface for better privacy-related decisions. Journal of Information Security and Applications, 34, 8-26. https://doi.org/10.1016/j.jisa.2016.10.003

Hansen, M., Berlich, P., Camenisch, J., Clauß, J., Pfitzmann, A., and Waidner, M. (2004). Privacy-enhancing identity management. Information Security Technical Report, 9, 35-44. https://doi.org/10.1016/j.istr.2008.06.003

IDC (2021). Marktanteile der Betriebssysteme am Absatz vom Smartphones weltweit in den Jahren 2010 bis 2020 und Prognose bis 2025. In Statista - Das Statistik-Portal. https://de.statista.com/statistik/daten/studie/182363/umfrage/prognostizierte-marktanteile-bei-smartphone-betriebssystemen/ (Accessed 28th of October 2022)

ISO 9241-11 (1998). Ergonomic Requirements For Office Work With Visual Display Terminal (VDT)S – Part 11: Guidance On Usability. https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-1:v1:en (Accessed 28th of October 2022)

ISO 9241-210 (2010). Ergonomics of Human-System Interaction – Part 210: Human-Centred Design for Interactive Systems. https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-1:v1:en (Accessed 28th of October 2022)

Joeckel, S., and Dogruel, L. (2020). Default effects in app selection: German adolescents' tendency to adhere to privacy or social relatedness features in smartphone apps. Mobile Media and Communication, 8(1), 22-41. https://doi.org/10.1177/2050157918819616

Karapanos, E., Zimmerman, J., Forlizzi, J., and Martens, J. (2009). User experience over time. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '09: CHI Conference on Human Factors in Computing Systems. ACM. https://doi.org/10.1145/1518701.1518814

Karrer, K., Glaser, C., Clemens, C., and Bruder, C. (2009). Technikaffinität erfassen - der Fragebogen TA-EG. Der Mensch im Mittelpunkt technischer Systeme, 8, 196-201.

Kelley, P. G., Cranor, L. F., and Sadeh, N. (2013). Privacy as part of the app decision-making process. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 3393-3402. ACM. http://dx.doi.org/10.1145/2470654.2466466

Kelley, P. G., Consolvo, S., Cranor, L. F., Jung, J., Sadeh, N., and Wetherall, D. (2012). A Conundrum of Permissions: Installing Applications on an Android Smartphone. In J. Blyth, S. Dietrich, and L. J. Camp (Eds.), Financial Cryptography and Data Security, 68-79. Berlin, Heidelberg: Springer. http://dx.doi.org/10.1007/978-3-642-34638-5_6

King, J. (2012). How Come I'm Allowing Strangers to Go Through My Phone? Smartphones and Privacy Expectations. http://dx.doi.org/10.2139/ssrn.2493412

Van Kleek, M., Liccardi, I., Binns, R., Zhao, J., Weitzner, D. J., and Shadbolt, N. (2017). Better the devil you know: Exposing the data sharing practices of smartphone apps. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 5208-5220. ACM. https://doi.org/10.1145/3025453.3025556

Van Kleek, M., Binns, R., Zhao, J., Slack, A., Lee, S., Ottewell, D., and Shadbolt, N. (2018). X-Ray Refine: Supporting the Exploration and Refinement of Information Exposure Resulting from Smartphone Apps. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 393. ACM. https://doi.org/10.1145/3173574.3173967

Knijnenburg, B. P. (2015). A user-tailored approach to privacy decision support, Ph.D. Thesis, University of California, Irvine, Irvine, CA.

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., and Sinnelä, A. (2011). UX Curve: A method for evaluating long-term user experience, Interacting with Computers, Volume 23, Issue 5, 473–483. https://doi.org/10.1016/j.intcom.2011.06.005

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, Vol. 33, No.1, 159-174.

Laugwitz, B., Held, Th., and Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In Symposium of the Austrian HCI and usability engineering group, 63-76. Springer, Berlin, Heidelberg.

Lewis, J. R., and Sauro, J. (2018). Item benchmarks for the system usability scale. Journal of Usability Studies, 13(3).

Liu, B., Andersen, M. S., Schaub, F., Almuhimedi, H., Zhang, S., Sadeh, N., Acquisti, A., and Agarwal, Y. (2016). Follow my recommendations: A personalized privacy assistant for mobile app permissions. in SOUPS'16 Proceedings of the Twelfth USENIX Conference on Usable Privacy and Security, 27-41.

Malhotra, N. K., Kim, S. S., and Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. Information systems research, 15(4), 336-355.

Mayring, P. (2014). Qualitative Content Analysis: Theoretical Foundation, Basic Procedures and Software Solution. Retrieved from: http://nbn-resolving.de/urn:nbn:de:0168-ssoar-395173 (Accessed 28th of October 2022)

Nissenbaum, H. (2009). Privacy in context: Technology, policy, and the integrity of social life. Stanford University Press.

Norvell, D. C., Dettori, J. R., and Chapman, J. R. (2016). Enhancing clinical study retention rates to avoid follow-up bias: how do we keep our study participants from "the land of the lost"? Global spine journal, 6(05), 519-521. https://doi.org/10.1055/s-0036-1584928

Von Rauchhaupt, J. (2017). GfK-Halbjahresanalyse: Mobile First ist gelebte Nutzungsrealität. Retrieved from https://www.adzine.de/2017/11/gfk-halbjahresanalyse-mobile-first-ist-gelebte-nutzungsrealitaet/ (Accessed 28th of October 2022)

United Nations, Department of Economic and Social Affairs, Population Division (2015). World Population Prospects: The 2015 Revision. New York: United Nations.

Schrepp, M., Thomaschewski, J., and Hinderks, A. (2017). Construction of a Benchmark for the User Experience Questionnaire (UEQ). In International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 4, Issue 4, 40. Universidad Internacional de La Rioja. https://doi.org/10.9781/ijimai.2017.445

Shen, B., Wei, L., Xiang, C., Wu, Y., Shen, M., Zhou¸Y., and Jin, X. (2021). Can Systems Explain Permissions Better: Understanding Users' Misperceptions under Smartphone Runtime Permission Model. In 30'h USENIX Security Symposium USENIX Security 21.

Shklovski, I., Mainwaring, S. D., Skúladóttir, H. H., and Borgthorsson, H. (2014). Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In Proceedings of the 32nd annual ACM conference on Human factors in computing systems, 2347-2356. ACM. https://doi.org/10.1145/2556288.2557421

Smullen, D., Feng, Y, Zhang, S. A., and Sadeh, N. (2020). The best of both worlds: Mitigating trade-offs between accuracy and user burden in capturing mobile app privacy preferences. Proceedings on Privacy Enhancing Technologies, 2020(1), 195-215.

Statistisches Bundesamt Deutschland 2015. Statistisches Jahrbuch Deutschland und Internationales. Statistisches Bundesamt, Wiesbaden.

Statistisches Bundesamt 2022. Institut für Arbeitsmarkt- und Berufsforschung. IAB-Arbeitszeitrechnung. http://doku.iab.de/arbeitsmarktdaten/AZ_Komponenten.xlsx (Accessed 28th of October 2022)

Thaler, R. H., and Sunstein, C. R. (2008). Nudge: Improving decisions about health, wealth, and happiness. Constitutional Political Economy, 19(4), 356-360.

Trepte, S., Masur, P. K., and Von Pape, T. (2018). Privatheit im Wandel. Eine repräsentative Umfrage und eine Inhaltsanalyse zur Wahrnehmung und Beurteilung von Privatheit. Retrieved October 28th, 2022 from https://www.forum-privatheit.de/wp-content/uploads/Trepte_Privatheit-aus-psycholog.-Perspektive_Kick-Off-Forpri_2014-10-17.pdf

Xu, H., Gupta, S., Rosson, M. B., and Carroll, J. M. (2012). Measuring Mobile Users´ Concerns for Information Privacy. Paper presented at the Thirty-third International Conference on Information Systems, Orlando, FL.

YouGov. (2017). Wie viele Apps haben Sie auf Ihrem Smartphone installiert? In Statista - Das Statistik-Portal. https://de.statista.com/statistik/daten/studie/857931/umfrage/anzahl-installierter-smartphone-apps-in-deutschland/. (Accessed 28th of October 2022)

Appendices

A.1 Overview of Measurement Times and Variables

Table A1. Overview of measurement times and variables.

Phase	No.	Duration in min rnd.	Conduction	Dependent Variables
Baseline	T0	*M* = 28.00 *SD* = 10.00	Online	List of installed applications, app usage behavior and competence, usage and knowledge of permission function
	T1	*M* = 23.00 *SD* = 18.00	Online	List of installed apps, reasons for app uninstallation or installation of a specific app (individually adapted), permission granting or withdrawal for a specific app during last week (individually adapted)
	T2.1	*M* = 3.00 *SD* = 2.00	Face-to-Face (laboratory)	List of installed apps, reasons for app uninstallation or installation of a specific app (individually adapted), permission granting or withdrawal for a specific app during last week (individually adapted)
Trial	T2.2	*M* = 16.00 *SD* = 90.0	Face-to-Face (laboratory)	Usability and user experience evaluation of the tool perceived advantages and disadvantages of the tool (open-ended)
	T3	*M* = 16.00 *SD* = 7.00	Online	List of installed apps, reasons for app uninstallation or installation of a specific app (individually adapted), permission granting or withdrawal for a specific app during last week (individually adapted), usage context of the tool
	T4	*M* = 29.00 *SD* = 19.00	Online	List of installed apps, reasons for app uninstallation or installation of a specific app (individually adapted), permission granting or withdrawal for a specific app during last week (individually adapted), usability and user experience evaluation of the tool, usage context of the tool, perceived advantages and disadvantages of the tool (open-ended)
	T5	*M* = 17.00 *SD* = 13.00	Online	List of installed apps, reasons for app uninstallation or installation of a specific app (individually adapted), permission granting or withdrawal for a specific app during last week (individually adapted), usage context of the tool
	T6	*M* = 36.00 *SD* = 17.00	Online	List of installed apps, smartphone app knowledge, and usage behavior, usage/knowledge of permission function (general), reasons for app uninstallation or installation of a specific app (individually adapted), permission granting or withdrawal for a specific app during last week (individually adapted), usage context of the tool, usability and user experience of the tool, perceived advantages and disadvantages of the tool (open-ended), perceived changes of own behavior

A.2 User Interface AndProtect App

Figure A1. AndProtect app user interface; a) Start page; b) Upload screen for an unevaluated app; c) Analysis report of a certain app; d) Data access information; e) Individualization function of risk valuation, terms are translated into English and app-icons and names are blinded.

DOI: 10.2478/9788366675896-003 • Technikpsychologie • 31-40

sciendo

Proceedings of the International Symposium on Technikpsychologie

Users' Privacy Literacy, Motivation to use Tor and Further Privacy Protecting Behavior

Florian Platzer[1,†], Alexandra Lux[2,3]

[1]*Fraunhofer SITATHENE Rheinstraße 75 64295 Darmstadt, Germany*
[2]*TU Darmstadt, Germany University of Hohenheim, Germany*
[3]*University of Hohenheim Schloss Hohenheim 1 70599 Stuttgart, Germany*

Abstract

The concepts of** Tor **and** darknets **are usually associated with general anonymity. But this is not always the case. A more nuanced approach is needed that takes into consideration the resources of the potentially deanonymizing party. A connection to the Tor network can be established by using the Tor browser. However, different actors are still able to detect that one is connecting to the Tor network and can thus be traced. Furthermore the execution by the respective users also plays a key role in preserving their privacy. We carried out a study with the goal to gain primary insight into the Tor user group with regard to online privacy literacy and literacy of Tor in particular. Furthermore we examined the relation of motivations to use Tor and further privacy protective behavior through the employment of additional technologies in combination to Tor. Our results show that users who use Tor to protect their data from systems and organizations, have a higher Tor literacy score than Tor users with other motivations for using Tor. Regardless of the motivations for using Tor, encryption such as PGP and the use of Tails/Whonix are the most common additional technologies in combination with Tor.

Keywords

Privacy • Tor • User Behaviour

1. Introduction

Due to an exponential growth of data and the emergence of new technologies within short time frames, it is becoming increasingly difficult for individuals to keep track of what is happening with their data. Consequently, many people are concerned about their privacy [Auxier et al., 2019]. However, this does not necessarily mean that they adapt their behavior accordingly. Previous studies indicate a seeming discrepancy between the concern for privacy and the handling of one's own data [Acquisti and Gross, 2006, Paine et al., 2007, Wills and Zeljkovic, 2011], better known as the *Privacy Paradox* [Barnes, 2006]. While literature offers various plausible explanations for this, a commonly referred to is the *Knowledge Gap Hypothesis*. It states that users are concerned about their privacy but do not have the appropriate skills to adjust their behavior accordingly to better protect their privacy [Trepte et al., 2015, Debatin et al., 2009]. *Online Privacy Literacy* is thus a means to protect one's privacy [Trepte et al., 2015, Park, 2013, Baruh et al., 2017]. It encompasses knowledge of technical possibilities, related regulations and institutional practices for achieving online privacy as well as knowledge of their correct application [Trepte et al., 2015].

Previous research pointed to inconclusive results concerning the relationship between privacy concern and literacy and its effects on privacy behavior [Brough and Martin, 2020]. While some studies indicate that privacy literacy as opposed to concern is associated with privacy protection behavior [Park et al., 2012], others however were able to show a positive correlation between the level of privacy concern and privacy behavior [Weinberger et al., 2017]. A recent meta analysis indicated "privacy concerns predict the extent to which individuals use online services and engage in privacy management" [Baruh et al., 2017, p.20]. Thus, the relation between users' online privacy literacy and their motivation to use Tor is crucial in order to have a better understanding of users' privacy behavior [Brough and Martin, 2020]. One specific technical option to better protect one's privacy is the usage of privacy- enhancing technologies (PETs), like Tor. The anonymity network Tor offers Internet users both anonymity and privacy protection in the online context.

[†]*Corresponding author: Florian Platzer*
E-mail: florian.platzer@sit.fraunhofer.de

Following previous work, the goal of this study is to gain a better understanding of the Tor user group. With the correct application of privacy enhancing technology of key importance, we are particularly interested to not only examine user's online privacy literacy, but also their literacy regarding Tor. Just because people are using Tor does not automatically indicate best practice and thus maximum privacy. This is important as misuse due to a lack of literacy can result in less privacy than not using Tor at all [Lux and Platzer, 2022]. To the best of our knowledge no previous study has yet examined user's literacy of Tor. Further, we want to add to the existing body of literature and examine user's motivation to use Tor, as well as further privacy protecting behavior. Here we are especially interested in the relationship to other technologies that are used in addition to Tor to further enhance one's privacy and protect against deanonymization from different parties.

This paper is structured as follows: Section 2 discusses related work around online privacy literacy, Tor and provides the necessary background about technologies that can be applied in addition to Tor. Section 3 describes the methodology used to collect the data through an online questionnaire. Data is analyzed and evaluated in Section 4. Section 5 concludes this paper and discusses limitations, as well as future work.

2. Related Work and Background

In the following, we will elaborate on previous work on online privacy literacy in 2.1. In particular, we will present related work concerning online privacy literacy of user' of Tor. Subsequently, in 2.2, we will explicate the additional technologies that can be used in combination with Tor to further enhance users' privacy.

2.1 Related Work

The concept of privacy has been conceptualized in various ways. For example, conceptualization from various scholarly backgrounds includes such as "negative freedom (discourse in political philosophy), treated as a residual category of the public sphere (discourse in sociology), defined as a form of seclusion or condition of limited access (legal discourse), and finally described as selective or secondary control (psychological discourse and particularly the discourse on informational privacy)" [Masur, 2018, p.311-312]. The three key concepts that are commonly referred to in the context of privacy are constituted by the works of Westin [1967], Altman [1975] and Burgoon [1982]. Trepte and Dienlin [2014] include them in the following definition: "Privacy is an individual state of seclusion and intimacy [Westin, 1967], subject to a constant regulation of too much and too little privacy [Altman, 1975], where four different privacy dimensions can be distinguished at any point in time: informational, social, psychological, and physical privacy [Burgoon, 1982]" [Trepte and Dienlin, 2014, p.56].

Online privacy literacy encompasses the combination of declarative (knowing what) and procedural (knowing how) knowledge about privacy on the Internet. It can be measured using the Online Privacy Literacy Scale (OPLIS) [Masur et al., 2017a], which contains 20 questions that comprise the following four aspects (1) knowledge about institutional practices, (2) knowledge about technical aspects of data protection, (3) knowledge about data protection law, and (4) knowledge about data protection strategies. Concerning the online privacy literacy of Tor users, recent studies indicate that users of Tor have a higher online privacy literacy than regular internet users [Harborth and Pape, 2020b, Lux and Platzer, 2022]. A study by Harborth and Pape [2020b] suggested that privacy literacy can positively influence online behavior. Based on the results of an online questionnaire (N=124), they showed that "online privacy literacy has a positive effect on the trusting beliefs in Tor" and that "trusting beliefs in Tor have a positive effect on the behavioral intention to use Tor" [Harborth and Pape, 2020b, p. 63].

Additionally, with regard to online privacy literacy of Tor users another study by Lux and Platzer [2022] showed that there is neither a correlation between the usage frequency of Tor nor the context of whether Tor is being used to surf the clear- and/or darknet and the online privacy literacy of the users. However, they showed that frequency of usage and network type correlate in terms of knowledge of additional technology that can be used to enhance privacy. Users interested in *anonymity* have a higher score in online privacy literacy than users that did not indicate interest in the topic of anonymity. Users that applied Tor for *marketplaces* appeared to be less literate than users that did not use Tor to surf *marketplaces*. [Lux and Platzer, 2022]. However, neither of the above mentioned works did not consider users literacy of Tor.

Other studies focused on the usage of anonymization technologies on the Internet [Li et al., 2013, Danezis and Diaz, 2008, Winkler and Zeadally, 2015]. Here, networks such as Tor, I2P or JAP (JonDo) are understood as a means to protect users privacy or anonymity. However, this protection can be further optimized by applying additional technologies. Thus, we are investigating the use of technologies that can be added to Tor in order to protect one's privacy or anonymity from possible deanonymization. Previous works showed that different actors can deanonymize a Tor user: e.g. network administrator, Tor node provider, internet or web service provider, internet exchange point or government [Juen et al., 2014, Johnson et al., 2013, Ries et al., 2011, Lux and Platzer, 2022, Mani et al., 2018]. This depends on the resources that the respective actor has available. Depending on the actor

Tor users want to be protected from, they can use additional technologies to prevent deanonymization or to better protect their privacy.
In the following section, we will explicate the additional technologies that we considered in our study.

2.2 Background

Tor is an anonymity network that aims to protect a user's anonymity and hence affords them privacy. In the context of this work this particularly describes the condition where no user data can be collected (e.g. for profiling or tracking) unless the user agrees to it. Thus, user activities cannot be traced. Anonymity means that no one can determine the identity of an Internet user - for example, by collecting identifying information such as the user's IP address.
Internet users can connect to the Tor network by using the Tor browser. All data sent via the Tor browser is anonymized through the Tor network by obfuscating the IP address. The Tor network has over 6,000 publicly known Tor nodes [Project, 2022]. Due to the onion routing used in Tor, all data traffic is multiple times encrypted and routed through a data path of normally three Tor nodes. The first node in a data path is called *guard node*, and the last one is called *exit node*. Each Tor node in the data path knows only its predecessor and successor. No one knows the entire path. Thus, nobody can trace who is communicating with whom and about what. This means that when using Tor, users can surf the Internet anonymously. But there is no guarantee of not being deanonymized. Technologies that are added to the use of Tor can potentially increase resistance to deanonymization or better protect privacy. In this work we are in particular taking into account the technologies listed below.

Guard node. The guard node is the first Tor node in a data path to which a Tor user connects directly. Because of the direct connection, the guard node knows the IP address of the Tor user [Hopper et al., 2010]. As a result, the guard node may be a potential threat for attacks against the Tor user, which may lead to deanonymization under certain con- ditions [Cambiaso et al., 2019] (e.g. end-to-end correlation attacks [Basyoni et al., 2020, Abbott et al., 2007]). In order to protect themselves, users can operate their own Tor node and select it as the guard node for all Tor connections.

Bridges and PTs. Bridge nodes are Tor nodes that are not publicly listed and can thus serve as additional entry points into the Tor network if, for example, a government blocks all publicly known Tor nodes [Dunna et al., 2018]. Tor traffic has a unique structure [Saputra et al., 2016], so it is possible to detect Tor users that have connected to the Tor network. Pluggable Transports (PTs) can be used to obfuscate traffic [Khattak et al., 2014] so that this traffic is no longer detectable as containing Tor packets. Thus, Bridges and PTs provide protection from certain actors such as the internet service provider (ISP) or the government. Bridges and PTs prevent these actors from detecting that a Tor user is using Tor.

Tails/Whonix. With an operating system such as Tails or Whonix, all Internet connections are routed through the Tor network by default. If Tails or Whonix is not applied and the Tor browser is used on a regular operating system, only the data that the Tor user sends directly via the Tor browser is sent anonymously via the Tor network. All other data is still sent over clearnet connections and is therefore not anonymized.

VPN. In a virtual private network, a virtual tunnel is set up to a VPN provider through which all Internet data packets are encrypted. The VPN provider then sends all data packets to the actual destination. Connecting a VPN before the Tor network can provide additional protection against the ISP or against the first Tor node (*guard node*) within the data path while connecting a VPN after the Tor network can provide additional protection against the last Tor node (*exit node*).

PGP. PGP stands for Pretty Good Privacy and is used to encrypt files or messages. In this work, PGP is used as an example of additional encryption when communicating over the Tor network. PGP is not for anonymity because it just protects the content of messages. However, messages can have information that can deanonymize users. Due to the additional encryption, no data or messages are stored unencrypted on web servers. The recommendation to use PGP in addition to Tor is widespread among darknet market users [Dwyer et al., 2022].
It is important to note that not every technology necessarily improves a Tor user's protection. Every Tor user must be aware of which actor they want/need to protect themselves from. For example, a VPN provider can pose more risks than any protection. Choosing a VPN needs to be carefully considered. The risk of VPN providers selling or otherwise abusing their users' activity and data is well documented [Khan et al., 2018]. Using a VPN after the Tor network can also make more complete target profiles more visible to a single adversary, either the VPN or someone watching it.
Using Tor through the Tor browser or Tails/Whonix does not protect against detection that an Internet user is using Tor. Some actors, such as the ISP or the government, may still detect this. If the use of Tor is not allowed, this can lead to trouble for the user.

3. Methodology

Our user data was collected through an online questionnaire on various forums and message boards on the clear- and darknet.

In order to maximize our sample size, we chose channels that indicated user activity and in which people talk about the darknet. In addition to choosing platforms known to us, we relied on channels known from previous studies on the matter [Harborth and Pape, 2020a]. The survey distribution channels are listed in Table 1.

In order to comply with many channels rules, we asked permission from the moderators to distribute our survey and provided them with a short description of the study. Upon permission, we published the link to the survey along with a short description of the study and the invitation to participate.

The questionnaire consisted of a total of 31 questions. 20 of those asked about online privacy literacy and Tor literacy. One question referred to the knowledge and the use about additional technologies. Five questions asked about the usage of Tor and the according motivation for it. Four other questions asked for socio-demographic information. A filter question was placed at the beginning, asking whether the participant had ever used Tor. If this question was answered negatively, the survey ended immediately. Answering a question was mandatory to progress to the next question. Solely the questions on socio-demographics were kept optional.

Users' online privacy literacy was measured through the Online Privacy Literacy Scale (OPLIS) [Masur et al., 2017a]. Since our survey was aimed at Tor users worldwide, the part "knowledge of data protection law" was omitted due to the fact that this was based on German/European law. Thus, the survey of online privacy literacy comprised 15 questions instead of 20. Following recent studies [Harborth and Pape, 2020b], the score was extrapolated from 15 to 20 questions. This way results can be compared to score reports of previous studies.

Analogous to the OPLIS questionnaire [Trepte et al., 2015, Masur et al., 2017b], we operationalized declarative and procedural knowledge about Tor through five questions. These questions were created based on the best practices and FAQs posted on the Tor support website [Project, 2021]. In particular, they consisted of questions concerning the following aspects: security issues (TOR_01: checking manipulated version), the functionality of the Tor browser (TOR_02 & TOR_3: JavaScript, New Identity) and a general understanding of the Tor technology (TOR_04 & TOR_05: anonymous surfing). Answer options were provided for all questions, no free text was used.

Motivations for using Tor were derived from recent literature on the matter [Gehl, 2018, Gehl, 2016, Jardine, 2018]. In particular, we extracted the following four groups of motivations for using Tor: (1) data protection, (2) political reasons, (3) illegal material, (4) in- formative exchange. Further, based on [Lux and Platzer, 2022] we considered the following options for additional technologies that can be used in combination with Tor to better protect anonymity and privacy:

1. Using a VPN before connecting to Tor
2. Using a VPN after Tor
3. Using operating systems such as Tails or Whonix
4. Using Bridges with Pluggable Transports
5. Using one's own Tor relay as guard node
6. Using additional encryption such as PGP

In order to account for differences of knowledge and usage of additional technologies, participants were asked to indicate whether the technologies listed are (a) known and used, (b) known but not used, or (c) unknown.

4. Results

The questionnaire was accessed a total of 238 times, of which N=120 participants filled in completely. 206 participants stated that they had used Tor before. 11 participants stated that they had never used the Tor browser and thus ended the survey. 21 users ended the survey without answering the filter question. Subsequently, we first explicate the descriptives of our data in Section 4.1, before sharing the results of the statistical analysis in Section 4.2 and our evaluation in Section 4.3.

Table 1. Survey distribution channels.

Board/Forum	Type	D/C	Board/Forum	Type	D/C
The Hub	Forum	D	Galaxy3	SN	D
DiDw	Forum	D	Stronghold Past	Pastebin	D
Hidden Answer	Forum	D	reddit	Forum	C
DNM Avengers	Forum	D	Facebook	SN	C
crimenetwork	Forum	D	Twitter	SN	C
dread	Forum	D	LinkedIn	SN	C
envoy	Forum	D	OpenLab Augsburg	Chat	C
Connect	SN	D			

SN: Social Network, D: Darknet, C: Clearnet

4.1 Descriptives

Of the 120 people who completed the questionnaire, the age group between 20 and 39 years represented the majority of participants (69.17%) 14.17 % belonged to the age group between 14 and 19 years, followed by 12.5% in the age group between 40 and 59 years. The rest reported being under 14, over 60, or did not indicate their age. With 84.17% of the participants providing information about their gender, 74.17% were male, 6.7% were female, and 3.33% identified themselves with the category diverse. Almost half of all participants have a universal qualification as highest level of education (53.33%) . 17.5% have a high school diploma (university entrance qualification without studies) and the remaining 29.17% either have some high school certificate without a diploma, have a secondary school certificate or have no school leaving certificate. All values are shown in Table 2.

4.1.1 Online Privacy Literacy and Tor Literacy

Our analysis shows that based on our sample, Tor users answer on average 82% of the questions on the Online Privacy Literacy Scale (OPLIS) correctly.

The results of users' Tor literacy are depicted in Figure 1a. 41% of all participants answered four of the five questions correctly. 51% of all who answered four questions correctly belong to the group with a university degree and 22% belong to the group with a high school diploma. 57% of all who answered five questions correctly belong to the group with a university degree followed by 14% who belong to the group with a high school diploma. Only 14% of all participants answered two or fewer questions correctly.

As shown in Figure 1b, the two questions about the functionality of the Tor browser were answered incorrectly most often (TOR_02, TOR_03). 90% of all participants were able to correctly answer the security issue (TOR_01).

Table 2. Distribution of participants by gender, age group and highest level of education

Gender		Total	Age Group						Education				
			<14	14-19	20-39	40-59	>60	N/A	nS	sHS	SecS	HS	U
Male	(#)	89	-	17	59	11	2	-	13	6	8	17	45
	(%)	74.17	0.00	14.17	49.17	9.17	1.67	0.00	10.83	5.50	6.67	14.17	37.5
Female	(#)	8	1	-	6	1	-	-	1	2	-	2	3
	(%)	6.67	0.83	0.00	5.00	0.83	0.00	0.00	0.83	1.67	0.00	1.67	2.50
Diverse	(#)	4	1	-	3	-	-	-	2	-	-	-	2
	(%)	3.33	0.83	0.00	2.50	0.00	0.00	0.00	1.67	0.00	0.00	0.00	1.67
N/A	(#)	19	-	-	15	3	-	1	1	2	-	2	14
	(%)	15.83	0.00	0.00	12.5	2.50	0.00	0.83	0.83	1.67	0.00	1.67	11.67
Total	(#)	120	2	17	83	15	2	1	17	10	8	21	64
	(%)	100.00	1.67	14.17	69.17	12.50	1.67	0.83	14.17	8.33	6.67	17.50	53.33

Legend: nS: no school certificate. sHS: some high school certificate. SecS: secondary school certificate.
HS: high school diploma. U: university degree.

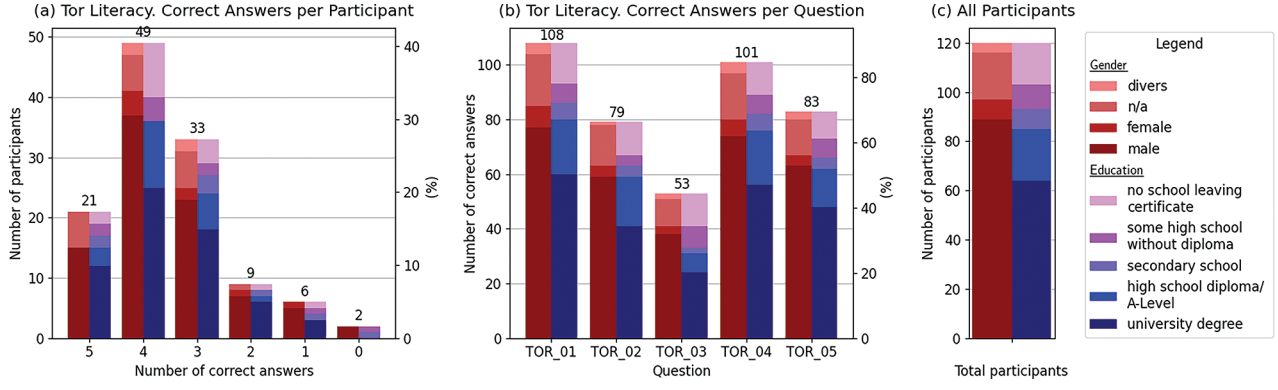


Figure 1. Tor literacy.

Table 3. Motivations for using the Tor network

Label	Motivations for using the Tor network	# Answers	(%)
MOT_01	I want to protect my data from systems and organizations that evaluate personal information and/or pass it on to third parties.	101	84.17
MOT_02	I use Tor for political reasons.	35	29.17
MOT_03	I would like to offer and/or consume material that is potentially illegal.	25	20.83
MOT_04	I would like to have an informative exchange on topics with a specific target group.	20	16.67

4.1.2 Motivations to Use Tor

Table 3 shows the motivations for using Tor. The majority of the N=120 participants (84%) indicated their motivation to be protected from systems and organizations that evaluate personal information and/or pass it on to third parties (MOT_01). 29% indicated their motivation to use Tor to be political reasons (MOT_02). 21% indicated their motivation to use Tor to be able to offer or consume material that is potentially illegal (MOT_03), and 17% indicated that their motivation as to share information about specific topics with specific groups (MOT_04). When asked about the motivation to use Tor, multiple choices were possible.

4.1.3 Additional Technologies

As depicted in Figure 2, most participants use PGP and/or operating systems such as Tails or Whonix as additional technologies in combination with Tor. Further, based on our sample, the majority are familiar with the possibility of switching a VPN before or switching a VPN after the Tor network, but only a fraction of those actively use it.

4.2 Analysis

In the following section, we present the results of our analysis concerning the relations between online privacy and Tor literacy, the motivation to use Tor and the additional technologies used.

In a first step, we analyzed online privacy literacy and Tor literacy. Spearman's ρ correlation coefficient was used to assess the relationship between users' online privacy literacy and Tor literacy. There was a weak positive correlation between the two variables, $\rho = .20$, $p < .05$. Thus, users who score higher in OPLIS also demonstrate a higher literacy concerning Tor.

Further, we were interested in examining the relation of users' motivation to use Tor and their Tor literacy. Our analysis indicates a statistically significant correlation between the motive to protect one's data from third parties (MOT_01) and users' Tor literacy ($\chi^2(5) = 14.67$, $p = .012$, $V = 0.35$). 29% of all participants who have the motivation to protect their data from third parties (MOT_01) have a Tor literacy score of three (three correctly answered questions out of five), just like 21.1% who do not have this motivation. In contrast, a score of four have 43.6% of those who use Tor to protect their data and only 26.3% who do not have this motivation. The highest score of five have 18.8% of those who have the motivation to protect their data and only 10.5% who don't have this motivation. Thus, users who use Tor to protect their data from third parties also have a higher Tor literacy score. For all other motivations, no significant effect could be observed in relation to Tor literacy.

Additionally, we analyzed how the use of the various additional technologies are distributed over the different motivations.

Protection of personal data from organizations and third parties (MOT_01). As shown in Figure 3, over 57% of all participants who use Tor to protect themselves from organizations that collect personal data (MOT_01) use PGP as an additional technology in combination with Tor. Over 45% of them use operating systems such as Tails or Whonix, and almost 24% use Bridges and Pluggable Transports. 21% use VPN before connecting to Tor in order to protect themselves from data collection and analysis on the Internet. PGP and Tails/Whonix are also the most commonly used for all other motivations given. This is followed by the use of Bridges+PTs.

Political Reasons (MOT_02). Over 74% of all participants that use Tor for political reasons (MOT_02) use PGP. 46% of them use operating systems such as Tails or Whonix, and 31% use Bridges and Pluggable Transports in addition to Tor.

Offer and/or consume potentially illegal material (MOT_03). Of the participants who stated their motivation to be to offer and/or consume potentially illegal material (MOT_03), most use PGP, followed by Tails/Whonix and the use of a VPN before connecting to the Tor network.

Informative exchange with a specific target group (MOT_04). 75% of those who stated that their motivation is to have an informative exchange on topics with specific target groups (MOT_04) use operation systems such as Tails or Whonix and 70% use PGP. Bridges and Pluggable Transports are used by 60%.

Subsequently, we analyzed the usage of additional technologies in relation to the motivation to use Tor. All of the significant effects occur in relation to the motivation *to have an informative exchange on topics with a specific target group* (MOT_04). For all other motivations, no significant effect

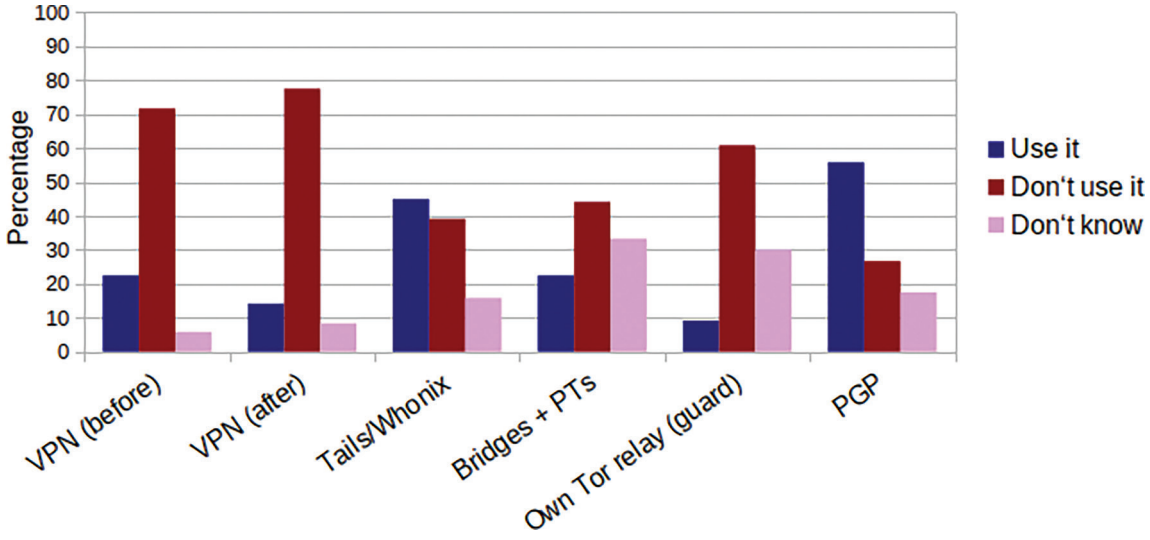


Figure 2. Knowledge and use of technologies.

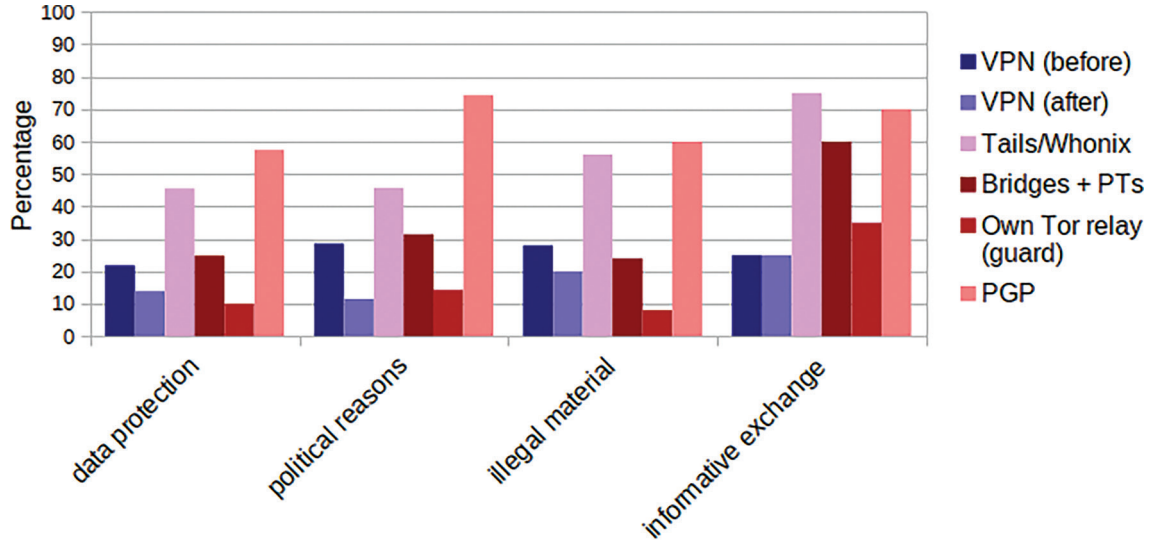


Figure 3. Motivation for using Tor and applied technologies.

could be observed. In the following, we present all effects by technologies: **Tails/Whonix** ($\chi^2(1) = 8.73$, $p = .003$, $V = 0.27$). The column percentages of the crosstabulation shows that 75% of all participants who indicated the motivation *informative exchange* use Tails/Whonix, as opposed to only 39% of participants who did not indicate this motivation.

Tails/Whonix is not used by 25% of the participants who indicated *informative exchange* as motivation and by 61% of the participants who do not use Tor for *informative exchange*. Thus, a significant effect could be observed on users who use Tor for information exchange and the usage of operation systems such as Tails or Whonix.

Bridges+PTs ($\chi^2(1) = 19.35$, $p < .0001$, $V = 0.40$). 60% of all participants who indicated the motivation *informative exchange* use Bridges+PTs as additional technology. Among the participants who did not indicate this motivation, only 15% use this technology.

Bridges+PTs are not used by 40% of participants who indicated *informative exchange* as motivation and by 85% of participants who did not indicate this motivation. Thus, a significant effect could be observed to users who use Tor for information exchange and the usage of Bridges and Pluggable Transports.

Own Guard Node ($\chi^2(1) = 19.24$, $p < .0001$, $V = 0.40$). 35% of all participants who indicated the motivation *informative*

exchange operate their own Tor node as a guard node. Among participants who did not indicate this motivation, only 4% use their own guard node. 65% of participants with this motivation and 96% of participants without this motivation do not operate a Tor node as their own guard node. Thus, a significant effect could be observed to users who use Tor for information exchange and the usage of their own guard node.

Evaluation

Results of recent studies indicate that Tor users have a higher online privacy literacy than regular Internet users [Harborth and Pape, 2020b, Lux and Platzer, 2022]. Our data indicates that users who score higher in OPLIS also demonstrate a higher literacy concerning Tor.

Concerning users' Tor literacy, our data suggests that users do not extensively engage with the functionality of the Tor browser. Some users do not know that JavaScript can be temporarily enabled for selected pages when JavaScript is globally disabled in the Tor browser. This could lead to Tor users not disabling JavaScript, as disabling JavaScript leads to limitations in the browsing experience. JavaScript makes it possible to identify web browsers that can be misused for tracking [Mulazzani et al., 2013]. JavaScript should be disabled by default and temporarily enabled only for trusted services or pages [Abbott et al., 2007]. However, by default, the Tor browser has JavaScript enabled.

Results further indicate that the majority of users have the knowledge how to check that they have not installed a manipulated version of Tor. This verification prevents Tor users from being deanonymized by a counterfeit version of Tor.

As already mentioned in Section 4.1.3, most Tor users from our survey use encryption such as PGP and operating systems such as Tails or Whonix in addition to Tor. A Tor user might accidentally establish an unpredictable connection via the clearnet when using the Tor browser instead of Tails or Whonix. By connecting via the clearnet, users of Tor would undermine their anonymity by revealing their IP addresses. Operating systems such as Tails or Whonix are easy ways to protect against a non-anonymous Internet connection. PGP prevents unauthorized people from reading the contents of a communication. Many darknet forums recommend using PGP in addition to Tor. Dwyer et al. showed that PGP is also regularly used in darknet marketplaces and is actively recommended [Dwyer et al., 2022]. Our work shows that PGP is not mainly used by users who would like to offer and/or consume material that is potentially illegal. In any motivation group, PGP is one of the most used technologies in addition to Tor. However, in order to protect oneself from actors such as the ISP, technologies such as VPN or Bridges must be used. From a purely technical point of view, VPNs could provide a remedy [Khan et al., 2018]. However, a VPN also carries the risk of a single point of failure. Another instance is added into the system, which must be trusted [Khan et al., 2018, Ramadhani, 2018]. This must be viewed critically. Within the Tor community there are different opinions on this issue. The Tor developers do not recommend using a VPN[11]. Our data shows that many users have knowledge about this technology, but do not use a VPN. By using Bridges and Pluggable Transports, actors can neither recognize who is communicating with whom, nor that they are communicating via the Tor network. Therefore, these technologies would be suitable for protection against Internet node operators, the Internet service provider and even against the government [Lux and Platzer, 2022].

Overall, users who use Tor in order to protect data from systems and organizations that evaluate personal information and/or pass it on to third parties, have a higher Tor literacy. User groups who exchange information with a specific target group use Tails/Whonix, Bridges+PTs or their own guard node more often than users who do not primarily use Tor to exchange information.

Since over 74% of all participants were male and only 6.6% were female, no statement could be made about gender and online privacy literacy or Tor literacy.

Conclusion

The goal of this study was to gain further insight into the Tor user group with regard to online privacy literacy, Tor literacy and the relations of motivations for using Tor and further privacy protecting behavior through the use of additional technologies in combination to Tor to further enhance user's privacy.

We showed that people who use Tor to protect their data from systems and organizations that evaluate personal information and/or pass it on to third parties have a higher Tor literacy score than Tor users with other motivations for using Tor. This group of users use encryption like PGP and operating systems such as Tails or Whonix as additional technologies in combination with Tor. Our data indicates that regardless of the motivations for using Tor and the users' literacy, PGP and operating systems such as Tails or Whonix are most commonly applied. While these already provide additional protection for privacy and anonymity, a more nuanced examination in regards to the possible deanonymizing actor is necessary. Especially Bridges and Pluggable Transports (PTs) serve as a countermeasure against deanonymization by potent actors such as ISPs or even governments. The results show that users who use Tor for informative exchange with a specific target group are more likely to use Bridges+PTs, Tails/Whonix or their own guard nodes, than users with other

[1] see e.g.: How can we help? | Tor Project | Support: https://support.torproject.org/#faq-5 (Accessed: 25.03.2022) and TorPlusVPN: https://gitlab.torproject.org/legacy/trac/-/wikis/doc/TorPlusVPN (Accessed: 25.03.2022)

motivations for using Tor. This might indicate that those users are generally more interested in *anonymity* and thus are more privacy literate.

Limitation. This study is limited in regards to the sample size (N=120). Though the sample size is similar to other studies on this matter [Harborth and Pape, 2020b], certain aspects are underrepresented and do thus not allow generalizability. This is mainly due to the fact that access to the field is not easily achieved. Disclosing information in this context can be considered against the norm and is thus met with distrust. Therefore, it is difficult to get participants to answer a questionnaire about their browsing behavior on the darknet when at the same time, based on anonymity, they use a darknet to keep just such information anonymous. Further, we cannot guarantee that participants will answer truthfully such as about offering or consuming illegal materials due to privacy concerns.

Future work. Future work should consider a more detailed investigation of users' backgrounds in terms of culture and legislation. This would help explain differences in the constellation of anonymizing technologies and actors for their own privacy protection. Besides online privacy literacy and Tor literacy, we could further investigate the user's literacy regarding the technologies used to improve privacy.

Acknowledgment

The joint project PANDA on which this publication is based was funded by the Federal Ministry of Education and Research under the funding codes 13N14355 and 13N14356. The authors are responsible for the content of this publication.

References

Abbott, T. G., Lai, K. J., Lieberman, M. R., and Price, E. C. (2007). Browser-based attacks on tor. In *International Workshop on Privacy Enhancing Technologies*, pages 184–199. Springer.

Acquisti, A. and Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the facebook. In *International workshop on privacy enhancing technologies*, pages 36–58. Springer.

Altman, I. (1975). The environment and social behavior: Privacy, personal space, territory, and crowding.

Auxier, B., Rainie, L., Anderson, M., Perrin, A., Kumar, M., and Turner, E. (2019). Americans and privacy: Concerned, confused and feeling lack of control over their personal information.

Barnes, S. B. (2006). A privacy paradox: Social networking in the united states. *First Monday*.

Baruh, L., Secinti, E., and Cemalcilar, Z. (2017). Online privacy concerns and privacy management: A meta-analytical review. *Journal of Communication*, 67(1):26–53.

Basyoni, L., Fetais, N., Erbad, A., Mohamed, A., and Guizani, M. (2020). Traffic analysis attacks on tor: a survey. In *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, pages 183–188. IEEE.

Brough, A. R. and Martin, K. D. (2020). Critical roles of knowledge and motivation in privacy research. *Current opinion in psychology*, 31:11–15.

Burgoon, J. K. (1982). Privacy and communication. *Annals of the International Communication Association*, 6(1):206–249.

Cambiaso, E., Vaccari, I., Patti, L., and Aiello, M. (2019). Darknet security: A categorization of attacks to the tor network. In *ITASEC*, pages 1–12.

Danezis, G. and Diaz, C. (2008). A survey of anonymous communication channels. Technical report, Technical Report MSR-TR-2008-35, Microsoft Research.

Debatin, B., Lovejoy, J. P., Horn, A.-K., and Hughes, B. N. (2009). Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of computer-mediated communication*, 15(1):83–108.

Dunna, A., O'Brien, C., and Gill, P. (2018). Analyzing china's blocking of unpublished tor bridges. In *8th USENIX Workshop on Free and Open Communications on the Internet (FOCI 18)*.

Dwyer, A. C., Hallett, J., Peersman, C., Edwards, M., Davidson, B. I., and Rashid, A. (2022). How darknet market users learned to worry more and love pgp: Analysis of security advice on darknet marketplaces. *arXiv preprint arXiv:2203.08557*.

Gehl, R. W. (2016). Power/freedom on the dark web: A digital ethnography of the dark web social network. *new media & society*, 18(7):1219–1235.

Gehl, R. W. (2018). *Weaving the dark web: legitimacy on freenet, Tor, and I2P*. MIT Press.

Harborth, D. and Pape, S. (2020a). Dataset on actual users of the privacy-enhancing technology tor.

Harborth, D. and Pape, S. (2020b). How privacy concerns, trust and risk beliefs, and privacy literacy influence users' intentions to use privacy-enhancing technologies: The case of tor. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, 51(1):51–69.

Hopper, N., Vasserman, E. Y., and Chan-Tin, E. (2010). How much anonymity does network latency leak? *ACM Transactions on Information and System Security (TISSEC)*, 13(2):1–28.

Jardine, E. (2018). Privacy, censorship, data breaches and internet freedom: The drivers of support and opposition to dark web technologies. *new media & society*, 20(8):2824–2843.

Johnson, A., Wacek, C., Jansen, R., Sherr, M., and Syverson, P. (2013). Users get routed: Traffic correlation on tor by realistic adversaries. In *Proceedings of the 2013 ACM SIGSAC*

conference on Computer & communications security, pages 337–348.

Juen, J., Johnson, A., Das, A., Borisov, N., and Caesar, M. (2014). Defending tor from network adversaries: A case study of network path prediction. *arXiv preprint arXiv:1410.1823*.

Khan, M. T., DeBlasio, J., Voelker, G. M., Snoeren, A. C., Kanich, C., and Vallina-Rodriguez, N. (2018). An empirical analysis of the commercial vpn ecosystem. In *Proceedings of the Internet Measurement Conference 2018*, pages 443–456.

Khattak, S., Simon, L., and Murdoch, S. J. (2014). Systemization of pluggable transports for censorship resistance. *arXiv preprint arXiv:1412.7448*.

Li, B., Erdin, E., Gunes, M. H., Bebis, G., and Shipley, T. (2013). An overview of anonymity technology usage. *Computer Communications*, 36(12):1269–1283.

Lux, A. and Platzer, F. (2022). Online-privatheitskompetenz und möglichkeiten der technischen umsetzung mit dem anonymisierungsnetzwerk tor. *Selbstbestimmung, Privatheit und Datenschutz*, page 129.

Mani, A., Wilson-Brown, T., Jansen, R., Johnson, A., and Sherr, M. (2018). Understanding tor usage with privacy-preserving measurement. In *Proceedings of the Internet Measurement Conference 2018*, pages 175–187.

Masur, P. K. (2018). *Situational privacy and self-disclosure: Communication processes in online environments*. Springer.

Masur, P. K., Teutsch, D., and Trepte, S. (2017a). Entwicklung und validierung der online-privatheitskompetenzskala (oplis). *Diagnostica*.

Masur, P. K., Teutsch, D., and Trepte, S. (2017b). Oplis. online privacy literacy scale. https://www.oplis.de/docs/OPLIS_Translation_english. pdf. Accessed: 2022-03-25.

Mulazzani, M., Reschl, P., Huber, M., Leithner, M., Schrittwieser, S., Weippl, E., and Wien, F. (2013). Fast and reliable browser identification with javascript engine fingerprinting. In *Web 2.0 Workshop on Security and Privacy (W2SP)*, volume 5, page 4. Citeseer.

Paine, C., Reips, U.-D., Stieger, S., Joinson, A., and Buchanan, T. (2007). Internet users' perceptions of 'privacy concerns' and 'privacy actions'. *International Journal of Human-Computer Studies*, 65(6):526–536.

Park, Y. J. (2013). Digital literacy and privacy behavior online. *Communication Research*, 40(2):215–236.

Park, Y. J., Campbell, S. W., and Kwak, N. (2012). Affect, cognition and reward: Predictors of privacy protection online. *Computers in Human Behavior*, 28(3):1019–1027.

Project, T. (2021). How can we help? support. https://support. torproject.org/. Accessed: 2021-07-13.

Project, T. (2022). Servers - tor metrics. https://metrics.torproject. org/networksize.html. Accessed: 2022-24-11.

Ramadhani, E. (2018). Anonymity communication vpn and tor: a comparative study. In *Journal of Physics: Conference Series*, volume 983, page 012060. IOP Publishing.

Ries, T., Panchenko, A., Engel, T., et al. (2011). Comparison of low-latency anonymous communication systems-practical usage and performance. In *Ninth Australasian Information Security Conference*, pages 77–86. ACS.

Saputra, F. A., Nadhori, I. U., and Barry, B. F. (2016). Detecting and blocking onion router traffic using deep packet inspection. In *2016 International Electronics Symposium (IES)*, pages 283–288. IEEE.

Trepte, S. and Dienlin, T. (2014). Privatsphäre im internet. *Neue Medien und deren Schatten. Mediennutzung, Medienwirkung und Medienkompetenz*, pages 53–79.

Trepte, S., Teutsch, D., Masur, P. K., Eicher, C., Fischer, M., Hennhöfer, A., and Lind, F. (2015). Do people know about privacy and data protection strategies? towards the "online privacy literacy scale"(oplis). In *Reforming European data protection law*, pages 333–365. Springer.

Weinberger, M., Bouhnik, D., and Zhitomirsky-Geffet, M. (2017). Factors affecting students' privacy paradox and privacy protection behavior. *Open Information Science*, 1(1):3–20.

Westin, A. (1967). Privacy and freedom new york atheneum, 1967. *Privacy and Personnel Records," The Civil Liberties Review (Jan./Feb., 1976)*, pages 28–34.

Wills, C. E. and Zeljkovic, M. (2011). A personalized approach to web privacy: awareness, attitudes and actions. *Information Management & Computer Security*.

Winkler, S. and Zeadally, S. (2015). An analysis of tools for online anonymity. *International Journal of Pervasive Computing and Communications*.

DOI: 10.2478/9788366675896-004 • Technikpsychologie • 41-54

Proceedings of the International Symposium on Technikpsychologie

Defeating Dark Patterns: The impact of supporting information on dark patterns and cookie privacy decisions

Jennifer Klütsch[1,†], Christian Böffel[1,†], Sophia von Salm-Hoogstraeten[1,2,†], Sabine J. Schlittmeier[1,†]

[1]RWTH Aachen Professorship for Work and Engineering Psychology Jägerstr. 17-19 52056 Aachen, Germany
[2]Saint Gobain Research Germany Glasstraße 152134 Herzogenrath, Germany

Abstract

In our daily usage of websites, we face a multitude of privacy decisions. However, as users often lack the knowledge, motivation or time for these decisions, they are vulnerable to privacy-unfriendly designs, so-called dark patterns. To promote deliberate cookie decisions even in presence of dark patterns, the present study investigates, within an experimental online-study (N = 207), the effectiveness of a knowledge intervention. During cookie decisions, either a knowledge intervention in the form of decision-supporting information ("cookie assistance") or no such information (instead an add) were presented. The intervention effects on deliberation, reaction time and cookie activation were investigated, in the presence and absence of dark patterns. Results indicate that supporting information neither stimulated more deliberate and slower reactions nor encouraged participants to make privacy-friendlier decisions in the presence of dark patterns. Instead, evidence for highly conditioned decision-making was found. Practical implications for future privacy-interventions and legal regulations are discussed.

Keywords

privacy • cookies • dark patterns • information

1. Introduction

Users face a multitude of privacy decisions in their daily life. They decide on privacy policies when creating user accounts, on access authorizations when using apps or on web cookies when visiting websites. However, even though users are expected to make deliberate privacy decisions, with the General Data Protection Regulation (short: GDPR) specifying that their consent must be "freely given, specific, informed and unambiguous" (GDPR, 2016, p. 6), researchers argue that most users still tend to make uninformed decisions on privacy policies or web cookies (e.g., Graßl et al., 2021; Gray et al., 2018; Machuletz & Böhme, 2020; Utz et al., 2019). This could be due to the fact, that users face difficulties in balancing perceived risks (e.g., release of personal data) and benefits (e.g., social affiliation, discounts) of privacy decisions equally, as assumed by the privacy calculus theory by Culnan and Armstrong (1999, see, Barth & de Jong, 2017; Kokolakis, 2017, for a review). One reason could be that the storage and processing of personal data are often not users' primary task (Acquisti et al., 2017) or that users lack motivation, knowledge, ability, or time for deliberated risk assessments (Bösch et al., 2016). Thus, users could rely on automatic, fast, and effortless processing rather than slow, effortful, and deliberate processing (see, Barth & de Jong, 2017, for a review). They might base their decisions on incomplete information, information asymmetries, heuristics, biases and other context-dependent factors (see, Acquisti et al., 2017; Kokolakis, 2017, for a review). For instance, during large-scaled analysis of users' behavior on Facebook, Gross and Acquisti (2005) found first evidence that privacy decisions are influenced by default settings. Even if participants had the option to change their default settings and limit their personal data release, most kept the default option and provided large amounts of personal data to Facebook (Gross & Acquisti, 2005). Moreover, Acquisti et al. (2013) demonstrated that users' privacy decisions are biased through loss aversion or order effects as participants appreciated the receipt of personal data (potential loss) more than the protection of public data (potential

[†]Corresponding author: Jennifer Klütsch; Christian Böffel; Sophia von Salm-Hoogstraeten; Sabine J. Schlittmeier
E-mail: kluetsch@psych.rwth-aachen.de; christian.boeffel@psych. rwth-aachen.de; sophia.von.salm@ outlook.de; sabine.schlittmeier@ psych.rwth-aachen.de

gain) and this behavior was influenced by the order of the data offer. To improve informed and deliberate privacy decisions by the user, the present study examines the effectiveness of supporting information during the cookie decision itself and the role of so-called *dark patterns*. Such deliberately chosen design choices represent major security threats and could tempt users to increased disclosure of personal data.

Dark patterns and web cookies

By hiding information to discourage certain options or by redirected expected functionalities (Gray et al., 2018), *dark patterns* manipulate users' accessible number of choices or their accessible information (Mathur et al., 2021) to guide them to the privacy-unfriendly option. In other words, they "(trick) users into performing unintended and unwanted actions, based on a misleading interface design" (Bösch et al., 2016, p. 239). *Dark patterns* are one type of nudge (Acquisti et al., 2017) in which a nudge describe "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" (Thaler & Sunstein, 2008, p. 6). Consequently, these *dark patterns* counteract legal efforts by the European General Data Protection Regulation (GDPR) as they support unintended rather than informed privacy decisions (e.g., Graßl et al., 2021; Machuletz & Böhme, 2020; Nouwens et al., 2020; Utz et al., 2019).

Unfortunately, *dark patterns* are common practice on various online services, e.g., within web cookie banners. For instance, during large-scaled analysis of 11,000 e-commerce websites, more than 11% implemented at least one dark pattern (Mathur et al., 2019). Further, even if websites should, according to the GDPR, include cookie banners where (1) consent needs to be explicit, (2) consent and withdrawal are equally difficult, as well as (3) no non-necessary purposes or vendors are preselected (Nouwens et al., 2020), Nouwens et al. (2020) found evidence that only about one tenth of examined websites adhered to all three requirements. As web cookies represent the most used tracking method and are able to track online-behavior for advertisement or analytics (Sanchez-Rola et al., 2019), they represent a major security gap in users' lives that is particularly devious when combined with dark patterns. For example, hiding the reject-button or showing bulk options increased users' cookie acceptance (Nouwens et al., 2020). Additionally, a large-scaled study by Utz et al. (2019) demonstrated that cookie banner positions and certain cookie types increased interaction rate or cookie activation. For instance, the researchers observed increased cookie activation when dark patterns such as pre-selected checkboxes were presented (Utz et al., 2019).

The multitude of privacy risks associated with web cookies and implemented dark patterns underscore the importance to investigate why dark patterns influence privacy behavior and to derive interventions that counteract their impact. The Dual Process Model by Kahneman (2003) is a framework that can be used to explain how dark patterns impact decision making. It distinguishes two systems of cognitive processes: (1) *System 1*, referring to automatic, fast, and effortless thinking and reasoning processes, and (2) *System 2*, referring to slower, effortful, and deliberated processes (Kahneman, 2003). As users often lack the motivation, ability, knowledge, or time for privacy decisions, most privacy decisions rely on automatic, fast, and effortless *System 1* (Bösch et al., 2016). For instance, when users visit a website, they are interested in using a service whereas the associated cookie decision is rather incidental. Users often make such a decision fast and without much effort, because their motivation and time to engage with the cookie banner are limited. Additionally, information provided by cookie banners is often needlessly complex or detailed, users' knowledge about the topic limited and their ability to make a deliberate privacy decision low. However, as users do not deliberate their decision in *System 1* processing, they are influenced by heuristics and biases and tend to focus on the highlighted and simplest option. They are also less aware to notice dark patterns and, thus, cannot counteract them (Bösch et al., 2016). First empirical results on users' awareness of dark patterns emphasize the assumption that users rely on *System 1* processing. Although, users are generally aware of deceptive design choices (Maier & Harr, 2020) and their impact on users' behavior (Bongard-Blanchy et al., 2021), they are unable to detect or unsure about dark patterns that prevail during their privacy decisions and underestimate the harm to themselves (Bongard-Blanchy et al., 2021; Di Geronimo et al., 2020). However, their detection of dark patterns is improved when they are primed or informed (Di Geronimo et al., 2020).

Improving privacy-friendly decisions by interventions

Integrating the theoretical and empirical insights, interventions would have to either guide *System 1* processing into privacy-friendly directions (e.g., Dennis & Minas, 2018; Sunstein & Reisch, 2019) or trigger *System 2* processing by increasing motivation as well as knowledge (e.g., Bösch et al., 2016; Sunstein & Reisch, 2019). Besides, interventions could encourage users' general competencies as suggested by Hertwig and Grüne-Yanoff (2017). Thus, Graßl et al. (2021) highlight the importance of three intervention approaches: (1) *non-educative nudges* to guide users' System 1 processing to the privacy-friendly option, (2) *educative nudges* to induce *System 2* processing (Sunstein & Reisch, 2019) or (3) *long-term boosts* to generally enhance users' competencies (Hertwig & Grüne-Yanoff, 2017).

The first intervention approach—*non-educative nudges*—refers to nudges that guide users' behavior towards one desired

outcome without intending to increase users' ability to choose with privacy-friendly settings (Sunstein & Reisch, 2019). One example for *non-educative nudges* are *bright patterns* that guide users to the privacy-friendly option and quickly and effectively reduce privacy risks (Graßl et al., 2021). For instance, Graßl et al. (2021) reversed default options or obstructions in cookie banners to bright patterns and demonstrated a corresponding increase on users' cookie rejections as an indicator for privacy-friendly behavior—. Nevertheless, like *dark patterns*, non-educative nudges exploit cognitive heuristics and biases (Hertwig & Grüne-Yanoff, 2017). Rather than enhancing users' reflective thinking, interventions such as *bright patterns* trigger unreflective *System 1* processing (Graßl et al., 2021; Hertwig & Grüne-Yanoff, 2017). Yet even *bright patterns* are successful interventions to encourage more privacy-friendly behavior, the negative effects of *dark patterns* are not reduced the next time they appear.

Therefore, the present study focuses on so-called *educative nudges* to counteract the impact of dark patterns. *Educative nudges* such as reminders or warnings are designed to trigger users' *System 2* processing (Graßl et al., 2021; Sunstein & Reisch, 2019). Examples for *educative nudges* could be interruptions that emphasize potential dangers (fear appeal), highlight unintentional data disclosures and its potential counter measurements (warnings), highlight the most important information (attractors) (Distler et al., 2020), or prepare comprehensible information (disclosure requirements) (Sunstein, 2016). First empirical results showed their effectiveness.; for example, Harbach et al. (2014) visualized Android app permissions with personalized examples (e.g., users' shared photos or locations) as well as highlighted accompanied dangers and found evidence that these modified permissions caused fear and guided users to the privacy-friendly decision. Nevertheless, it should be noted that these *educative nudges* (e.g., fear appeals) raise ethical questions as it is unclear whether a deliberated or a fear-led decision is encouraged (Distler et al., 2020). Rather than inducing fear, *educative nudges* can provide users with knowledge to make a deliberated cookie decision (Sunstein, 2016; Sunstein & Reisch, 2019), for instance, by providing feedback on consequences of certain cookie decisions (Graßl et al., 2021). However, it remains unclear how these types of educative nudges should be implemented as well as how and which kind of information should be presented to improve knowledge and motivation (cf. Graßl et al., 2021).

The present study

Overall, the multitude of dark patterns implemented by website operators as well as the associated privacy risks highlight the need for further research on how to enable deliberate privacy decisions (e.g., Di Geronimo et al., 2020; Mathur et al., 2019; Nouwens et al., 2020). One possible approach is to examine interventions, which promote *System 2* processing (*educative nudges*). With its exploratory approach, the present study aims to gain insights on one *educative nudge,* in form of a knowledge intervention, and its effectiveness during web cookie privacy decisions, in particular in the presence of dark patterns. Thus, it contributes to present research on educative nudges by incorporating insights from the Dual Process Model and provides first ideas on how educative nudges might be implemented as well as how and which kind of information could be presented. For this approach, participants were instructed to indicate their cookie privacy decision before assessing one cooking webpage to state their first impression (as cover story). The associated knowledge intervention consisted of supporting information on three cookie types (session, third-party, tracking) and optional cookies during the privacy decision, where we systematically varied two aspects in an online experiment: (1) whether participants received supporting information on cookie banners or not in (2) the presence or absence of dark patterns. We examined participants' deliberation on and time required for the cookie decisions as well as their cookie activation.

We expected, that supporting information about cookie banners promotes deliberate decisions while prolonging decision times on cookie options. Derived from the Dual Process Model by Kahneman (2003), dark patterns could influence users' privacy decisions due to automatic *System 1* processing (Bösch et al., 2016). According to this idea, *educative nudges* such as supporting information within a knowledge intervention could help users to counteract dark patterns if they enhance *System 2* processing. Consequently, the impact of dark patterns could be reduced leading to privacy-friendly decisions (Bösch et al., 2016; Sunstein & Reisch, 2019). We expected, that participants are less likely to activate cookies if dark patterns are presented with supporting information than without supporting information. Current research found evidence that users are more likely to activate cookies when dark instead of no patterns were presented (e.g., Machuletz & Böhme, 2020; Nouwens et al., 2020; Utz et al., 2019). However, as *System 2* processing should be accompanied with privacy decisions independent of dark patterns impact (Bösch et al., 2016), we expected the difference in cookie activation between dark and no patterns to be reduced when supporting information is presented.

2. Method

Participants

A total of 207 participants between 18 and 64 years (M = 26.3 years, SD = 7.7) completed the experiment (76 males, 129 females, 2 non-binary). Whether participants received trials with (cookie assistant) or without (app

advertisement) information was randomly assigned between participants resulting in 47.8 % of participants in the "no information" condition and 52.2 % in the "information" condition. With respect to their digital media usage, more than half of all participants (53.1 %) indicated to use digital media more than 40 hours per week. All participants gave informed consent and agreed to data collection and were recruited through e-mail, social media or university mailing lists. Participation was voluntary and without any payment. Psychology students of RWTH Aachen University received course credits for participation. The inclusion criterion was an age of at least 18 years. Data collection was carried out from 7 June to 19 July 2021.

An a-priori power analysis using G*Power Tool (Faul et al., 2007) was conducted to determine the required sample size on the between-subjects comparison between the information and no information groups. For a power of .80 and assuming a medium effect of d =.50 (Cohen, 1988) with an alpha level of α = .05, a total sample size of N = 128 or larger was determined.

Material and apparatus

The experiment was created with PsychoPy 2020.2.10 (Peirce et al., 2019) and hosted on Pavlovia (Bridges et al., 2020). For participation, a laptop or computer with a commonly used browser (Chrome, Edge, Safari, Firefox) was mandatory. Stimuli were centered and adapted to the participants' screen and its size (adaptable size for each stimuli). The stimulus material consisted of 14 mock-up cooking webpages and four—pattern (2) x information (2)—cookie banners per webpage with a screen resolution of 1384 x 1080 Pixel. All resulting 56 cookie banners included a counterbalanced amount of either dark and no patterns as within-subject factor as well as either supporting information via a cookie assistant or no information via an app advertisement as between-subjects factor (see Figure 1 for examples). This resulted in 28 cookie banners with supporting information (cookie assistant) or without information (app advertisement), in which each presented 14 cookie banners with dark and 14 cookie banners with no patterns. Dark patterns were implemented by highlighting only the accept-all button with color whereas no patterns included cookie banners where both (accept/reject) buttons were highlighted with color (i.e., interface interference or aesthetic manipulation; Fansher et al., 2018; Graßl et al., 2021). Additionally, within the information condition, supporting information on three cookie types (session, third-party, tracking) and optional cookies were presented by a cookie assistant whereas, within the no information condition, an app advertisement for the respective website was presented. The cookie information within the information condition was based on the German Consumer Association (Verbraucherzentrale, 2021). Both conditions (information/no information) consisted of an approximately equal number of words as pop-up to the left and above the cookie banner. Additionally, to encourage a deliberate rather than a fear-led decision, fear-inducing language was avoided when designing the information condition (see Figure 1 for examples of stimuli webpages and Appendix A for its English translation). Each cookie banner was presented as a layer above the subsequent website and participants were able to click on the buttons "Disagree to all" or "Agree to all". The subsequent cooking webpages varied for color (green/blue/red/orange/brown/ grey), design ("cooking planet"-"Kochplanet" /"cooking friends"-"Kochfreunde") and recipe (salad/pasta) and were presented once to each participant (independently of their cookie decision). Within the experimental condition, participants' cookie decision (accept/decline) and reaction time were measured. Further, participants' first impression for the cooking webpage on a Likert scale from 1 (terrible) to 7 (very good) and, optionally, one word that describes their first impression were queried (see Appendix B). The first impression task served as a cover story and was not recorded.

Further, demographic data (age, gender), digital-media usage and a post-hoc questionnaire were queried. The post-hoc questionnaire contained questions regarding *deliberation* of the cookie decision and *privacy attitude*. Furthermore, *treatment checks* for the knowledge intervention were included. Both, the experimental phase and the questionnaire, were tested for comprehensibility with three one-on-one tests using think aloud techniques.

Deliberation was measured through six items on a Likert scale from 1 (I do not agree at all) to 7 (I totally agree). As the Dual Process Model distinguishes *System 1* and *System 2* processing (Kahneman, 2003), items related to either intuitive or rational decision-making (see, Hamilton et al., 2016). Two rational and three intuitive items by Hamilton et al. (2016) and one rational item by Graßl et al. (2021) served as template. Both were translated and adapted to cookie decisions. *Privacy attitude* was measured with seven items on a Likert scale from 1 to 7 where each scale's legend varied. Six items by Dienlin and Trepte (2015) on general attitude towards privacy were reformulated for information dissemination via cookie banners and used as templates together with one item by Machuletz and Böhme (2020) on perceived importance. It should be noted that privacy attitude is not reported or included within the results as it does not belong to the reported research question. To verify a knowledge improvement by the cookie assistant in comparison to the app-advertisement (*treatment check*), participants had to examine cookie-statements as true or wrong and identify which cookie banners are mandatory or optional. All questions

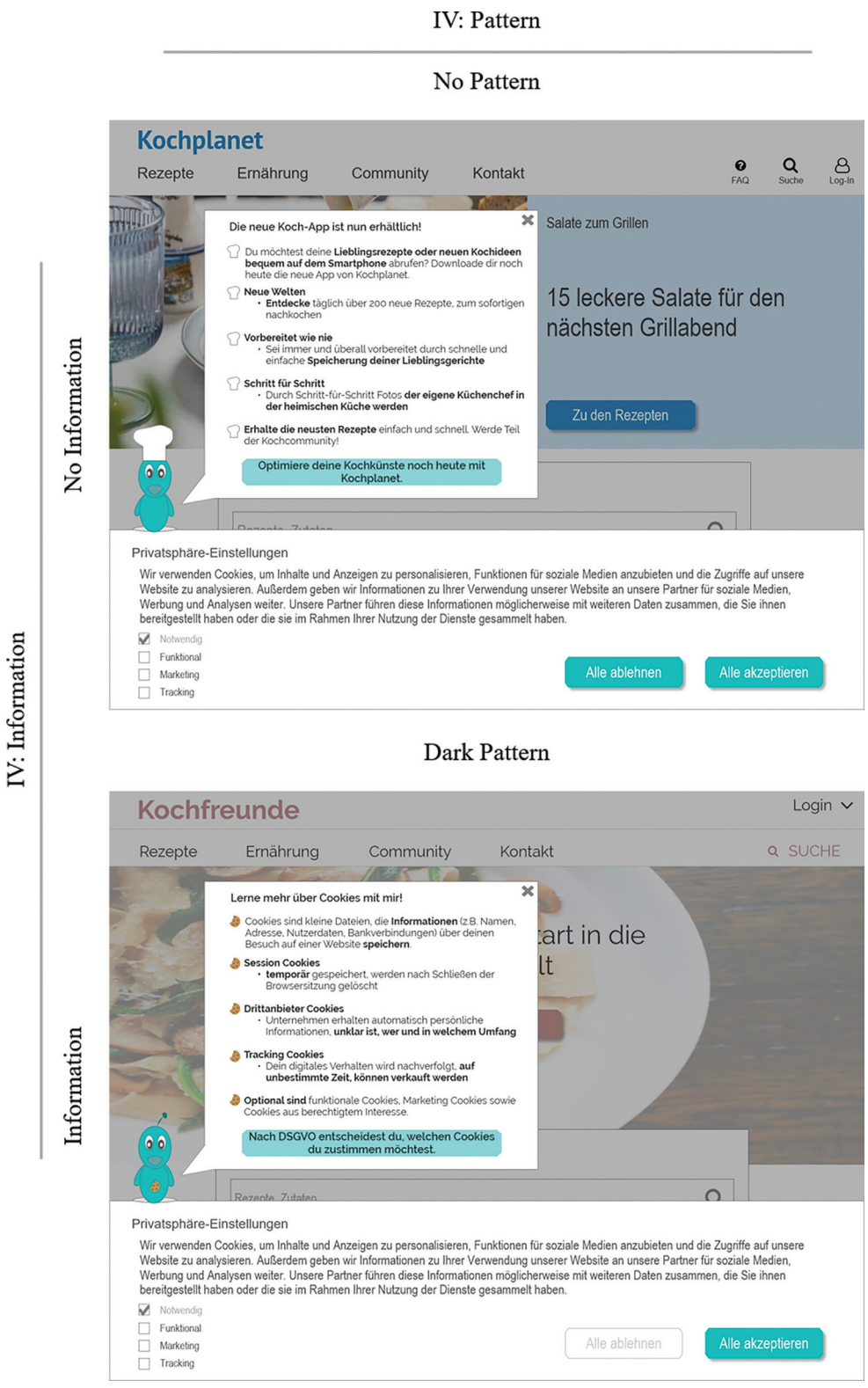


Figure 1. Stimuli examples for no or dark pattern as well as (no) information condition for blue, cooking planet, salad cooking webpage as well as brown, cooking friends, pasta cooking webpage (translations of stimuli webpages for both conditions (no information, information) are provided in Appendix A).

were answerable with the cookie assistant. Further, within the information condition, participants should rate their perception of the cookie assistant and its informational level with two items on a Likert scale from 1 (I do not agree at all) to 7 (I totally agree). The questionnaires used are listed in Appendix B.

Procedure

Participants needed 15 – 20 minutes to complete the experiment. They gave informed consent and agreed to the privacy declaration before participating. After consenting, participants reported their demographic data (age, gender) and digital media usage per week. Then, the experimental phase started. Participants were instructed to either agree or disagree to cookies and, afterwards, rate a website indicating their first impression on a Likert scale from 1 (terrible) to 7 (very good) as well as, optionally, one word that describes their first impression. Before each trial, participants had to click on "Go" (German: "Los") to calibrate the mouse to the screen middle. In each trial, participants were presented, first, a cookie banner until consent or rejection, then, the underlying website for three seconds and, lastly, a questionnaire querying their first impression (see Figure 2). Reaction time (RT) and cookie activation were measured whereas the first impression task was not recorded. In total, 28 cookie banners were randomized presented per participant. To familiarize participants with their tasks, one practice trial was carried out. After completing all trials, the study purpose was queried to identify whether the cover story worked as intended. Then, participants had to rate items on *deliberation* of their cookie decision, their respective knowledge about cookies (*treatment check*) as well as their *privacy attitude* (post-hoc questionnaire). Lastly, participants, in the information condition, were asked to rate their perception of the cookie assistant and its informational level (*treatment check*). Then, within the debriefing, the purpose of the study was disclosed and explained, participants received contact information and could, if desired, obtain the results after data completion.

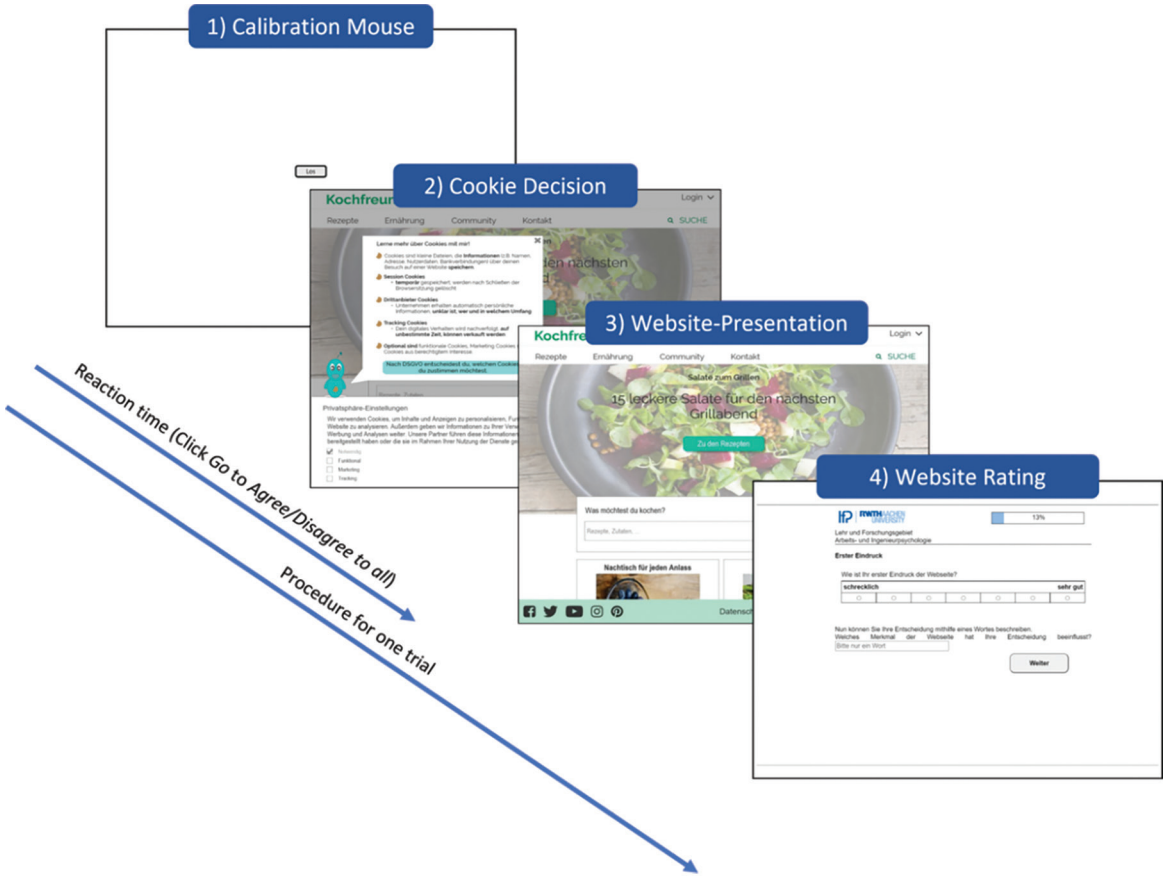


Figure 2. Experimental procedure for one trial representing information condition (cookie-assistant) and cooking webpage (green, cooking friends, salad).

Research ethics

All participants gave written informed consent and participation was voluntary. No undue physical or psychological stress was anticipated as well as participants did not take risks by participating in this study. Participants were able to end data collection at any time and without giving any reason or receiving any disadvantages. Furthermore, participants could skip questions regarding personal information (e.g., age), did not interact with real cookie banners and stimulus material was selected to not raise strong emotions (e.g., fear). Even if, participants were instructed that they should agree or disagree to cookie banners and that their usage behavior will be measured, using the cover story overlaid the full research purpose until the end of the study. However, to enable participants to revoke data storage, data was exclusively saved if participants clicked on "Continue and store data" after debriefing participants about the full research purpose and the applied cover story.

3. Results

For all analyses, the statistic software IBM SPSS Statistics (27.0) with an alpha level of α = .05 was used.

Deliberation and response time

For deliberation and RT analysis, two-sided *t*-tests for independent samples were conducted. Before investigating effects for deliberation, reliability analysis for the deliberation questionnaire was calculated and revealed questionable internal consistency (0.62). Since removing one specific item would not have increased internal consistency, all items were used in the following analysis and classified under this aspect in the discussion. To examine whether participants who received supporting information were more likely to make deliberate decisions on a cookie banner, a two-sided *t*-test for independent samples was conducted. No significant difference between the no information (M = 2.67, SD = 1.21) and the information condition (M = 2.58, SD = 1.12) was found regarding participants' deliberation, $t(205) = 0.61$, $p = .55$, $d = 0.08$.

For RT analysis, RTs were *z*-transformed per participant and trials with *z*-scores of -3/+3 (Field, 2013) as well as trials over 120 seconds served as exclusion criteria. Considering these criteria, a total of 2.8 % trials (164 out of 5796 trials) were excluded from RT analysis. Overall, participants took on average three seconds for a cookie decision (M = 2.90, SD = 3.18). A two-sided *t*-test for independent samples on RTs revealed no significant difference between both information conditions, $t(205) = 1.31$, $p = .19$, $d = 0.18$. Thus, participants, who received supporting information, did not take longer for their decision on a cookie banner (M = 2753 ms, SD = 1716 ms) than participants who did not receive supporting information (M = 3063 ms, SD = 1684 ms). Further, it was examined whether RT was influenced by the presence of dark patterns. However, a 2x2 mixed ANOVA with repeated

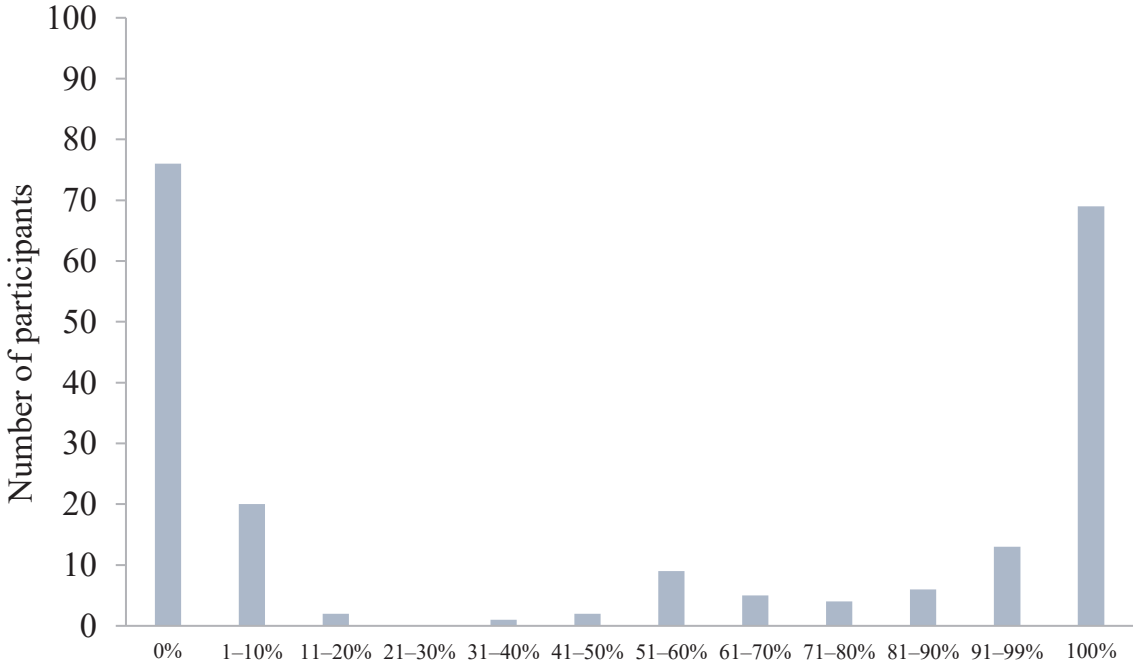


Figure 3. Frequency distribution of participants' cookie acceptance rate across

measures, capturing pattern (dark vs. no pattern) as within-subject factor and information condition (with vs. no information) as between-subjects' factors, did not reveal a significant main effect of pattern on RTs, $F<1$, indicating no dark pattern impact on RT to cookie banners. Further, no main effect of condition nor interaction between condition and pattern was found ($F \leq 1.72$; $p's \geq .19$).

Cookie activation

Before investigating effects on cookie activation, the relative frequency (in %) of cookie activation was averaged across all trials per participant. The frequency distribution (see Figure 3) presented that 36.7 % (N = 76) participants revealed constant cookie rejection behavior (0 % cookie acceptance) and 33.3 % (N = 69) participants revealed constant cookie acceptance behavior (100 % cookie acceptance). In total, this resulted in 70.0 % (N = 145) participants with constant cookie activation behavior. All participants were included in the following analysis. Afterwards, it was analyzed whether cookie activation differs for dark patterns in dependence of their assigned information condition. For trials including dark patterns, a two-sided t-test for independent samples revealed no significant difference in cookie activation when information was presented (M = 52.3 %, SD = 48.1) in comparison to the no information condition (M = 48.3 %, SD = 47.5), $t(205) = -0.60$, $p = .55$, $d = -.08$. Thus, no evidence was found that, in the presence of dark patterns, participants who received supporting information were less susceptible to cookies than participants without assisting information.

Further, it was assumed that cookie activation decreases when supporting information is presented in particular when dark patterns are present. To test this hypothesis, a 2x2 mixed ANOVA, with pattern (dark vs. no pattern) as within-subject factor and information (with vs. no information) as between-subjects factor on cookie activation was calculated. The ANOVA revealed the frequently shown impact of dark patterns on cookie activation, $F(1, 205) = 9.59$, $p < .01$, $\eta_p^2 = .05$. This significant main effect indicates that cookie banners with dark patterns were activated on average in M = 50.4 % (SD = 47.7) of all trials per participant whereas cookie banners with no patterns were only activated on average in M = 46.4 % (SD = 47.0) of all trials per participant. However, there was no significant interaction between pattern and information, $F(1, 205) = 1.06$, $p = .31$, $\eta_p^2 = .01$, and with this no evidence that supporting information reduced the influence of dark patterns on cookie activation (see also Figure 4). Descriptively, the difference even seems to proceed in the opposite direction. Further, no significant main effect for condition was found (no information: M = 46.4%, SD = 47.0; information: M = 50.4, SD = 47.7), $F < 1$.

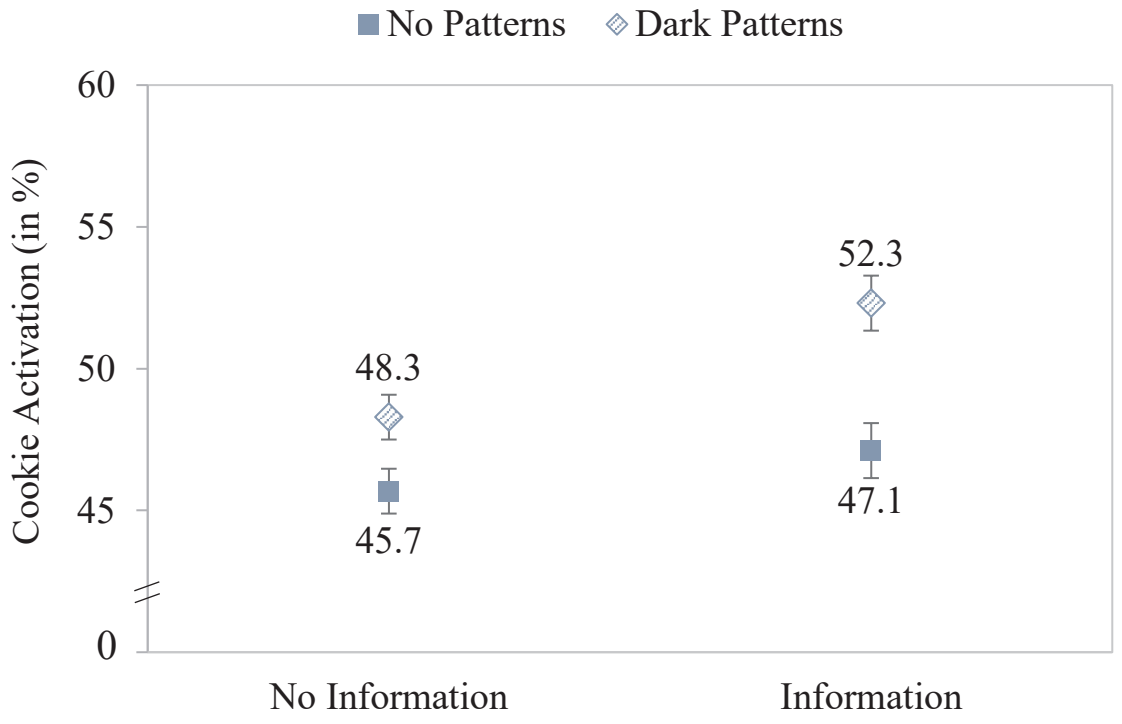


Figure 4. Cookie activation (in %) by information condition (between-subjects) and pattern (within-subject); error bars indicate standard errors within-subject as computed according to Cousineau (2005).

Treatment checks

To identify whether the cover story worked as intended, open answers on study purpose were coded and revealed that 163 participants indicated websites, e.g., its design, as study purpose compared to 41 participants who indicated cookies as study purpose.

To examine if information via the cookie assistant was read, it was analyzed whether participants' knowledge differed between conditions (information vs. no information). In accordance to Strycharz et al. (2021), correct answers were coded as 1 and incorrect answers were coded as 0. The calculated mean represented the knowledge score. However, participants in the information condition (M = 0.68, SD = 0.09) did not have significantly greater knowledge than participants who received no information (M = 0.68, SD = 0.10), $t(205)$ = -0.07, p = .94, d = -.01.

To measure cookie assistance and its informational level, participants' impression of the cookie assistant was recorded with two items (α = .78). Results revealed "medium" consent to the cookie assistants' informational level with M = 3.08 (SD = 1.73) on a scale from 1 (I do not agree at all) to 7 (I totally agree).

4. Discussion

The present online experiment examined the influence of a knowledge intervention and its effectiveness to encourage deliberate privacy decisions when dealing with dark patterns in web cookie banners. Additionally, it was examined whether an increase in users' knowledge can empower them to make privacy-friendly decisions regardless of dark patterns. The empirical findings, however, revealed no evidence for a successful intervention as neither more deliberate and slower nor more privacy-friendly decisions were made when participants received supporting information via a cookie assistant. By integrating current theoretical and empirical findings on dark patterns, the results, potential influencing factors and practical implications are discussed below.

Consonant with various studies (e.g., Machuletz & Böhme, 2020; Nouwens et al., 2020; Utz et al., 2019), we found evidence that dark patterns increased participants' acceptance rate of cookies, underlining their impact on privacy decisions. More important, the present study lends further support to the problematic nature of dark patterns as results demonstrate their resilience against interventions. Focusing on participants' deliberation and time required for cookie decisions, we expected that the knowledge intervention should benefit users' *System 2* processing and encourage their deliberation on and time required for a cookie decision. However, no evidence was found in this respect. On the contrary, independent of whether or not cookie assistance was given, participants had low to medium deliberation scores and made fast cookie decisions. Consistent with the findings on reaction time and deliberation, but contrary to our expectations, no impact of the given information on cookie activation in general nor on the influence of dark patterns was given. However, it should be noted that the majority of our participants (about 70 %) showed constant cookie related behavior, i.e. they either accepted or rejected all cookies. This might be an indicator for previously defined cookie strategies or conditioned behavior. Overall, no evidence was found that the developed knowledge intervention promoted users' deliberate *System 2* processing or enhanced users' ability to counteract dark patterns. Even if the informational level of the cookie assistant revealed medium scores and the cover story worked as intended (163 participants indicated the websites as study purpose), these results and similar knowledge scores between both conditions demonstrated that the implemented knowledge intervention did not work as intended.

One reason could be, that the present study conveys unsuitable information or fails to convey information at all. As the knowledge intervention referred to topic-relevant knowledge, we believe that the information itself should have been able to increase users' performance on the knowledge questions, especially since the fit between its content and the control questions was large. In contrast to this expectation, the subsequent questionnaire did not show differences in participants' knowledge level between participants who received the information and those who did not. This might indicate, that the former participants might have ignored the information because it lacked visual salience, was too extensive or they wanted to get rid of it as quickly as possible to return to the primary task: rating the webpages. As mentioned above, users in *System 1* processing do not deliberate their decision and tend to focus on the highlighted and simplest option, which is one reason why dark pattern work as intended (Bösch et al., 2016). However, the present results demonstrate how difficult it is to lead users to *System 2* processing in the first place and knowledge interventions alone might not be the right tool to combat highly automated behavior such as clicking on cookie banners. Other visualizations for knowledge interventions must be considered that have higher visual salience and are able to grab users' attention even under these difficult circumstances. To achieve this goal, the intervention should "pop-out", e.g. by using vivid and high contrast colors or animations. It also has to be visually different from irrelevant information, for example adds, so that the intervention is not mistaken as distractor. One approach could be to integrate *privacy-by-design* approaches (e.g., Barth et al., 2021) and present privacy locks above the cookie buttons or information presented right above the cookie banner.

As about 70 % of the participants revealed constant response behavior, it could be argued that participants developed strongly conditioned behavior on cookie requests over time. The

important role of conditioned behavior is also highlighted by Graßl et al. (2021) who found indications that users developed accepting default behavior on cookie banners independent of dark patterns. Such highly conditioned behavior raises the question whether knowledge interventions alone can be effective, as users must overcome their conditioned behavior to interact with knowledge interventions. As motivation is described as one critical factor to trigger *System 2* processing (Bösch et al., 2016; Brough & Martin, 2020; Terpstra et al., 2019) or to change users' behavior in general (Fogg, 2009), Brough and Martin (2020), for example, argue that knowledge differences need to be combined with motivation to influence users' privacy behavior (see also Fogg, 2009; Terpstra et al., 2019). To encourage users' privacy decisions, Terpstra et al. (2019) further claim that three components must be fulfilled: (1) an interruption of the user's task or goal, (2) an explanation on, e.g., how and why data is stored, and (3) the option to make own choices. Although participants received supporting information to indicate how data is stored and the ability to make their own cookie decisions in the present study, the missing effectiveness of the knowledge intervention could be reasoned in conditioned behavior that cannot be conquered due to insufficient motivation or stimulation (e.g. through an interruption) to engage in reflective thinking. To test this assumption, follow-up investigations could examine whether inquiries (e.g., "Are you sure you want your privacy settings to be saved?") (Terpstra et al., 2019), precise risk communication (e.g. specific risk scenarios, Gerber et al., 2019) or fear appeals (Brough & Martin, 2020; Distler et al., 2020; Harbach et al., 2014) in combination with knowledge interventions could raise awareness and trigger users' *System 2* processing. As some participants who focused on the cover story (website ratings) indicated that the study increased their motivation to change cookie decisions or to learn more about cookies, it could further be examined whether induced discrepancies between users' attitude and behavior could improve motivation for privacy decisions.

Limitations

Besides these alternative explanations and research suggestions, the given results should be considered in view of potential methodological limitations. First, it could be argued that the knowledge intervention did not show the desired effects as the research context provided a certain level of security. Rather than showing their regular privacy behavior, the resulting feeling of safety in the context of a scientific experiment could provoke participants to mainly focus on the primary task (rating websites). However, as participants used their own devices, and as it was not indicated that faked cookies were used until the end of the experiment, this alternative explanation cannot fully explain the results. Moreover, it cannot be ruled out that the present study showed too little internal consistency in the deliberation questionnaire. However, as no effect of the knowledge intervention was found on RT or cookie activation as well as the majority of our participants showed constant cookie related behavior, these methodological limitations do not seem to strongly affect the present results.

Practical implications

As results indicate that participants made highly conditioned, undeliberated, and fast decisions on the ubiquitous consent requests of cookie banners, the question arises whether state-of-the-art repetitive cookie requests are purposeful. Even if the GDPR attempts to provide users' with "freely given, specific, informed and unambiguous" (GDPR, 2016, p. 6) privacy decisions, the resulting repetitive cookie requests could be associated to conditioned rather than informed privacy decisions (see, Graßl et al., 2021). Such results underline the importance of new legal developments such as *PIMS* (*P*ersonal *I*nformation *M*anagement *S*ystems), where users declare their consent once and website operators will be informed on their consent (see also, John & Rennert, 2022). Such one-time privacy decisions could reduce the effort required and increase users' motivation to deliberately decide on their privacy settings. As website operators would still be entitled to request individual and customizable privacy decisions under the new legislation (see also, John & Rennert, 2022), it should be ensured that these customizable cookie decisions will not be enforced to visit a website or connected to dark patterns in order to gain users' data.

Further, the present study indicates that participants lacked the knowledge to make privacy decisions on cookie requests. These insights into users' cookie decisions highlight the need for an increase of users' knowledge on privacy and cookie requests. Hence, legal regulations should not only ensure informative and easily understandable consent requests through *educative nudges* (Sunstein & Reisch, 2019) but also improve users' digital literacy with specific trainings (*long-term boosts*) (Hertwig & Grüne-Yanoff, 2017). Further trainings (*long-term boosts*) on, e.g., users' digital literacy could be developed as educational measure. For example, app-based privacy trainings (e.g. FoxIt; Gerber et al., 2018) or serious games (as developed, for example, within the project A-DigiKomp: https://www.a-digikomp.rwth-aachen.de) might be suitable to playfully improve digital literacy.

Conclusion

By demonstrating that participants make undeliberated, fast as well as privacy-unfriendly decisions regardless of supporting information, the present results question the effectiveness of, so far, implemented consent requests on cookie banners and the presented knowledge intervention. As users seem to develop highly conditioned behavior on

repetitive privacy decisions, we argue for a reformation of current cookie requests in particular and more research on privacy-enhancing interventions by taking simplified knowledge as well as motivation into account.

References

Acquisti, A., Adjerid, I., Balebako, R., Brandimarte, L., Cranor, L. F., Komanduri, S., Leon, P. G., Sadeh, N., Schaub, F., Sleeper, M., Wang, Y., & Wilson, S. (2017). Nudges for privacy and security: Understanding and assisting users' choices online. *ACM Computing Surveys*, *50*(3), 1–41. https://doi.org/10.1145/3054926

Acquisti, A., John, L. K., & Loewenstein, G. (2013). What is privacy worth? *The Journal of Legal Studies*, *42*(2), 249–274. https://doi.org/10.1086/671754

Barth, S., & de Jong, M. D. T. (2017). The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review. *Telematics and Informatics*, *34*(7), 1038–1058. https://doi.org/10.1016/j.tele.2017.04.013

Barth, S., Ionita, D., De Jong, M. D., Hartel, P. H., & Junger, M. (2021). Privacy rating: a user-centered approach for visualizing data handling practices of online services. *IEEE transactions on professional communication*, *64*(4), 354–373. https://doi.org/10.1109/TPC.2021.3110617

Bongard-Blanchy, K., Rossi, A., Rivas, S., Doublet, S., Koenig, V., & Lenzini, G. (2021). "I am definitely manipulated, even when I am aware of it. It's ridiculous!" – Dark patterns from the end-user perspective. In W. Ju, L. Oehlberg, S. Follmer, S. Fox, S. Kuznetsov (Eds.), *Designing Interactive Systems Conference 2021 (DIS'21)* (pp. 763–776). Association for Computing Machinery. https://doi.org/10.1145/3461778.3462086

Bösch, C., Erb, B., Kargl, F., Kopp, H., & Pfattheicher, S. (2016). Tales from the dark side: Privacy dark strategies and privacy dark patterns. *Proceedings on Privacy Enhancing Technologies*, *2016*(4), 237–254. https://doi.org/10.1515/popets-2016-0038

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414. https://doi.org/10.7717/peerj.9414

Brough, A. R., & Martin K. D. (2020). Critical roles of knowledge and motivation in privacy research. *Current Opinion in Psychology*, *31*, 11-15. https://doi.org/10.1016/j.copsyc.2019.06.021

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Lawrence Erlbaum Associates.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Culnan, M. J., & Armstrong, P. (1999). Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization Science*, *10*(1), 104-115. https://doi.org/10.1287/orsc.10.1.104

Dennis, A. R., & Minas, R. K. (2018). Security on autopilot: Why current security theories hijack our thinking and lead us astray. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *49*(SI), 15–38. https://doi.org/10.1145/3210530.3210533

Di Geronimo, L., Braz, L., Fregnan, E., Palomba, F., & Bacchelli, A. (2020). UI dark patterns and where to find them: A study on mobile applications and user perception. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)* (pp. 1–14). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376600

Dienlin, T., & Trepte, S. (2015). Is the privacy paradox a relic of the past? An in-depth analysis of privacy attitudes and privacy behaviors. *European Journal of Social Psychology*, *45*(3), 285–297. https://doi.org/10.1002/ejsp.2049

Distler, V., Lenzini, G., Lallemand, C., & Koenig, V. (2020). The framework of security-enhancing friction: How UX can help users behave more securely. In *New Security Paradigms Workshop 2020 (NSPW '20)* (pp. 45–58). Association for Computing Machinery. https://doi.org/10.1145/3442167.3442173

Fansher, M., Chivukula, S. S., & Gray, C. M. (2018). #darkpatterns: UX practitioner conversations about ethical design. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)* (pp. 1–6). Association for Computing Machinery. https://doi.org/10.1145/3170427.3188553

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/BF03193146

Field, A. (2013). *Discovering statistics using IBM SPSS statistics.* (3rd ed.). Sage.

Fogg, B. J. (2009). A behavior model for persuasive design. In *Proceedings of the 4th International Conference on Persuasive Technology (Persuasive '09)* (pp. 1–7). Association for Computing Machinery. https://doi.org/10.1145/1541948.1541999

General Data Protection Regulation (GDPR) (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=DE

Gerber, N., Reinheimer, B., & Volkamer, M. (2019). Investigating People's Privacy Risk Perception. *Proceedings on Privacy Enhancing Technologies*, 2019(3), 267-288. https://doi.org/10.2478/popets-2019-0047

Gerber, N., Gerber, P., Drews, H., Kirchner, E., Schlegel, N., Schmidt, T., & Scholz, L. (2018). FoxIT: enhancing mobile users' privacy behavior by increasing knowledge and awareness. In *Proceedings of the 7th Workshop on Socio-Technical Aspects in Security and Trust* (pp. 53-63).

Graßl, P., Schraffenberger, H., Borgesius, F. Z., & Buijzen, M. (2021). Dark and bright patterns in cookie consent requests. *Journal of Digital Social Research*, *3*(1), 1–38. https://doi.org/10.31234/osf.io/gqs5h

Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018). The dark (patterns) side of UX design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)* (pp. 1–14). Association for Computing Machinery. https://doi.org/10.1145/3173574.3174108

Gross, R., & Acquisti, A. (2005). Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society (WPES '05)* (pp. 71–80). Association for Computing Machinery. https://doi.org/10.1145/1102199.1102214

Hamilton, K., Shih, S.-I., & Mohammed, S. (2016). The development and validation of the rational and intuitive decision styles scale. *Journal of Personality Assessment*, *98*(5), 523–535. https://doi.org/10.1080/00223891.2015.1132426

Harbach, M., Hettig, M., Weber, S., & Smith, M. (2014). Using personal examples to improve risk communication for security and privacy decisions. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '14)* (pp. 2647–2656). Association for Computing Machinery. https://doi.org/10.1145/2556288.2556978

Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, *12*(6), 973–986. https://doi.org/10.1177/1745691617702496

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, *93*(5), 1449–1475. https://doi.org/10.1257/000282803322655392

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, *64*, 122–134. https://doi.org/10.1016/j.cose.2015.07.002

Machuletz, D., & Böhme, R. (2020). Multiple purposes, multiple problems: A user study of consent dialogs after GDPR. *Proceedings on Privacy Enhancing Technologies*, *2020*(2), 481–498. https://doi.org/10.2478/popets-2020-0037

Maier, M., & Harr, R. (2020). Dark design patterns: An end-user perspective. *Human Technology*, *16*(2), 170–199. https://doi.org/10.17011/ht/urn.202008245641

Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–32. https://doi.org/10.1145/3359183

Mathur, A., Kshirsagar, M., & Mayer, J. (2021). What makes a dark pattern... dark?: Design attributes, normative considerations, and measurement methods. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)* (pp. 1–18). Association for Computing Machinery. https://doi.org/10.1145/3411764.3445610

Nouwens, M., Liccardi, I., Veale, M., Karger, D., & Kagal, L. (2020). Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)* (pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376321

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Sanchez-Rola, I., Dell'Amico, M., Kotzias, P., Balzarotti, D., Bilge, L., Vervier, P.-A., & Santos, I. (2019). Can I opt out yet?: GDPR and the global illusion of cookie control. In *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security (Asia CCS '19)* (pp. 340–351). Association for Computing Machinery. https://doi.org/10.1145/3321705.3329806

Strycharz, J., Smit, E., Helberger, N., & van Noort, G. (2021). No to cookies: Empowering impact of technical and legal knowledge on rejecting tracking cookies. *Computers in Human Behavior*, *120*, 106750. https://doi.org/10.1016/j.chb.2021.106750

Sunstein, C. R. (2016). *The ethics of influence: Government in the age of behavioral science*. Cambridge University Press. https://doi.org/10.1017/CBO9781316493021

Sunstein, C. R., & Reisch, L. A. (2019). Educative nudges and noneducative nudges. In C. R. Sunstein & L. A. Reisch (Eds.) *Trusting Nudges* (1st ed., pp. 95-118). Routledge. https://doi.org/10.4324/9780429451645

Terpstra, A., Schouten, A. P., de Rooij, A., & Leenes, R. E. (2019). Improving privacy choice through design: How designing for reflection could support privacy self-management. *First Monday*, *24*(7), 1–19. https://doi.org/10.5210/fm.v24i7.9358

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Utz, C., Degeling, M., Fahl, S., Schaub, F., & Holz, T. (2019). (Un)informed consent: Studying GDPR consent notices in the field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS '19)* (pp. 973–990). Association for Computing Machinery. https://doi.org/10.1145/3319535.3354212

Verbraucherzentrale. (2021). *Cookies kontrollieren und verwalten*. Verbraucherzentrale.de. https://www.verbraucherzentrale.de/wissen/digitale-welt/datenschutz/cookies-kontrollieren-und-verwalten-11996.

Appendix A

Information text (translated to English)

Table Appendix A:

Condition	Information text	Source
Information (Cookie Assistant)	Learn about cookies with me! – Cookies are small files that *store information* (e.g. name, address, user data, bank details) about your visit on a website. – **Session cookies** ○ **Temporarily** stored, are deleted after closing the browser session – **Third party cookies** ○ Companies automatically receive personal information, it is unclear who and to what extent – **Tracking cookies** ○ Your digital behavior is tracked, for an **indefinite period, can be sold** – **Optional are** functional cookies, marketing cookies and cookies for legitimate interest. According to DSGVO, you decide to which cookies you want to agree.	The cookie information was based on the German Consumer Association (Verbraucherzentrale, 2021)
No Information (App Advertisment)	The new cooking app is now available! – You want to access your f**avorite recipes or new cooking ideas conveniently on your smartphone?** Download the new app from Kochplanet/Kochfreude today. – **New worlds** ○ **Discover** over 200 new recipes every day for instant re-cooking – **Prepared as never before** ○ Be prepared anytime, anywhere with quick and easy **storage of your favorite dishes** – **Step by step** ○ **Become your own chef in the home kitchen** through step-by-step photos – **Get the newest recipes** easily and quickly. Become part of the cooking community! Optimize your cooking skills today with Kochplanet/Kochfreunde.	

Appendix B

Questionnaires (translated to English)

Table Appendix B:

Variable	Measurement Item	Source
First impression	*[Likert scale, (1) terrible to (7) very good; open answer]* – What is your first impression of the website? – optional: What feature of the website influenced your decision?	
Study purpose	*[open answer]* – What do you think this study is about?	
Deliberation	*[Likert scale, (1) I do not agree at all to (7) I totally agree]* – Before selecting a cookie option, I considered all possible cookie options. – Before choosing a cookie option, I took time to weigh the advantages and disadvantages of my decision. – Before selecting a cookie option, I carefully read the cookie policy. – I intuitively selected a cookie option. – I selected the cookie option, which corresponds to my first impression. – I trusted in my feeling when I chose a cookie option.	Hamilton et al. (2016) Graßl et al. (2021) served as template
Knowledge	*[single choice; true, wrong]* Cookies are … … security patches designed to close security issues on the computer. … text information that is stored on a device when a website is visited. … another word for privacy policies. Session cookies… … are deleted after each browser session. … are stored temporarily. … automatically share personal information with other companies. … can be sold to other companies. … can store digital behavior over multiple browser sessions. Third party cookies… … are deleted after each browser session.	Encoding was adopted from Strycharz et al. (2021) Cookie information was based on Verbraucherzentrale (2021)
Knowledge	… are stored temporarily. … automatically share personal information with other companies. … can be sold to other companies. … can store digital behavior over multiple browser sessions. Tracking cookies… … are deleted after each browser session. … are stored temporarily. … automatically share personal information with other companies. … can be sold to other companies. … can store digital behavior over multiple browser sessions. What types of cookies are optional? … necessary cookies … functional cookies … marketing cookies … cookies for legitimate interest	Encoding was adopted from Strycharz et al. (2021) Cookie information was based on Verbraucherzentrale (2021)
Privacy Attitude	*[Likert scale; 1 to 7, alternating response format]* I think that sharing personal information via cookies on a website ... … is not meaningful (1) to meaningful (7). … is disadvantageous (1) to advantageous (7). … is questionable (1) to harmless (7). … is dangerous (1) to not dangerous (7). … is reckless (1) to deliberate (7). … is bad (1) to good (7). The protection of my privacy on the Internet is... … unimportant (1) to important (7).	Dienlin and Trepte (2015) Machuletz and Böhme (2020) served as template
Cookie Assistant	*[Likert scale, (1) I do not agree at all to (7) I totally agree]* – The Privacy Assistant helped me to select a cookie option. – Through the Privacy Assistant, I have received helpful information on the subject of cookies.	

DOI: 10.2478/9788366675896-005 • Technikpsychologie • 55-65

sciendo

Proceedings of the International Symposium on Technikpsychologie

The Technology Acceptance Model (TAM) and its Importance for Digitalization Research: A Review

Angela Schorr†

Department of Psychology, Faculty of Life Sciences, Media and Educational Psychology, Media Psychology Lab, University of Siegen, Adolf-Reichwein-Str. 2a, 57072 Siegen, Germany

Abstract

Davis' Technology Acceptance Model (TAM) plays a major role in research on the acceptance of digital technologies (Davis, 1986; 2015; Venkatesh & Davis, 2000; Venkatesh et al., 2012). It is heavily influenced by the Theory of Planned Behavior by Fishbein and Ajzen (Ajzen, 2020) and is still rated "the most popular theoretical framework" for research on technology adoption (Feng et al., 2021). Davis was one of the first to deal with the topic of user reactions to digital technologies in the 1980s. During his career, his main topic of research was to improve the understanding of user acceptance processes, thereby providing a theoretical basis for practical technology acceptance testing. The evolution of TAM1 to TAM3 is presented briefly in this review. Also, alternative technology acceptance models are reviewed, including the Unified Theory of Acceptance and Use of Technology (UTAUT) that evolved from TAM3 (Venkatesh et al., 2003; 2012). Until today, both theories TAM and UTAUT are used extensively in digital transformation research in business, media, education and health. While the peaceful coexistence of both theories is amazing, this success story also raises a number of questions. A closer analysis focusing on the TAM reveals conceptual and methodological weaknesses characteristic of both models, which urgently need to be fixed to make progress in the entire field of digital technology acceptance research.

Keywords

technology acceptance model • digital technology • digitalization • TAM • UTAUT

1. Technology and technology acceptance

Before turning to the topic of digital technology acceptance, it is useful to briefly address the concept of technology itself. Standard definitions equate *technology* with applied science as the basis for products and production processes. In 2009, the economist W. Brian Arthur published a broader, future-oriented view of the concept of technology in the monograph "The Nature of Technology": Arthur defines technology as "a means to fulfill a human purpose" and adds to that, "as a means, a technology may be a method or process or device" (Arthur, 2009, p. 28). According to Arthur, people place a lot of hope in technology, but don't really trust it. They are aware of the difference between "technology as enslaving our nature versus technology as extending our nature" (Arthur, 2009, p. 215-216). Driven by the high pace of digitalization, the understanding of technology is changing inside and outside of science: In everyday life, users expect technologies to work reliably. They must meet requirements such as utility, usability, and safety. To achieve some practical result, the development of a technology may draw upon many fields of knowledge.

At the beginning of technology acceptance research the failure rate of information technology applications was very high; large software companies made huge financial losses from systems that were rejected (Davis 2015). Davis, an MIT graduate in business engineering, was one of the first to deal with the topic of user reactions to *digital* technologies ("computer-based information systems", Davis, 1986, p. 7). His aim was to find out how system design characteristics influence user motivation. The Technology Acceptance Model was intended to provide the theoretical basis for practical user acceptance testing developers of new hardware and/or software could use to evaluate digital technologies prior to and during implementation (Davis, 2015).

Fred Davis' original theory on technology acceptance had a simple structure: While reflecting on the interviews with end-users he realized that two dominant reasons were given to reject a new system. These two key beliefs, *perceived usefulness* and *perceived ease of use* guided his further research: The two beliefs form the attitude

†*Corresponding author*: *Angela Schorr*
E-mail: angela.schorr@uni-siegen.de

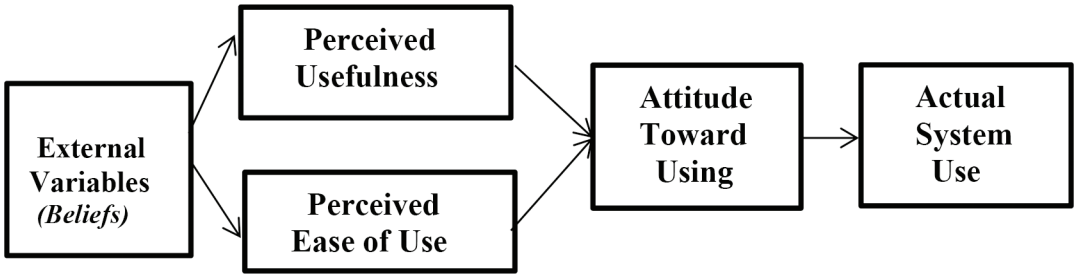


Figure 1. First version of the Technology Acceptance Model (TAM; Davis, 1986).

(*attitude toward using*), which is followed by a *behavioral reaction*, i.e., the *actual use* of the technology. These four variables make up technology acceptance, i.e. "how users are motivated to use the system" (Fig.1; Davis, 1986, p. 11). In an effort to provide further conceptual support for the model and to improve the model's predictive power with the help of the Theory of Reasoned Action (TRA, Fishbein & Ajzen, 1975; Ajzen & Fishbein, 1980) and its successor, the Theory of Planned Behavior (TRB, Ajzen, 1985; 2020), major efforts were made in subsequent years to investigate the influence of additional "determinants" on the two key beliefs. In contrast to the approach of the TRA/TRB authors, Davis did not determine their relative influence on user behavior by leaving the weighting to the test subjects, but determined it using regression analyses. With many collaborators, most notably Visvanath Venkatesh, he expanded the basic model (TAM 1-TAM3) and triggered a flood of empirical studies from all over the world that has not abated to this day.

Today there is a whole range of technology adoption models that are used in organizational change projects, in digitization research, as well as in consumer, health, and educational research. There are numerous concepts that overlap in terms of content with the concept of technology acceptance. These include technology readiness, resistance to change, technology threat avoidance, technostress, digitalization anxiety, technology commitment, readiness for technology, technology affinity, etc. Table 1 provides an overview of theoretical models of digital technology adoption research, as well as the basic theories from various sciences that were used for their development. *However, none of these models has generated as much scientific research on digital technology acceptance as the TAM (TAM1-3) and the UTAUT (1/2)*.

Table 1. Models in Digital Technology Adoption Research

1. **Basic theories of psychology relevant to technology acceptance research:**
   Theory of Reasoned Action, TRA (Fishbein & Ajzen, 1975; 1980)
   Theory of Planned Behavior, TPB (Ajzen, 1991; 2020)
   Social Cognitive Theory of Self-Regulation, SCT (Bandura, 1977, 2001)
2. **Theories from communication studies, information sciences, ergonomics, and economics:**
   Diffusion of Innovations Theory, DIT (Rogers, 2003)
   Usability Framework (Eason, 1984)
   Task-Technology-Fit Model, TTF (Goodhue & Thompson, 1995; Goodhue, 1998)
3. **Psychological theories on the acceptance of digital technologies:**
   Technology Acceptance Model, TAM (Davis, 1986; Davis, Bagozzi, & Warshaw, 1992)
   Technology Acceptance Model, TAM2 (Venkatesh & Davis, 2000)
   Technology Acceptance Model, TAM3 (Venkatesh & Bala, 2008)
   Decomposed Theory of Planned Behavior (combining TAM and TPB), DTPB (Taylor & Todd, 1995a)
   Unified Theory of Acceptance and Use of Technology, UTAUT (Venkatesh, Morris, Davis, & Davis, 2003)
   Unified Theory of Acceptance and Use of Technology, UTAUT2 (Venkatesh, Thong, & Xu, 2012)
4. **Theories dealing with subtopics of digital technology acceptance:**
   Technology Readiness (Parasuraman, 2000; Blut & Wang, 2020)
   Resistance to IT Change Index, RTCI (Davis & Songer, 2008)
   Technology Threat Avoidance Theory, TTAT (Liang & Xue, 2009)

2. TAM extensions: the next steps

In Davis' first study on technology acceptance from 1986, Davis could not objectively record the frequency of use of the software in the MIT data center and asked the test subjects about it (self-reported use). At the time, however, his declared goal was to objectively record the use of technology. In order to test the model structure of the TAM and to compare the prediction accuracy of TRA and TAM, Davis, Bagozzi and Warshaw (1989) carried out a project in which components of both models were to be examined for their usefulness in predicting technology acceptance. It turned out that there was a significant correlation between the TRA variable *behavioral intention to use* and *user behavior*. At all measurement times (it was a repeated measurement design) the outstanding role of *perceived usefulness* as a determinant of *behavioral intention to use* was confirmed. Furthermore, the key variable *perceived ease of use* had a direct impact on the *behavioral intention to use*. It was therefore decided to include the variable *behavioral intention to use* in the model (see Fig. 2) – a decision that triggered a lot of controversy from a methodological point of view.

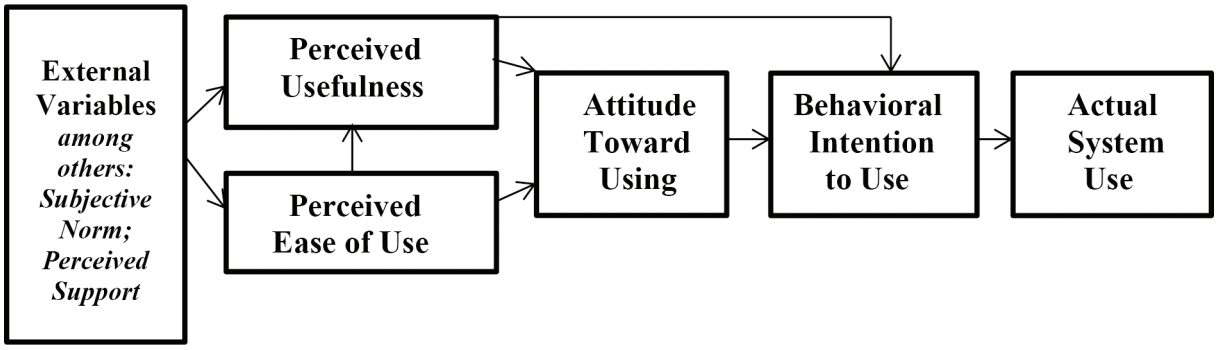


Figure 2. Technology Acceptance Model (TAM2; Davis, Bagozzi, & Warshaw, 1989).

To this day, the standard procedure in research on digital technology acceptance is to rely on survey-based, subjective information on user behavior. Particular attention is attached to the measurement of the *behavioral intention to use*. In the majority of research projects on technology adoption research, the behavioral intention is recorded. Occasionally, the *use* of the system is alternatively queried (thereby mostly relying on "self-reported current usage"). Many important hypotheses on the acceptance of digital technology were confirmed in this way, always under the premise that the information on use/usage behavior is based on the subjective feedback from test subjects. Davis himself defended his decision to rely on subjective information on user behavior by pointing out that the user's subjective perception is more relevant for his/her decision whether or not to use the system (Davis, 1989; 2015). Presumably, however, practical considerations may also have influenced this decision: It was easier to test the acceptance of digital systems during the implementation phase in companies, schools and elsewhere by self-report (Turner et al., 2010). Also, research results were easier to compare across studies if the same instruments (self-report scales) were used.

Critics of the survey strategy found differences between the subjective information provided by the user and the objective recording of use in their data (Szajna, 1996; Horton et al., 2001; Sharma & Yetton 2001; Olbrecht, 2010; Lai, 2017; Rahimi et al., 2018). For example, Szajna found that while the self-reported TAM key beliefs are valuable tools to predict the (self-reported) *intention to use an information system*, self-reported usage "may not be an appropriate surrogate measure for actual usage" (Szajna, 1996, p. 85). She rated the moderately significant correlations between the two variables she found in her longitudinal study as "relatively weak support for the convergent validity of self-report usage with actual usage" (Szajna, 1996, p. 89). Similarly, Horton, Buck, Waterson, and Clegg (2001, p. 237) concluded that "self-report and actual measures of usage are not interchangeable when applying such a model". Evaluating the data from 32 TAM studies, Sharma and Yetton (2001) confirmed that there are significant correlations in the moderate range between behavioral measures and subjective data.

In 2010 Turner et al. published a systematic review starting with the critical question "Does TAM predict actual use?" Based on the analysis of 79 empirical studies – of which, however, only a few had recorded actual use behavior – they came to the conclusion that the variable *behavioral intention to use* correlates reliably with *actual usage*, comparable to the two key beliefs in this respect (Turner, Kitchenham, Brereton, Charters, & Budgen, 2010). Hence, it can be concluded that the technology acceptance model is suitable for predicting the *behavioral intention to use* or the *self-reported use* by the participants, the actual use can deviate from it. In future research, to have a comprehensive understanding of technology acceptance and use, it may make sense to use multiple methods of data collection. Technically, recording user behavior/using learning analytics while testing a new digital system is no longer a problem today (Mothukuri et al., 2017).

3. Abandoning the variable *attitude toward using* (TAM 2/3)

New variables that Davis, Venkatesh and their team put to the test in addition to the key variables *perceived ease of use*, *perceived usefulness* and *behavioral intention to use* in TAM2 (presented in Davis et al., 1989; Venkatesh & Davis, 2000) and TAM3 (presented in Venkatesh & Bala, 2008) are among others the variables *subjective norm*, *voluntariness*, *facilitating conditions* and *self-efficacy*.

The variable the research group *decided to omit* was the TRA/TPB variable *attitude toward using*, a core element of the original TAM. This variable was assumed to be a major determinant of whether a user would use or reject a system. However, the expected high correlations with the two key beliefs *ease of use* and *perceived usefulness* were not

confirmed. In order to keep the model simple, Davis decided to dispense with this variable (Davis, 1989; Marangunić & Granić, 2015).

Research results of recent studies call this decision into question: Fussell and Truong (2022) were able to prove in their study on dynamic learning in VR (flight training) that the original TAM variables, namely *ease of use*, *perceived usefulness* and *attitude towards using* had the strongest relationships to *behavioral intention to use*. Similarly, a meta-analytic review by Feng et al. (2021) discovered a strong effect of *attitudes toward use* on *use intentions*. Additionally, a proposal for the revision of the related UTAUT model by Dwivedi et al. (2019, p.719) included the attitude variable as "central to behavior intentions and usage behaviors". This conclusion was based on an extensive meta-analysis. The authors of the study are convinced that the *attitude toward using* plays a central role in acceptance and use of IS/IT innovations (Dwivedi, Rana, Jeyaraj, Clement, & Williams, 2019; see also Marangunić & Granić, 2015). In summary, recent studies have shown that the attitude towards using variable has a strong relationship with behavioral intention to use, and it has been proposed to include this variable in this and related models, highlighting its importance in understanding users' behavior when interacting with digital technologies.

4. Extensions of TAM by including the variables *Subjective Norm* and *Voluntariness* (TAM 2/3)

In order to improve the prediction quality of the TAM, Venkatesh and Davis (2000) extended the technology acceptance model, the new TAM2, by adding a bundle of new variables: subjective norm, image, job relevance, output quality, and result demonstrability. Ajzen and Fishbein defined *subjective norm* as "the person's (...) perception that most people who are important to him think he or she should or should not perform the behavior in question" (Ajzen & Fishbein, 1980, p. 57). According to these authors, both personal attitudes and the perceived social environment can influence the subjective norm. The empirical results of this first revision of the Technology Acceptance Model confirmed Venkatesh and Davis' selection of the new variables. They were fascinated by the influence of the subjective norm on the key beliefs, which led to new discoveries: Building on insights from Taylor and Todd (1995a, 1995b), Venkatesh and Davis became aware of the influence of *voluntariness* on the process of acceptance. *They assumed that the subjective norm has a direct influence on the behavioral intention to use if the use of a system is not voluntary, but prescribed.* In case it is prescribed, they suspected that the influence of the subjective norm on the variable behavioral intention to use and perceived usefulness decreases with increasing experience.

As part of a field test, they conducted four longitudinal studies (systems use mandatory in two studies and voluntary in the other two), each with four measurement times (immediately after familiarization with a new information system and one, three, and five months after its implementation). Both hypotheses were confirmed: The subjective norm is moderated by two other variables, the variable *voluntariness* (voluntary use) and the variable *experience*. *Their results showed that the subjective norm only exerts a significant influence with mandatory use, but that this influence decreases with increasing use experience* (Venkatesh & Davis, 2000). Many studies later Olbrecht (2010) and Lai (2017) in their research reviews independently concluded that the variable subjective norm is not a robust construct of technology acceptance because it is highly situation-, time- and context-dependent (not surprising since it is a central property of the variable!). Some years later, both, the variable *subjective norm* and the variable *voluntariness* became core variables in TAM3 too.

5. Extensions of TAM by including the variables *Facilitating Conditions* and *Self-Efficacy*

Another TAM2 variable characterized by a strong situation, time, and context dependency (also unsurprising since it is a central property of this variable too) is that of *facilitating conditions*. Venkatesh et al. (2008, p. 485) define this variable as "individual perceptions of the availability of technological and/or organizational resources (...) that can remove barriers to using a system". Based on their research results, the authors admit that "in the presence of incomplete information and/or uncertainty regarding a behavior, facilitating conditions may not be a good predictor of the behavior" (Venkatesh, Brown, Maruping, & Bala, 2008, p. 485).

When designing the variable *facilitating conditions*, for the first time conceptual irregularities in TAM research appeared: Davis, Venkatesh and their team were not consistent in characterizing the content of this variable: the 3-4 items of the scale deal with knowledge, technical resources, compatibility of systems (Venkatesh, 2000). Occasionally, items for measuring *facilitating conditions* – also called *perceptions of external control* in the new TAM models – are found with the same wording in the scales for measuring *perceived behavioral control* (e.g., Venkatesh, Morris, & Ackerman, 2005); often, both scales are similar in terms of content. The starting point for this confusion was the research published by Taylor and Todd in 1995 (Taylor & Todd, 1995a, 1995b).

The *Decomposed Theory of Planned Behavior* presented by Taylor and Todd, which consisted of TAM variables and new concepts, had a *"(perceived) control structure"* under which these authors grouped four variables: a variable borrowed from Albert Bandura's Social Cognitive Theory

(Bandura, 1977; 2001) called *self-efficacy* (in the sense of *computer self-efficacy*), the variable *facilitating conditions/ technology* (lack of compatibility of the equipment), the variable *facilitating conditions/resources* (financial resources; equipment), and the TPB variable *perceived behavioral control*. Taylor and Todd (1995a; 1995b) changed the items for these three dimensions several times and made the scales sometimes shorter, sometimes longer, in order to be able test different theories simultaneously and comparatively. Due to the impressive explanation of variance of R^2 = 0.60 for the prediction of the (missing!) behavioral intention to use made possible by these additions in Taylor and Todd's model, Davis and Venkatesh decided to adopt these concepts and reclassified them. *But applying one theory to the other in empirical projects had its pitfalls and resulted in inaccuracies and modifications*.

The original *facilitating conditions concept* Davis and Venkatesh borrowed from Thomson, Higgins, & Howell (1991) contained items with a focus on guidance, assistance, specialized instruction by specific persons/experts/help pages in case of software and hardware problems. This is how it was later used by educational researchers such as Teo (Teo, 2011; Teo, Lee, & Chai, 2008) and Schorr (Schorr, 2020; Schorr & Gorovoj, under review). Venkatesh et al. (2008) define *facilitating conditions* broadly as "individual perceptions of the availability of technological and/or organizational resources (i.e., knowledge, resources, and opportunities) that can remove barriers to using a system" (Venkatesh, Brown, Maruping, & Bala, 2008, p.485). The original concept (instructional support face-to-face or digital) in some studies is recorded with one item only (e.g., Venkatesh, Brown, Maruping, & Bala, 2008). In their 2003 study, they point out that the influence of *facilitating conditions* on *user behavior* depends on the person's age and experience, i.e. the older and more experienced the person is, the stronger the influence (Venkatesh, Morris, Davis & Davis, 2003).

Banduras' concept of *self-efficacy* can be found in both TAM3 and UTAUT. Applied to technology acceptance research, Venkatesh et al. (2003) defined *(computer) self-efficacy* as "judgement of one's ability to use a technology (e.g., a computer) to accomplish a particular job or task" (Venkatesh, Morris, Davis, & Davis, 2003, p.432).

Venkatesh and Davis (1996) were able to prove that it makes sense and is also practically useful to deal with *computer self-efficacy* as one of the theoretical determinants of the TAM key beliefs. They realized that "over time, users built their computer self-efficacy beliefs based on the use of search systems" (Venkatesh et al., 2003). According to the authors, computer self-efficacy beliefs, both positive and negative, and particularly among knowledge workers have "a continuing significant impact on their perceptions of ease of use about any computer system" (Venkatesh & Davis, 1996, p. 472). *From their point of view, computer self-efficacy is a user-specific and system-independent characteristic that needs to be built up in training courses in order to generate increased acceptance of system/digital technologies* (see also Soror & Davis, 2014).

Many TAM and UTAUT researchers use the variable *computer self-efficacy* when exploring the determinants of the key beliefs (*perceived ease of use*, *perceived usefulness*). Unfortunately, Ajzen's statement that the TPB variable *perceived behavioral control* and the concept of *self-efficacy* are "very similar in content" (Ajzen, 2020, p. 316) fostered confusion and inaccuracy. Ajzens' (1991, p. 188) definition of the concept of *behavioral control* as "the perceived ease or difficulty of performing the behavior" is redefined much broader in the context of IS research by Venkatesth et al. (2003, p. 429) as "perceptions of internal and external constraints on behavior". In the context of the first presentation of the new Unified Theory of Acceptance and Use of Technology (UTAUT) by Venkatesh, Morris, Davis and Davis (2003), the newly included TPB concept *perceived behavioral control* (Ajzen 1991) corresponds in content (due to new items) the concept of (computer) self-efficacy, while the *concept of facilitating conditions* takes on the former guidance, assistance, specialized instruction focus (e.g., Teo, 2011). In the context of the new theory, Venkatesh et al. (2003) initially summarized both concepts and the *system compatibility issues* (facilitating conditions/technology) under the overarching label "facilitating conditions".

6. Theorizing in fast forward: TAM3 and UTAUT

The fact that the two technology acceptance models TAM3 and UTAUT are not positioned in competition with each other is not surprising given the timing of the publications. It is striking that the *Unified Theory of Acceptance and Use of Technology (UTAUT)*, first presented in 2003, can be attributed to Venkatesh, and that Davis was involved in this publication as a co-author (Venkatesh, Morris, Davis, & Davis, 2003). Five years later, without further explanation, Venkatesh introduces a new version of the TAM (TAM3) originally developed by Davis (Venkatesh & Bala, 2008). Both theories can be attributed to the major players, Davis and Venkatesh. Both models produce results.

The UTAUT, based on new variables and new data, achieved a high level of explained variance in its first application (adj. R^2 = 0.70; see Venkatesh, Morris, Davis, & Davis, 2003). It draws from eight different theory models, including TRA, TPB, TAM and Rogers' *Diffusion of Innovations Theory* (see Table 1). In a field study, Venkatesh and Bala used all eight models to test them under the same conditions and found that four variables directly influenced *behavioral intention to use*

or *user behavior*: *performance expectancy*, *effort expectancy*, *social influence* and *facilitating conditions*. Additional variables, such as *experience*, *voluntariness*, *gender* and *age*, were also found to be important and proved that access to resources and assistance are especially important for older users. They could also prove that users discover ways to find support to facilitate the use of the system as a result of increasing *experience* (Venkatesh, Morris, Davis, & Davis, 2003; Venkatesh, Brown, Maruping, & Bala, 2008).

7. Demands for more parsimony in theory development

Both, TAM3 and UTAUT enjoy great popularity and are equally controversial (cf. Bagozzi, 2007; Lai, 2017; van Raaij & Schepers, 2008, Williams et al. 2011, Williams et al. 2015, Tamilmani et al. 2017). Despite reports of success accompanied by high publication activity, critics accuse the authors of the leading technology acceptance models TAM3 and UTAUT of not having achieved any significant advantages with the new models compared to the original Technology Acceptance Model (Agudo-Peregrina et al., 2014; Marangunić & Granić, 2015; Tamilmani et al., 2017; Park et al., 2022). They take the position that the original, sparse technology acceptance model is much better suited for practical use and is even preferable to other complex alternatives such as the Theory of Reasoned Action and the Theory of Planned Behavior (Bagozzi, 2007; Chuttur, 2009).

The Technology Acceptance Model (TAM 3) presented by Venkatesh and Bala (2008) originally was developed to determine new influencing variables on the key variable of *perceived ease of use*. However, as Agudo-Peregrina et al. (2014) state "the increased complexity of TAM3 does not result in a significant improvement in the explanation of the acceptance and use process when compared to prior and more simple TAM-based models" (Agudo-Peregrina et al., 2014, p. 313).

As for the UTAUT, Bagozzi (2007) criticizes the high degree of complexity of the model due to a large number of variables, so that "in the end we are left with a model with 41 independent variables for predicting intentions and at least eight independent variables for predicting behavior" (Bagozzi, 2007, p. 245). Raaij and Schepers (2008) question the high variance explained: "UTAUT's high R^2 is only achieved when moderating the key relationships with up to four variables (gender, age, experience and voluntariness) in order to yield more significant coefficients" (van Raaij & Schepers, p. 840). In 2011, Williams et al. first time analyzed the hype surrounding the UTAUT. In their systematic review, they come to the conclusion that although the UTAUT is cited disproportionately often, it is used much less frequently. Out of 450 publications, only 43 used the UTAUT or UTAUT elements (Williams, Rana, Dwivedi & Lal, 2011).

In retrospect, the "fathers" of these theories agree with the critics: from the outset, Fred Davis preferred a sparse theory based on a few dimensions, explicitly referring to Fishbein and Ajzen in this matter (Davis, 1986, 1989; Davis, 2015). In 2020, Izhak Ajzen warns against adding more predictors to his theory (TPB) on the ground that this usually explains little more variance than with the basic variables. Looking back on his research on the Theory of Planned Behavior, which became so important for the TAM, he demands that "for the sake of parsimony, additional predictors should be proposed and included in the TPB with caution, and only after careful deliberation and empirical exploration". Compared to the TPB, Ajzen rated Davis' Technology Acceptance Model as a "content-specific model" that is thematically geared towards "the acceptance of computer-related technology in the workplace" (Ajzen, 2020, p. 317).

Venkatesh, Thong and Xu (2012) revised the UTAUT and presented the UTAUT2 in order to extend the theory from an information systems theory to the consumer context (cf. Chang, 2012; Venkatesh et al., 2012). The concept of *voluntariness* became obsolete and was dropped (since consumer research is based on voluntary participation?), and new concepts such as *hedonic motivation*, *price value*, and *habit* were added. The complexity of the model increased, its practicality decreased. In 2016, Venkatesh, Thong and Xu therefore presented a 48-page, highly differentiated review of the results of the UTAUT research. Many tables and many careful analyses lead them to the conclusion: "The main effects in UTAUT/UTAUT2 should serve as the baseline model of future research for parsimony and refining current context effects and/or identifying new context effects." (Venkatesh, Thong, & Xu (2016, p. 346). Using different models to evaluate theories, they finally admit that the UTAUT has a "relatively low parsimony". To streamline UTAUT for future research, Venkatesh recommends to omit the moderation effects of *age*, *gender*, *experience*, and *voluntariness* from the baseline model (Venkatesh, Thong, & Xu, 2016) – a recommendation that raises new problems for user research in business, media, education, and health.

8. The TAM – internationally reviewed

Over the years, many efforts were made to systematize and evaluate the abundance of research results based on TAM/UTAUT and to name future research tasks. Some of these reviews are briefly outlined below, sorted by publication date.

A typical review from the early years was published by Sharp (2007), who evaluated the research results of the TAM for the area of information systems education. Sharp optimistically praised the new approach based on complex

research designs and applied statistics. He identified its flexibility and applicableness on the basis of "numerous direct determinants and external variables that have been added to the model and the various technologies to which it has been applied" as the models' greatest strength (Sharp, 2007, p. 10).

For their literature review in 2015, Marangunić and Granić (2015) selected 85 TAM publications from the years 1986 to 2013 as well as 7 existing, extensive TAM literature reviews. They give an overview of the development of the TAM and also report on the fast growing number of "extended TAMs". Finally they evaluate consistent and inconsistent research results and supplement the knowledge gained by structuring notes for further research. In 2019, Granić and Marangunić published a second review focusing on TAM publications in the context of education. They rate the TAM as a "leading scientific paradigm and credible model for facilitating assessment of diverse technological deployments in educational context". But they also complain that "the state of current research on technology acceptance model application in educational context lacks comprehensive reviews addressing variety of learning domains, learning technologies and types of users" (Granić & Marangunić, p. 2572/3).

Tang and Hsiao (2016) focused their review on the publication development on TAM. They identified 4,571 TAM related articles with over 9,000 researchers involved. Based on the journals they tabulated, they stated that TAM research is important both in business and education research. The initial question of whether there is an overuse or misuse of TAM is left unanswered in their review (Tang & Hsiao, 2016). In 2018, Rahimi et al. (2018) published a systematic review of the technology acceptance model in the field of health informatics. They identified 134 articles from the years 1999-2017 that addressed clinic staff and patients' technology acceptance. The majority of the work was based on TAM model extensions. The authors reported progress with the proviso that there are still areas that can be improved to increase the predictive performance of TAM (Rahimi, Nadri, Afshar, & Timpka, 2018).

Overall, these and many other reviews are basically positive. What is striking is that in view of the abundance of publications on new "extended TAMs", it is becoming increasingly difficult for reviewers to go beyond some superficial classifications (e.g., journals that publish TAM work, number of researchers working on the model, etc.).

9. Proposal for a strategy change

Standardization of instruments as a first step

A problem in TAM research is that, as a central indicator for the relevance of factors (variables) for the acceptance behavior, these are consistently checked for their influence on the three key variables *perceived ease of use*, *perceived usefulness* and *intention to use*. Statistical correlations and elaborate, increasingly complex structural equation models serve as guide for theoretical models whose further use often remains unclear. This rigid strategy precluded other approaches. Careful development and testing of the survey instruments was largely dispensed with. In terms of subject matter and content, repeatedly new variables were found, tested and subsequently determined to be relevant, stimulated by the active comparison with a wealth of theories. Variables with overlapping content caused confusion, the original instruments for recording these variables were modified and continued to be used without checking.

The theoretical backgrounds of concepts borrowed from various theories and the research on these concepts are not thoroughly investigated. Instead of striving to further expand the Technology Acceptance Model, i.e. by expanding the theory with new variables based on correlative data, the well-confirmed key variables that characterize digital technology acceptance should be summarized in a multidimensional scale measuring the technology acceptance variable to ensure a solid measurement. *The aim should no longer be to expand the theory, but to research the psychological prerequisites of digital technology acceptance in depth on the basis of the model and with the help of a solid instrument (criterion) based on it*. The same applies to scales that depict important dimensions of content from other fields of research; they should be developed and checked just as carefully.

This could be a new way of dealing with the topic of digital technology acceptance based on Davis' technology acceptance model. The measurement tools could be a starting point. There are too many different scales applied to those variables identified as meaningful. Since the structure of the content of the individual items should be relatively simple in order to enable a broad spectrum of technology (software and hardware) to be evaluated (i.e. by simply filling in the name of the respective software/hardware under test), it is important to clarify the theoretical background of the measuring instruments in detail in order to validate the scales and then use them consistently across research teams and user groups so that the results are comparable.

Conceptually, the TAM factors recognized as key variables – as handled in a previous study (Gorovoj & Schorr, 2020; Schorr, 2020) – could be summarized in a multifactorial, multidimensional scale for measuring (digital) technology acceptance, which in its entirety represents digital technology acceptance as a criterion. For detailed analysis, the sub-dimensions should also be reliably measurable. There are numerous examples of this approach by concepts close to technology acceptance (cf. Table 2). These and other scales

Table 2. Scales measuring technology acceptance and related concepts[1]

Name of the Scale	Author(s)	Content / Dimensions measured
Resistance to Change Index	Davis, K. A. & Songer, A. D. (2008)	Measuring the likelihood of an individual to accept or reject information technology change. 7 dimensions: (1) Attitudes towards computers and technology; (2) Motivation to use new technology; (3) Readiness for change; (4) Irrational ideas; (5) Defense mechanisms related to behavior of individual during change; (6) Perceived interpersonal power; and (7) Perceived support for change
Technology Threat Avoidance Scale	Liang & Xue (2010); Young et al. (2016); Samhan (2017); Carpenter et al. (2019); Choi et al. (2022)	Measuring technology threat avoidance. 8 dimensions: (1) Perceived Susceptibility; (2) Perceived Severity; (3) Perceived Threat; (4) Perceived Effectiveness; (5) Perceived Cost; (6) Self-efficacy; (7) Avoidance Motivation; (8) Avoidance Behavior
Technostress Scale	Ragu-Nathan et al. (2008)	Stress created by ICT use. Five dimensions: Overload, Invasion, Complexity, Privacy, Inclusion
Generalized Digitalisation Anxiety Scale (DAS)	Pfaffinger et al. (2021)	Measuring generalized digitalisation anxiety (DA). 4 dimensions: (1) Generalized DA; (2) Self-related DA; (3) Interaction and Leadership-related DA; (4) Implementation-related DA
Task-Technology Fit Instrument	Goodhue & Thompson (1995); Goodhue (1998)	Measuring task-technology fit. 12 dimensions: (1) Right Data, (2) Right Level of Detail; (3) Accuracy; (4) Compatibility; (5) Locatability; (6) Accessability; (7) Flexibility; (8) Meaning; (9) Assistance; (10) Ease of Use of Hardware & Software, (11) Systems Reliability; (12) Overall
Technology Readiness Index: TRI 2.0 (for the first version, see Parasuraman, 2000)	Parasuraman, A., & Colby, C. L. (2015), (1st version, 2000)	Measuring technology readiness. 4 dimensions: (1) Optimism; (2) Innovativeness; (3) Discomfort; (4) Insecurity
Digital Technology Acceptance Scale (DTAS)	Schorr, A. (2020), Gorovoj, A. & Schorr, A. (2020), based on Davis et al. (1989)	Measuring digital technology acceptance. 4 dimensions: (1) Ease of Use, (2) Perceived Usefulness, (3) Attitude towards Usage, (4) Behavioral Intention to Use

[1]There are two other, thematically related German-language scales which, in contrast to the scales in Table 2, measure affinity or readiness for technology as a personality trait: (1) The scale by Karrer, Glaser, Clemens, & Bruder (2009) on technology affinity TA-EG, which captures the four dimensions of enthusiasm for technology, competence in dealing with technology, positive consequences of technology, and negative consequences of technology. (2) The scale by Neyer, Felber, & Gebhardt (2012), which measures technology readiness with the three dimensions of technology acceptance, technology competence, and technology control beliefs (perceived technology control).

could also be used for the necessary validation of the new standardized instrument.

Bringing together experts from different disciplines as a next step

Bringing experts from different disciplines into this research field can be a next step. The majority of TAM researchers received interdisciplinary training throughout their academic education. They are business engineers, business informatics specialists, educators specialized in computer science, media IT specialists and health IT specialists. This diversity was of great advantage for the entire field of research, because in the development of new conceptual approaches there were no reservations to other disciplines. At the same time, it was important to clearly anchor these approaches in information science.

But in order to achieve new insights, it is necessary at a certain point – that has been reached now! – *to actually implement interdisciplinarity, i.e. to consult experts from those disciplines on which TAM research is based*. Davis correctly refers to the TAM as a "motivational model" (Davis, 1986; 2015). Many TAM concepts are borrowed from psychology and the acceptance of digital technologies is basically a psychological issue. Only one (co)author of the major players in current research on technology acceptance is a psychologist (Nikola Marangunić, University of Split; specialized on human-computer interaction). Communication scientists, educators and health researchers are also hardly represented in TAM research. Involving these experts more in technology acceptance research makes sense as they can assess the technical and theoretical background of concepts borrowed from various sciences. Furthermore, they have specialized research designs and appropriate technical procedures and tools that might be helpful for future research. Even if it seems like a contradiction: The problem of TAM research is not its interdisciplinary nature. That is why the formation of *interdisciplinary teams* could generate progress and solve some of the problems mentioned in this review.

The Technology Acceptance Model is an exciting project for psychology. Implementation phases of new digital technologies occur repeatedly at work and in people's daily life and trigger learning processes. The topic of *learning in adulthood* becomes increasingly important, and research on this topic and as well as on the topic of acceptance research needs to be reoriented. With regard to the design of training for defined user groups, a lively exchange of research findings with TAM researchers from information science is recommended. To study users' reactions to new digital technologies is not the only contribution of interest to science and industry that psychologists can make. The psychologists Richard Landers and Sebastian Marin (2021) carefully analyzed the interaction

between users and technological progress. Researchers in psychology, they appeal, should now take the next step „to explicitly model technology design" (Landers & Marin, 2021, p. 235).

10. Conclusions

Based on serious research and scientifically sound consulting, the concepts and tools produced by the TAM research teams have helped to successfully design the implementation phases of digital technologies for people in a wide range of industry sectors (automotive industry, banking industry, telecommunications industry, in public administrations, in the healthcare system and many more). Early on, the TAM researchers recognized the great opportunities that comprehensive digitization offers in all areas of society and reacted to them. Davis and his colleagues knew that with complex digital products, simple customization is not enough. Many successful careers in science and practice have been and continue to be founded within the framework of technology acceptance research. This research is characterized by sophisticated research designs and great openness to ideas and concepts from other disciplines.

This review is intended to give an impetus to standardize research on digital technology acceptance in relation to research instruments in order to open up the option of merging the ever-increasing amount of research results more easily. In addition, interdisciplinary teams, especially from information science and psychology, should be formed to advance the research field at the current state of these disciplines.

References

Agudo-Peregrina, Á. F., Hernández-García, Á., & Pascual-Miguel, F. J., (2014). Behavioral intention, use behavior and the acceptance of electronic learning systems: Differences between higher education and lifelong learning. *Computers in Human Behavior*, *34*, 301-314.

Ajzen, I., & Fishbein, M., (1980). *Understanding attitudes and predicting behavior*. Englewood Cliffs, N. J.: Prentice-Hall.

Ajzen, I., (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.) *Action control. From Cognition to Behavior* (pp. 11-39). Springer Verlag: Berlin, Heidelberg.

Ajzen, I., (1991).The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*. 50(2), 179–211.

Ajzen, I., (2020). The theory of planned behavior: Frequently asked questions. *Human Behavior and Emerging Technologies, 2*(4), 314-324.

Arthur, W. B., (2009). *The nature of technology: What it is and how it evolves*. New York, N.Y.: Simon & Schuster.

Bagozzi, R. P., (2007). The legacy of the technology acceptance model and a proposal for a paradigm shift. *Journal of the Association for Information Systems*, *8*(4), 244-254.

Bandura, A., (1977). Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review*. 84(2),191–215.

Bandura, A., (1997). *Self-efficacy: The Exercise of Control*. New York: Macmillan.

Bandura, A., (2001). Social cognitive theory: An agentic perspective. In S. T. Fiske (Ed.), *Annual Review of Psychology, 52*, pp. 1-26. Palo Alto: Annual Reviews, Inc.

Blut, M. & Wang, C., (2020). Technology readiness: a meta-analysis of conceptualizations of the construct and its impact on technology usage. *Journal of the Academy of Marketing Science, 48*, 649-669.

Carpenter, D., Young, D. K., Barrett, P., & McLeod, A. J., (2019). Refining technology threat avoidance theory. *Communications of the Association for Information Systems*, *44*.

Chang, A. (2012). UTAUT and UTAUT 2: A review and agenda for future research. *The Winners*, *13*(2), 10-14.

Choi, H. S., Carpenter, D., & Ko, M. S., (2022). Risk taking behaviors using public Wi-fi™. *Information Systems Frontiers*, *24*(3), 965-982.

Chuttur, M., (2009). Overview of the technology acceptance model: Origins, developments and future directions. *Sprout, 9*, 9-37. *All Sprouts Content*. 290; https://aisel.aisnet.org/sprouts_all/290

Davis, F. D., (1986). *A Technology Acceptance Model for Empirically Testing New End-User Information Systems. Theory and results*. Dissertation. Sloan School of Management, Massachusetts Institute of Technology.

Davis, F. D., (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319-340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R., (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, *35*(8), 982-1003.

Davis, F. D., & Venkatesh, V., (1996). A critical assessment of potential measurement biases in the technology acceptance model: three experiments. *International Journal of Human-Computer Studies*, *45*(1), 19-45.

Davis, K. A., & Songer, A. D., (2008). Resistance to IT change in the AEC industry: an individual assessment tool. ITcon Vol. 13 (2008), Davis & Songer, pg. 56-68. vtechworks.lib.vt.edu

Davis, F. D., (2015). On the relationship between HCI and technology acceptance research. In *Human-computer interaction and management information systems: Foundations* (pp. 409-415). Routledge.

Dwivedi, Y. K., Rana, N. P., Jeyaraj, A., Clement, M., & Williams, M. D., (2019). Re-examining the unified theory of acceptance and use of technology (UTAUT): Towards a revised theoretical model. *Information Systems Frontiers*, *21*(3), 719-734.

Eason, K. D., (1984). Towards the experimental study of usability. *Behavior & Information Technology*, *3*(2), 133-143.

Feng, G. C., Su, X., Lin, Z., He, Y., Luo, N., & Zhang, Y., (2021). Determinants of technology acceptance: Two model-based meta-analytic reviews. *Journalism & Mass Communication Quarterly*, *98*(1), 83-104.

Fishbein, M., & Ajzen, I., (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research.* Reading, MA: Addison-Wesley.

Fussell, S. G. & Truong, D., (2022). Using virtual reality for dynamic learning: An extended technology acceptance model. *Virtual Reality, 26*, 249-267.

Goodhue, D.L., & Thompson, R.L., (1995). Task-technology fit and individual performance. *MIS Quarterly 19*(2), 213-236.

Goodhue, D. L., (1998). Development and measurement validity of a task-technology fit instrument for user evaluations of information system. *Decision Sciences*, *29*(1), 105-138.

Gorovoj, A., & Schorr, A., (2020). Zur wahren Bedeutung von Einstellungs- und Persönlichkeitsfaktoren für die Akzeptanz Digitaler Medien (On the role of attitude and personality factors for the acceptance of digital media). Gesellschaft für Arbeitswissenschaft (Ed.). *Digitaler Wandel, Digitale Arbeit, Digitaler Mensch?* Dokumentation des 66. Arbeitswissenschaftlichen Kongresses, No.35, B.20.3,1-6.

Granić, A., & Marangunić, N., (2019). Technology acceptance model in educational context: A systematic literature review. *British Journal of Educational Technology*, *50*(5), 2572-2593.

Horton, R. P., Buck, T., Waterson, P. E., & Clegg, C. W., (2001). Explaining intranet use with the technology acceptance model. *Journal of Information Technology*, *16*(4), 237-249.

Karrer, K., Glaser, C., Clemens, C., & Bruder, C., (2009). Technikaffinität erfassen – der Fragebogen TA-EG. *Der Mensch im Mittelpunkt technischer Systeme*, *8*, 196-201.

Lai, P. C., (2017). The literature review of technology adoption models and theories for the novelty technology. *JISTEM-Journal of Information Systems and Technology Management*, *14*, 21-38.

Landers, R. N., & Marin, S., (2021). Theory and technology in organizational psychology: A review of technology integration paradigms and their effects on the validity of theory. *Annual Review of Organizational Psychology and Organizational Behavior, 8*, 235-258.

Liang, H., & Xue, Y. L., (2010). Understanding security behaviors in personal computer usage: A threat avoidance perspective. *Journal of the Association for Information Systems*, *11*(7), 394-413.

Marangunić, N., & Granić, A., (2015). Technology acceptance model: A literature review from 1986 to 2013. *Universal Access in the Information Society*, *14*(1), 81-95.

Mothukuri, U. K., Reddy, B. V., Reddy, P. N., Gutti, S., Mandula, K., Parupalli, R., & Magesh, E. (2017, August). Improvisation of learning experience using learning analytics in eLearning. In *2017 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH)* (pp. 1-6). IEEE.

Neyer, F. J., Felber, J., & Gebhardt, C., (2012). Entwicklung und Validierung einer Kurzskala zur Erfassung von Technikbereitschaft. *Diagnostica*, *58*(2), 87-99

Olbrecht, T., (2010). *Akzeptanz von E-Learning: Eine Auseinandersetzung mit dem Technologieakzeptanzmodell zur Analyse individueller und sozialer Einflussfaktoren* (Doctoral Dissertation, University of Jena, Germany).

Park, I., Kim, D., Moon, J., Kim, S., Kang, Y., & Bae, S., (2022). Searching for New Technology Acceptance Model under Social Context: Analyzing the Determinants of Acceptance of Intelligent Information Technology in Digital Transformation and Implications for the Requisites of Digital Sustainability. *Sustainability*, *14*(1), 579, 1-29.

Parasuraman, A., (2000). Technology Readiness Index (TRI) a multiple-item scale to measure readiness to embrace new technologies. *Journal of Service Research*, *2*(4), 307-320.

Parasuraman, A., & Colby, C. L. (2015). An updated and streamlined technology readiness index: TRI 2.0. *Journal of Service Research*, *18*(1), 59-74.

Pfaffinger, K. F., Reif, J. A., Spieß, E., & Berger, R. (2020). Anxiety in a digitalised work environment. *Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie (GIO)*, *51*(1), 25-35.

Pfaffinger, K. F., Reif, J. A., Huber, A. K., Eger, V. M., Dengler, M. K., Czakert, J. P., ... & Berger, R. (2021). Digitalisation anxiety: development and validation of a new scale. *Discover Mental Health*, *1*(1), 1-14.

Rahimi, B., Nadri, H., Afshar, H. L., & Timpka, T., (2018). A systematic review of the technology acceptance model in health informatics. *Applied Clinical Informatics*, *9*(03), 604-634.

Ragu-Nathan, T. S., Tarafdar, M., Ragu-Nathan, B. S., & Tu, Q., (2008). The consequences of technostress for end users in organizations: Conceptual development and empirical validation. *Information Systems Research*, *19*(4), 417-433.

Rogers, E. M. (2003). *Diffusion of innovations* (Social Science, 5th Edition, Free Press trade paperback edition). New York: Free Press. Retrieved from: http://www.loc.gov/catdir/bios/simon052/2003049022.html

Samhan, B. (2017, April). Security behaviors of healthcare providers using HIT outside of work: A technology threat avoidance perspective. In *2017 8th International Conference on Information and Communication Systems (ICICS)* (pp. 342-347). IEEE.

Schorr, A., (2020). Skala zur Erfassung der Digitalen Technologieakzeptanz – Weiterentwicklung zum testtheoretisch geprüften Instrument Gesellschaft für Arbeitswissenschaft (Ed.) *Digitaler Wandel, Digitale Arbeit, Digitaler Mensch?* Dokumentation des 66. Arbeitswissenschaftlichen Kongresses, No. 35, B.20.5, pp. 1-6.

Schorr, A. & Gorovoj, A. (submitted). Digital technology acceptance among adults of different age groups – a key factor for the success of the intergenerational contract. Journal of Adult Development.

Sharma, R., & Yetton, P., (2001). An evaluation of a major validity threat to the technology acceptance model. The 9th European Conference on Information Systems, Proceedings, pp.1170-1175.

Soror, A. & Davis, F., (2014). Using self-regulation theory to inform technology-based behavior change interventions. IEEE, 47the Hawaiian International Conference on System Science, 3004-3012. DOI 10.1109/HICSS.2014.373

Szajna, B., (1996). Empirical evaluation of the revised technology acceptance model. *Management science*, *42*(1), 85-92.

Tamilmani, K., Rana, N. P., & Dwivedi, Y. K., (2017). A systematic review of citations of UTAUT2 article and its usage trends. In *Conference on e-Business, e-Services and e-Society* (pp. 38-49). Springer, Cham.

Tang, K. Y., & Hsiao, C. H., (2016). The literature development of technology acceptance model. *International Journal of Conceptions on Management and Social Sciences*, *4*(1), 1-4.

Teo, T., Lee, C. B., & Chai, C. S., (2008). Understanding pre-service teachers' computer attitudes: applying and extending the technology acceptance model. *Journal of Computer Assisted Learning*, *24*(2), 128-143.

Teo, T., Ursavaş, Ö. F., & Bahçekapili, E. (2011). Efficiency of the technology acceptance model to explain pre-service teachers' intention to use technology: A Turkish study. *Campus-Wide Information Systems*.

Taylor, S. & Todd, P., (1995a). Decomposition and crossover effects in the theory of planned behavior: A study of consumer adoption intentions. *International Journal of Research in Marketing*, *12*(2), 137-155.

Taylor, S., & Todd, P. A., (1995b). Understanding information technology usage: A test of competing models. *Information Systems Research*, *6*(2), 144-176.

Thompson, R. L., Higgins, C. A., & Howell, J. M. (1991). Personal computing: Toward a conceptual model of utilization. *MIS Quarterly*, 125-143.

Turner, M., Kitchenham, B., Brereton, P., Charters, S., & Budgen, D. (2010). Does the technology acceptance model predict actual use? A systematic literature review. *Information and Software Technology*, *52*(5), 463-479.

Van Raaij, E. M., & Schepers, J. J., (2008). The acceptance and use of a virtual learning environment in China. *Computers & Education*, *50*(3), 838-852.

Venkatesh, V., & Davis, F. D., (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision Sciences*, *27*(3), 451-481.

Venkatesh, V., (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, *11*(4), 342-365.

Venkatesh, V., & Davis, F. D., (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, *46*(2), 186-204.

Venkatesh, V., Morris, M. G., & Ackerman, P. L., (2000). A longitudinal field investigation of gender differences in individual technology adoption decision-making processes. *Organizational Behavior and Human Decision Processes*, *83*(1), 33-60.

Venkatesh, V., Morris, M. G., Davis, G. B. & Davis, F. D., (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425–478.

Venkatesh, V., & Bala, H., (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences*, *39*(2), 273-315.

Venkatesh, V., Davis, F., & Morris, M. G. (2007). Dead or alive? The development, trajectory and future of technology adoption research. *Journal of the association for information systems*, *8*(4), 267-286.

Venkatesh, V., Brown, S. A., Maruping, L. M., & Bala, H., (2008). Predicting different conceptualizations of system use: The competing roles of behavioral intention, facilitating conditions, and behavioral expectation. *Management Information Systems Quarterly*,483-502.

Venkatesh, V., Thong, J. Y., & Xu, X., (2012). Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *Management Information Systems Quarterly, 36*(1), 157–178.

Venkatesh, V., Thong, J. Y., & Xu, X., (2016). Unified theory of acceptance and use of technology: A synthesis and the road ahead. *Journal of the Association for Information Systems*, *17*(5), 328-376.

Williams, M., Rana, N., Dwivedi, Y., & Lal, B., (2011). Is UTAUT really used or just cited for the sake of it? A systematic review of citations of UTAUT's originating article. aisel.aisnet.org

Williams, M. D., Rana, N. P., & Dwivedi, Y. K., (2015). The unified theory of acceptance and use of technology (UTAUT): a literature review. *Journal of Enterprise Information Management, 28*(3),443-488.

Young, D. K., Carpenter, D., & McLeod, A., (2016). Malware avoidance motivations and behaviors: A technology threat avoidance replication. *AIS Transactions on Replication Research*, *2*(1), 8, 1-17.

DOI: 10.2478/9788366675896-006 • Technikpsychologie • 66-81

Proceedings of the International Symposium on Technikpsychologie

Dual-use in volunteer operations? Attitudes of computer science students regarding the establishment of a cyber security volunteer force

Jasmin Haunschild†, Leon Jung, Christian Reuter

Technical University of Darmstadt Science and Technology for Peace and Security (PEASEC) Pankratiusstraße 264289, Darmstadt

Abstract

The digitalisation of critical infrastructure has increased the risk of large-scale cyber incidents. In contrast to the management of conventional emergencies by established civil protection organisations involving volunteers in Germany, few response capacities exist for these events. The concept of a volunteer force for cyber security could close this protection gap. However, such involvement also poses practical and ethical challenges. By conducting interviews with computer science students (N = 11), this paper analyses potential volunteers' attitudes towards ethical implications of a cyber volunteer force, as well as practical aspects that might motivate or hinder their participation. A qualitative content analysis reveals that students are largely unaware of potential dilemmas connected to vulnerabilities handling and national cybersecurity interests. Ethical guidelines and means of motivating and encouraging potential volunteers are discussed.

Keywords

volunteers • cybersecurity • attitude • interview

1. Introduction

We rely on Critical Infrastructure (CI), so that state and society function properly. In the course of advancing digitalisation, CI is no longer exclusively physical systems, but "cyber physical systems" (Jazdi, 2014; Ludwig et al., 2021), as they are digitally networked. This change entails new risks: While previously, CI operations could only be disrupted on a large scale by physical disasters (such as natural catastrophes), nowadays, digital threats, e.g., hacker attacks or malware infections, can also cause such large-scale disruptions. While organisations such as the German Federal Office for Information Security (BSI) advise and share information on security incidents to CI companies, there are hardly any operational capacities outside of individual companies to intervene in large-scale cyber incidents. At the same time, past events have shown that large-scale cyber incidents have been occurring. Due to the 2007 cyber attacks in Estonia, the Defence League Cyber Unit (CDL) was founded, which institutionalises the *ad hoc* expert network that was formed to counter the attacks (Cardash et al., 2013). A similar effort to include volunteers in cyber defence has also been promoted in Germany. The current government seeks to increase the involvement of volunteers in such activities as part of the larger volunteering organisation, the German Federal Agency for Technical Relief (Technisches Hilfswerk, THW). Yet, a separate organisation dedicated to cyber incidents is under discussion (AG KRITIS, 2020; Rundfeldt, 2020).

Due to the connection of cyber security response and IT vulnerabilities, ethical challenges arise regarding the dual-use potential of exploits. A cyber volunteer force's organisational structure might give states influence over the organisation and its processes with regard to identified vulnerabilities and exploits. Dual-use refers to research, technologies or artefacts which, according to the traditional definition, can serve civilian as well as military purposes (Code of Federal Regulations, 2013; European Commission, 2018), and, according to a broader definition, can have beneficial as well as harmful purposes (Forge, 2010; Lin, 2016; Miller, 2018). Exploits have a dual-use potential as they can be used for malicious or military purposes. The use of vulnerabilities by state actors is fiercely debated.

It is feared, e.g., that participation could indirectly support German law enforcement agencies – a security breach discovered during an operation could be withheld by a

†*Corresponding author*: *Jasmin Haunschild*
E-mail: haunschild@peasec.tu-darmstadt.de

directive of the superior authority (the Federal Ministry of the Interior), not be published and leaked to security agencies and intelligence services. The "responsible disclosure" process common in the IT security industry, i.e. "ethical hacking", would thus be circumvented (Schulze, 2019; Thomas et al., 2018). Yet, so far nothing is known about the IT professionals' attitude towards these ethical questions, about volunteering their time and expertise or the factors that would motivate or demotivate voluntary engagement. This contribution seeks to close this gap by answering the Research Question RQ: **"What are computer science students' attitudes on their participation in a cyber volunteer force?"**. We answer this by conducting and analysing interviews with computer science students (N = 11) about awareness of ethical challenges in IT, preferences regarding the organisational influence of security organisations and practical motivations and hindrances.

2. Literature Review

This literature review focuses on managing cyber incidents (section 2.1) and ethical issues related to cyber security (section 2.2). Section 2.3 identifies the research gap.

2.1 Disaster Management for Large-Scale Cyber Incidents

Protecting the population, warding off disasters and eliminating their consequences are tasks of disaster management (Wenzel et al., 2012). In Germany, disaster management is the mandate of the Federal Office of Civil Protection and Disaster Assistance (BBK) and is provided by the (volunteer) fire brigade, the state-mandated relief organisations (German Red Cross, German Life Saving Society, etc.), and the Federal Agency for Technical Relief (THW) (Terberl, 2015). They rely heavily on volunteers: More than 90 % of German civilian protection is carried out by volunteers (Haas, 2020). Current research in the field of disaster or crisis management focuses on how disaster management can be improved through digitalisation. Reuter and Kaufhold (2018) show that social media is used by citizens to self-organise during crises. In addition, volunteers are increasingly performing digital tasks in crisis response, such as mapping disaster areas (Fiedrich & Fathi, 2021). A study by Fathi et al. (2020) covers "Virtual Operations Support Teams" (VOSTs), analysing the first deployment of a German VOST by the THW in 2017. The VOST has since become an established part of the THW to obtain, process and present information from social networks to increase situational awareness in disaster situations (THW, 2022).

In contrast to conventional major emergencies, e.g. triggered by natural disasters, major IT incidents are a relatively recent phenomenon, that have been occurring more often. Not only are cyber attacks on companies increasing each year (Verizon, 2021), but cyber attacks on cities and their CI are also increasing in frequency and in the extent of the damage caused (Gedris et al., 2021). In a German study of 99 organisations, 50.5 % of participating CI cited having been the target of at least one cyber attack in 2017 (Lechner, 2018). Yet, the same disaster response capacities are not available for digital crises. In Germany, the BSI is the central cyber security authority and responsible for the IT security of CI (Dürig & Fischer, 2018; Schallbruch, 2017). However, it is mainly dedicated to knowledge sharing and supporting agencies and CI with regard to cyber incidents. Concerning response, the BSI has very limited resources within its "Mobile Incident Response Team" (BSI, 2022b; Herpig & Rupp, 2021). Similarly, Cyber Emergency Response Teams (CERTs), which have been established on the national and federal level, lack the resources for large-scale incident response (Riebe et al., 2021). This is problematic, since some CI are particularly vulnerable to cyber attacks: (1) CI areas are increasingly digital (AG KRITIS, 2020). (2) At the same time, components are very durable, typically have a lifespan of several decades and are thus rarely updated by replacements (AG KRITIS, 2020; Jasiu¯nas et al., 2021). (3) Due to the partly locally remote components of the infrastructure, a connection to the internet for remote maintenance is economically necessary (Jasiu¯nas et al., 2021). (4) In some cases, monocultures of hardware and software prevail. Due to the use of the same components, vulnerabilities in a single component lead to a large number of facilities being affected by the same vulnerabilities (Jasiu¯nas et al., 2021). (5) At the same time, only those facilities that cater to a large number of citizens are counted as CI and are required to adhere to certain cyber security procedures in Germany.

Confronted with similarly limited response capacities, Estonia established a "Cyber Defence League" (CDL) to deal with large-scale cyber incidents. The CDL is an expert network of volunteers from the IT sector who can be deployed in the event of large-scale IT incidents (Cardash et al., 2013; Collier, 2017; Kaska et al., 2013). It is part of the Estonian Defence League, a military volunteer organisation, under the Estonian Ministry of Defence (Kaska et al., 2013). Apart from VOSTs, which offer a virtual component for physical attacks, emergency telephone numbers for digital first responders are being tried out in Germany, offering a first consultation for small and medium enterprises (BSI, 2022a; Cyberwehr-bw.de, 2022). While large-scale incidents are rare, past low-probability crises, such as the COVID-19 pandemic, show the feasibility of preparing for such exceptional cases. Due to the large number of facilities that could be targeted simultaneously, the expert network *AG KRITIS* warns of large-scale cyber security incidents and their effects for civil society and suggests establishing a volunteer force dedicated to and trained for large-scale cyber incidents (AG KRITIS, 2020).

However, a number of ethical and practical questions remain open.

2.2 Ethics and Cyber Security

Major ethical challenges in computer science arise from the inherent "dual-use" character of computer science. In a narrow sense, dual-use refers to research, technologies or artefacts which can have civilian as well as military uses or more broadly positive as well as harmful uses (Forge, 2010; Lin, 2016). Regarding IT security, an ethical field of tension exists between national or internal security and individual security and privacy (Dunn Cavelty, 2014; Wenger et al., 2017). "Cyberspace" is seen by some as the newest military domain (Leinhos, 2019). In cyber conflicts, security vulnerabilities are central, as they allow to infiltrate digital systems. Currently, a digital arms race between states is suspected, with arms control in cyberspace already being discussed (Meyer, 2020; Reinhold, 2020). In addition to offensive attacks, operations such as so-called "hack backs" or "active cyber defence" have been discussed. These strategies entail attacking an attacker to stop the ongoing attack (Schallbruch, 2020), similar to offensive strategies that aim to harm opponents or to keep them busy, and rely on the existence of currently unpatched or unknown vulnerabilities and ways of exploiting them ("zero-day exploits"). Vulnerabilities can also be relevant for domestic security and surveillance: To enable digital criminal prosecution, "Remote Forensic Software", also called "State Trojans", is used. It is secretly installed by law enforcement agencies on suspects' end devices by exploiting open security gaps to enable access to the device (Schallbruch, 2020). Thus, vulnerabilities are deliberately kept secret or withheld (FragDenStaat.de, 2018; Netzpolitik.org, 2021) or even acquired on the black market (Ablon et al., 2014; Schulze, 2019). Similar tools have been used to surveil politicians, journalists and activists (Michaelsen, 2020; Scott-Railton et al., 2022).

A dilemma arises between ensuring privacy and civilian security or essentially weakening cyber security in favour of domestic and national security (Schallbruch, 2020). At the heart of the dilemma is how the state deals with IT vulnerabilities (Herpig, 2018; Schulze, 2019). A potential solution might be pre-defined procedures. A process that was developed by the hacking community is called "responsible disclosure" and entails informing vendors of their vulnerabilities and giving a set time frame to develop and publish a patch. Otherwise, it is publicly disclosed. Yet, the Vulnerabilities Equities Process (VEP) sets national criteria to evaluate whether a security vulnerability should be withheld or published (Schulze, 2019). How decisions are reached is often non-transparent and many countries lack such a procedure (Herpig, 2018).

From the perspective of IT security experts, dealing with vulnerabilities is an ethical dilemma, as they are often only identified through "ethical hacking" (Ding et al., 2019; Jaquet-Chiffelle & Loi, 2020) – meaning hacking that seeks to make software safer, instead of hacking to exploit the vulnerabilities. Ethical hacking, from a corporate perspective, consists of means to ensure responsible detection and reporting of a company's IT security vulnerabilities, while reducing the risk of such vulnerabilities being sold on the black market (Ding et al., 2019; Schulze, 2019). The Netherlands was the first country to publish an official Responsible Disclosure Guideline in 2013 (National Cyber Security Centre, 2018). The IT security community has thus created a process that defines ethical behaviour for IT security professionals. In Germany, identifying and making vulnerabilities public can be prosecuted as a criminal offence, and researchers are currently left to their own devices on how to disclose identified vulnerabilities, with limited disclosure to software producers as the recommended solution (Balaban et al., 2021).

Some research investigates the computer science professions' awareness of and attitudes towards ethical issues. Schneider (2013) conducted a study on technological consequences with computer science students and showed that all students in the sample followed the same moral reflection pattern. When asked about ethical problems in connection with the programming of a weaponised robot, which was to be used in a rescue scenario, the dual-use character was perceived by the students, but not considered. Only the use of this technology seemed morally relevant – and ethical reflection only became necessary in a case with a subjectively bad intention. Schneider (2013) compared his results with the research of sociologist Max Weber (1919), who distinguished an *ethic of ultimate ends* and an *ethic of responsibility*. Ethicists of ultimate ends only examine whether the intentions of an action are irreproachable. "If the consequences of an action stemming from pure sentiment are evil, [the ethicists of ultimate ends] do not hold the person acting responsible for it, but the world" (Weber, 1919). In the ethics of responsibility, on the other hand, "responsibility is assumed by those who base their actions on the consequences, regardless of the means that may have to be used in the process" (Weber, 1919).

Research also suggests that ethics are not prioritised in curricula. Casañ et al. (2020) examined the curriculum of the degree programme "Informatics Engineering" over a period of 29 years. They found that each new iteration of the curriculum places an increased focus on technical specialisation, while ethics or social responsibility are almost non-existent. Polmear et al. (2018) conducted a study with lecturers of courses on "ethical and social impact" and identified challenges in teaching, e.g., lack of support for faculties. Athey (1993) conducted a study with computer science students in which students' ethical beliefs were compared to those of experts, showing that students' ethical judgements differed from those

of experts. The authors suggest that this may be due to the experts' greater experience as well as economic insecurity, as many students believed they could lose their jobs if they challenged unethical behaviour. Hashemian and Loui (2010) found that students who successfully completed an ethics course considered more courses of action, answered consistently and showed a stronger sense of responsibility. In contrast, students who scored poorly in the ethics course or who did not take the course had a weaker sense of responsibility. The authors argue that an ethics course can strengthen the awareness of responsibility, the knowledge of how to deal with a difficult situation and the confidence in one's own actions.

2.3 Research Gap

The literature review shows that many gaps remain regarding ethical volunteering in cyber emergencies. Looking at the way that disaster relief is organised in Germany reveals that it strongly depends on volunteer-based organisations. At lot of research in crisis informatics has looked at how emergency organisations and individuals confront these situations, but this is rarely extended to cyber emergencies (Riebe et al., 2021). While private companies are largely held responsible for ensuring their own cyber security, recent incidents have shown the effects of attacks targeting CI. While large-scale cyber attacks have not yet occurred in Germany, they are a scenario that experts warn about and that could not be addressed with the current public cyber response resources. While involving volunteers not only in cyber situational awareness but also in cyber response has been used in Estonia and has gained some momentum, it remains unclear what aspects in the German context might be important from volunteers' perspectives. Investigating this aspect needs to take the national mode of organising cyber emergency response into account, since these vary widely between states and have ethical implications (Boeke, 2018).

Dilemmas concerning the state and IT experts arise. The extent to which potential volunteers are aware of the ethical challenges and the German cybersecurity architecture has not been researched. Against this background, we explore these gaps by addressing the following research questions:

- **RQ1a:** To what extent are computer science students *aware* of ethics in computer science, particularly concerning the management of security vulnerabilities?
- **RQ1b:** What are *ethical concerns* regarding establishing a cyber security volunteer force among computer science students?
- **RQ2:** Which *attitudes* do students have concerning a cyber security volunteer force's co-operation with other state security agencies?
- **RQ3:** Which *practical incentives* and *practical concerns* do computer science students see concerning establishing a cyber security volunteer force?

3. Research Method

To answer the research questions, we conducted interviews with N = 11 students of computer science. In the following, we describe our data collection process through guided interviews and the data analysis process using content analysis.

3.1 Data Collection through Guided Interviews

For data collection, we use qualitative interviews, as they offer more nuanced insights and allow for unexpected aspects to emerge, which is especially useful when investigating a research field about which little is currently known (Helfferich, 2011).

Interviewees were recruited through convenience sampling among students of computer science. The students came from different universities in the same region and were recruited through posts on university websites. Participants were offered to participate in a raffle for two gift certificates, each worth C15. To register, participants had to agree to the terms of the study, which included audio recorded interviews and anonymisation of the data during transcription. Data were hosted locally or on servers hosted by a German university. In the process, we recruited six bachelor and five master students of computer science between the ages of 21 and 26 (see Table 1). Despite several rounds of recruitment, regrettably, we did not have a female participant. As around 20 % of students starting German computer science programmes are female (Statistisches Bundesamt, 2022), the sample only represents 80 % of the students. The interviews were conducted via the Zoom video conferencing platform between 15 October 2021

Table 1. Study participants. CS = Computer Science, BIS = Business Information Systems.

Name	Age	Gender	Study programme	University
P01	25	male	M.Sc. CS	TU Darmstadt
P02	19	male	B.Sc. CS	TU Darmstadt
P03	25	male	M.Sc. CS	TU Darmstadt
P04	22	male	B.Sc. CS	TU Darmstadt
P05	26	male	M.Sc. CS	Hochschule Darmstadt
P06	25	male	M.Sc. BIS	TU Darmstadt
P07	26	male	B.Sc. CS	TU Darmstadt
P08	23	male	B.Sc. CS	TU Darmstadt
P09	23	male	M.Sc. BIS & M.Sc. CS	TU Darmstadt
P10	23	male	B.Sc. CS	JGU Mainz
P11	21	male	B.Sc. CS	Goethe Uni Frankfurt

and 17 January 2022, each lasting 50–60 minutes. During the interviews, the audio track was recorded. Based on this, anonymised transcripts were created, after which the audio recordings were deleted.

We chose guided interviews as they offer the necessary guidance, while leaving room for study participants to focus on issues most relevant to them (Helfferich, 2019). When addressing ethical challenges, it is important to pose the questions in a way that is neutral and encourages participants to express their true opinions, without feeling the need to follow a perceived social norm. When designing the interview questions, we therefore aimed for a neutral wording throughout the questionnaire. At the same time, in order to discuss later issues, it was also important to give background information and make participants aware of certain ethical challenges. We thus started by asking general questions about ethics in IT and then progressed to questions more closely related to cyber volunteering and its particular ethical challenges. As the participants cannot have any experience of participating in a cyber volunteer force, we used scenarios to encourage participants to imagine such participation under different conditions.

The interview guide consists of six sections in total (see Table 2).

3.2 Qualitative Content Analysis

Qualitative content analysis is the most frequently applied text-analytical method (Mayring & Fenzl, 2019). The method's aim is to work out the "latent meaning" (Mayring & Fenzl, 2019) and the interpretation of the content. To carry out the analysis, a coding system is used, which summarises the respective aspects of analysis (Mayring & Fenzl, 2019). The coding system can be derived in different ways; e.g., inductively based on the material, or deductively based on the underlying theory (Mayring & Fenzl, 2019). A coding system consists of different thematic categories, which are further differentiated into different codes (Kuckartz & Rädiker, 2020).

In this study, categories were derived from the topics that structure the interview and codes emerged inductively from the answers given by the participants. The data were coded by one researcher, with a second researcher verifying the results to increase their reliability. Any conflicting interpretations were discussed and resolved through consensus.

4. Research Findings

In the following, we present the findings of the content analysis.

4.1 Awareness of Ethical Issues Related to IT and Computer Science (RQ1a)

Regarding awareness about ethical aspects of computer science in general, it can first be stated that a large part of participants (n = 7) had rarely come into contact with ethics in computer science during their studies. Ethical issues that people were aware of primarily related to bias and explainability in artificial intelligence (n = 5). Fewer participants also mentioned the topic of 'surveillance technology' (n = 2), as well as the topics 'sustainability and energy consumption' (n = 1) and 'broken encryption through quantum computing' (n = 1). Asked more specifically about aspects of vulnerabilities management, participants were free to discuss elements that were relevant to them. Many mentioned being in favour of 'responsible disclosure' (n = 5), whereas a large majority expressed critical views on withholding vulnerabilities (n = 9). The use of State Trojans is viewed rather negatively by participants – one participant expressed support and three participants expressed criticism. On the subject of 'hack backs/active cyber defence', attitudes differ. While four participants expressed positive attitudes, two expressed negative attitudes – and seven participants made no explicit statement. The statements of P05 and P06 on this topic illustrate the different attitudes well: "*So hacker attacks [are] active warfare in my opinion. [...] As a civilian, I don't want to be involved in that.*" (P05)

"*If the German Parliament is now attacked, as in the case of the German Parliament hack ['Bundestagshack'], and you then try to find out where it came from by means of a hack back, that is, you are doing a bit of forensics. I think that is definitely acceptable.*" (P06)

As a result, **RQ1a** can be answered as follows: Asked about ethical aspects of computer science, the topic of 'artificial intelligence' was most present in students' minds. Two intervie- wees mentioned the export of 'State Trojans' as problematic. Issues related to vulnerabilities were not primary ethical concerns.

Explicitly asked about managing vulnerabilities management, interviewees considered withholding security vulnerabilities as critical, but showed different opinions on the topic of 'hack backs/active cyber defence'. Although there seems to be a general awareness of the issue of managing vulnerabilities, this is only transferred to a limited extent to topics in which vulnerabilities are exploited.

4.2 Ethical Issues Related to Cyber Volunteering (RQ1b)

Regarding ethical issues related to cyber-volunteering, participants primarily have security and privacy concerns: Because volunteers would need far-reaching permissions to navigate the affected IT systems, four participants cited a potential risk of abuse by volunteers – for example, they might pass on confidential information. One participant, similarly, fears that if volunteers needed to enter other people's homes for an intervention, they might invade their privacy. Three participants identified the risk that companies could shift responsibility for their cyber security to the volunteer force.

Table 2. Interview guide and contribution of each interview section to the research questions

Interview section	Contribution to RQ
1. Role of ethics in informatics	**RQ1a**
Introduction to the interview and general attitude towards ethics in IT, awareness of ethical issues related to IT. Questions asked:	
– *"What role has ethics played for you so far in computer science or in computer science studies?"*	
– *"What ethical questions have you encountered concerning IT outside of your studies and where?"*	
2. Examples of considerations in cyber security	**RQ1a, RQ1b**
Topics *"Vulnerability Management"* and *"Civilian Security and Volunteers"* are introduced. For each topic, participants are asked to locate themselves between two poles of a constructed spectrum – between supporting civilian security (privacy and responsible disclosure) vs. supporting more power for state security institutions (active cyber defence and surveillance); and between supporting disaster response based on volunteers vs. permanent employees. Questions asked:	
– *"Where do you locate yourself, and why?"*	
3. Concept of volunteer force	**RQ1b, RQ3**
Introduction of growing threat posed by increasing digitalisation of CI and introduction of the concept of cyber volunteers. Survey of ethical and practical concerns on establishing such a volunteer force, as well as perceived advantages and disadvantages of using volunteers in large-scale cyber emergencies. Questions asked:	
– *"What ethical issues does this touch on? In what way?"*	
– *"What would practically stop you from participating in something like this?"*	
– *"What would be appealing to you about participating in something like this?"*	
4. Scenario of a mission	**RQ1b, RQ3**
Two scenarios describe a situation in which volunteers discover a security breach using a previously unknown vulnerability during a mission. The vulnerabilities are handed over to superiors. A while later, the same vulnerability appears to have been used to attack another water facility in the own country (scenario 1) or an Iranian power plant abroad (scenario 2). The case of an attack on Iran was constructed in order to imply a political intention to withhold and use an exploit. This question evokes participants' reflection on the consequences of unpublished vulnerabilities in the context of a cyber volunteer force. Questions asked for each of the two scenarios:	
– *"What are your thoughts on this?"*	
– *"What do you think, could your actions have been insufficient?"*	
– *"Would this affect your view of the volunteer force or your work in it?"*	
5. Principles guiding co-operation with other security organisations	**RQ2**
Co-operation with other agencies could increase a volunteer force's contribution, but it can also decrease the organisations' independence. This is an issue that is also being criticised with regard to Germany's cyber security organisation, which is subordinated to the Ministry of the Interior, which has developed a State Trojan for policing (Meister, 2015). To find out whether different types of inter-organisational co-operation would be supported, we dicuss three versions of "guidelines" of a fictional cyber squad. The sections "Reason for joining the volunteer force", "Objective" as well as "Main task" are less controversial and therefore only briefly presented. In the sections "Deployment" and "Agency co-operation" three alternatives are presented each, which participants were then asked to assess. The alternatives differ in the extent to which the volunteer force is independent of the security authorities. Questions asked for each guideline section:	
– *"Which of these phrases would be your personal favourite and why?"*	
– *"What do you think of the other wordings? Why not the others?"*	
– *"With which wording would you abstain from volunteering?"*	
6. Final questions	**RQ1b, RQ3**
Summary and conclusion of the interview, opportunity to mention new aspects that have emerged. Questions asked:	
– *"How do you think the concept of the volunteer force would be received by your friends or fellow students?"*	
– *"Are there things that would now prevent you from joining the volunteer force?"*	
– *"Should it come to existence, would you volunteer in the volunteer force?"*	

Asked again at the end of the study whether anything might prevent participants from supporting a cyber volunteer force, there was, unsurprisingly, an increased awareness of issues relating to vulnerabilities: Two participants stated that the ethical implications of identifying vulnerabilities in the course of volunteering activities would not be obvious at first. Furthermore, two participants mentioned that they would be uncomfortable with the uncertainty about how an identified vulnerability would be handled and whether it might be withheld. An overview of the ethical concerns expressed, as well as example quotes, can be found in the appendix in Table A1.

In the further course of the interviews, two scenarios were presented to participants, one after the other. Both scenarios deal with a situation in which a vulnerability was found and reported to authorities after a previous volunteer force

deployment and was exploited again after some time – which could mean that the vulnerability may have been withheld. However, this possibility was not brought to participants' attention until scenario 2.

As a reaction to scenario 1, most participants (n = 9) showed negative emotions such as anger, frustration, demotivation and disappointment. The majority (n = 8) of interviewees' view of the volunteer group and their work would be negatively influenced by scenario 1. Only two participants stated that their view would not be influenced. Interviewees mainly felt that their work as volunteers was made futile and was not made good use of, if the vulnerability was not patched. At the same time, almost half of the participants (n = 5) saw their own actions of reporting the vulnerabilities to a higher authority as sufficient, only one participant expressed the opposite view, indicating that he would consider responsibly disclosing a responsibility himself if the authorities were inactive. Four of the five participants mentioned placing the responsibility in this scenario with the higher-level authority to which the security breaches were reported. Four other participants considered whether this authority might not have published the security vulnerabilities in this scenario due to overload, for example due to a lack of resources.

In scenario 2, interviewees were then made aware of the possibility that found and consequently reported security vulnerability had been withheld and exploited by attacking a facility in another country. Here, participants also expressed negative emotions – five interviewees expressed their anger and frustration. Another five participants expressed that it is the task of the volunteer force to help, not to cause harm. Three participants remarked that the attribution of the attack on the Iranian nuclear power plant was difficult to understand. Most participants (n = 8) expressed that under such conditions, they would refrain from participating. They mainly reasoned that such use of vulnerabilities would diminish rather than increase security and go against their motivation for joining. For example, one participant stated: *"Yes, it is this dual-use issue that comes up to some extent. That I myself have created a weapon for my intelligence service through my own knowledge, which can now partially attack others, especially the general population. Yes, so then I think I would drop out of the volunteer force if I found out that my knowledge and my volunteer work was being used to disrupt other systems"* (P06). One participant said he would still join the volunteer force, hoping that other vulnerabilities would be fixed. Two participants made no statement.

As a result, participants expressed the following ethical concerns regarding a volunteer cyber force (**RQ1b**): (1) Explicitly asked about ethical concerns, participants expressed the potential risk of abuse by volunteers (n = 4), shifting responsibility for cyber security to volunteers (n = 3) and invasion of privacy in a public volunteer force deployment (n = 1). Potential issues related to identified vulnerabilities were not mentioned at first. (2) However, withholding vulnerabilities would be regarded as highly problematic and disappointing, resulting in a large proportion of participants (n = 8) refraining from volunteering. (3) In general, two interviewees expressed that the ethical complications of withheld vulnerabilities would not be obvious at first. In addition, two participants saw the uncertainty about the consequences of one's actions as problematic.

4.3 Cyber Emergency Response and Security Agencies (RQ2)

In this section, participants are confronted with different possible guidelines for the volunteer force, which also provide information about its institutional design. This will discuss potential ethical and practical consequences of the "principles" of a volunteer force. The principles specify common reasons for deployment, objectives and main tasks. They vary regarding their possible deployments and co-operation with other security-related state institutions.

The alternatives of the principle on the use of the volunteer force are about when and how the volunteer force is used. In alternative 1, the volunteer squad serves civilian defence and is limited to peacetime actions. Cyber attacks, including hack backs and support of government agencies are excluded. In alternative 2, the squad also serves civilian defence, but hack backs are not excluded, and support of government agencies may be considered, e.g. by sharing tools. In alternative 3, the squad serves to protect CI in Germany. In the case of defence, this may include cyber attacks including hack backs and the support of state agencies.

Interviewees were now asked about their favourite alternative. Most interviewees (n = 8) named alternative 1 as their favourite. One participant favoured alternative 2, while two participants were undecided between alternative 1 and alternative 2 as their favourite. Regarding alternative 2, four participants expressed that they would rule out participation with such an organisation, while three interviewees would only participate with restrictions – e.g., if participation in hack backs were voluntary. Nine participants were critical of alternative 3 and all ruled out participation under these conditions. An overview with example quotes of the participants' statements can be found in the appendix in Table A2.

The three alternatives of the principle on inter-organisational co-operation deal with tools and information, such as knowledge of security vulnerabilities, and whether they may be passed on to security authorities. Under alternative 1, tools may not be shared with security authorities due to their presumed dual-use nature, and the retention of vulnerabilities is excluded. This also precludes the evaluation of vulnerabilities in a Vulnerabilities Equities Process (VEP),

which assesses whether a found vulnerability may be withheld for government use. In alternative 2, however, the volunteer group submits identified vulnerabilities to a VEP and can share knowledge with security organisations, e.g. through consulting. However, it cannot share any developed tools. In alternative 3, identified vulnerabilities are also judged in a VEP. But it also includes exceptions in a defence scenario. In case of an attack, not only may tools be passed on to security authorities, but relevant vulnerabilities can be withheld. All participants favoured alternative 1, indicating a clear preference for responsible disclosure. Nine participants stated that they would refrain from participating in the case of alternative 2 and alternative 3, while one participant would demand a say in the restraint of security breaches. The remaining two expressed that they attach particular importance to the defence case. An overview with example quotes can be found in the appendix in Table A3.

Regarding co-operation with other state agencies, we conclude (**RQ2**): (1) The majority of participants (n = 8) chose alternative 1 as their favourite, which precludes the support of state authorities. Yet, alternative 2 (which does not exclude this) is also considered by a part of participants – one participant chose this alternative as his favourite and only four interviewees explicitly stated that they would not participate in the volunteer force with this alternative. Thus, as in section 4.1, it can be seen that some participants see hack backs as acceptable and worthy of support. Alternative 3, however, where supporting other state agencies is expected, is ruled out by all interviewees. (2) With regard to the transfer of tools and information to other agencies, a clear tendency can be identified: All preferred to responsibly disclose all identified vulnerabilities. At the same time, a large proportion of participants (n = 9) excluded their participation in alternatives 2 and 3. Participants thus prefer independence from security authorities in this respect, and do not see a VEP as an adequate process for differentiating between vulnerabilities. While in a minority, two interviewees see the defence case as a special case and would support co-operation in such a case.

4.4 Incentives and Concerns (RQ3)

At first, we addressed participants' opinions concerning the feasibility of cyber volunteering generally. While many think that permanent staff has advantages (n = 7), there is also support for volunteering (n = 5), and a combination of both is favoured by a similar number (n = 4). Some participants mentioned advantages of using volunteers, including being able to mobilise more people (n = 1), greater flexibility (n = 1) and ensuring a higher level of assistance than with permanent staff alone (n = 1). The biggest practical concern relate to professional competences (n = 3), as IT and IT security are very complex topics.

When asked what would appeal to participants about joining a cyber security volunteer force, the majority (n = 7) cited altruistic motives, such as helping people, coping with disasters and generally doing good. Other incentives are getting to know new people and gaining experience (n = 4), having fun (n = 2) or experiencing a technical challenge (n = 1). Other positive aspects were described in the final interview section, when participants were asked about how the concept might be perceived by friends or fellow students. Two participants did not think that anyone would find the concept bad. Two other participants estimate that their fellow students would view the volunteer concept as positively as they do. Finally, one participant said that the concept of a volunteer force would allow for more public scrutiny and that it would be easier to find out if there was any wrongdoing in the organisation. In addition to positive aspects, two interviewees were critical of the organisations' position beneath the Ministry of the Interior. Another two interviewees stated that IT people generally do not have much trust in government IT security. The fact that their fellow students would view the concept as critically as they do is also estimated by two participants.

Asked about aspects that may prevent them from participating in the volunteer group, participants named professional, temporal and other obstacles. Among professional obstacles, many assume that that they would have insufficient knowledge (n = 6). One participant thinks that basic volunteer training would not suffice to adequately equip him to participate. Some (n = 2) have no interest in IT security. The main time-related obstacles mentioned were lack of time (n = 4), time-intensive training or further training (n = 2) and loss of working time at the full-time job (n = 1). In addition, one interviewee sees the duration of a volunteer team operation as a critical point – because in contrast to a fire brigade operation, which can be completed in half a day, IT incidents could last several days to weeks. Other obstacles include a fear of being overworked in the mission (n = 1), a fear of a bad working atmosphere in the volunteer force (n = 1) as well as a legal limitation on volunteering in more than two disaster relief organisations. When asked again, at the end of the interview, about obstacles to participate, mentions now primarily relate to the (ethical) use of the volunteer group, whereas at the beginning of the interview, participants mainly mentioned personal and organisational obstacles. Three participants stated that forwarding or withholding security vulnerabilities would prevent them from participating. Another three participants answered that it would depend on the final wording of the policies. Another three participants mentioned potential misuse of the volunteer group as an "instrument of the state".

Finally, in the last section of the interview, participants were asked explicitly whether they would volunteer in

such an organisation. Three interviewees were positive and said they would volunteer or could imagine doing so, while three other participants would not volunteer. Almost half of the participants (n = 5), on the other hand, would only volunteer under certain conditions. One participant, e.g., would want to talk to squad members beforehand to get an impression of it; another participant makes his commitment dependent on his personal environment and the assessment of the hacktivist NGO Chaos Computer Club (CCC). One interviewee stated that it would depend on the likeability of the people in the volunteer force, and another participant emphasised that his participation in another disaster volunteer force would currently legally prevent his participation. An overview of these answers with exemplary quotes can be found in the appendix in Table A4.

5. Discussion

This paper provides insights into computer science students' attitudes towards volunteering in cyber disaster response and their awareness of related ethical issues. Supporting previous findings, this study finds that students report having had only limited exposure to ethical issues in the course of their studies (Casañ et al., 2020; Polmear et al., 2018). Despite this self-reported gap, students expressed a surprisingly homogeneous attitude towards favouring responsible disclosures of software vulnerabilities. While we did not expressly explore this aspect, it is suggestive of a culture that has been influenced by the Chaos Computer Club (CCC), which has a long and influential history of activism and hacktivism related to decentralisation, open internet and privacy in Germany (Wagenknecht & Korn, 2016). However, another principle of the CCC is *"mistrust authority"* (Wagenknecht & Korn, 2016), which appeared to be less prevalent in our sample. This is indicated by the assumption of benign reasons for why an identified vulnerability might not have been patched. As a result, there is a discrepancy between the interviewees' explicit criticism of withholding vulnerabilities and a the perceived likelihood of states withholding vulnerabilities intentionally. One approach to explain this could be a lack of expertise regarding the strategic value of vulnerabilities. Another possibility might be that, similar to the study by Schneider (2013), students judge according to the "ethic of ultimate ends" (*"Gesinnungsethik"*) (Weber, 1919) and therefore consider the "good" intention of hacking a "bad" attacker using withheld vulnerabilities to be ethically acceptable. This is supported by some statements relating to "bad" states' activities or adversaries in war. However, this also suggests a lack of nuance regarding state activities and international relations, where interests, alliances and organisational structures are more relevant aspects, which the students did not seem to consider.

In the scenarios, most interviewees did not directly think that vulnerabilities could be abused. This only became clear to the majority of interviewees when scenario 2 was presented. Thus, awareness that security vulnerabilities can be withheld/exploited by the state was not directly present to all participants. Participants' statements could also be indicative of high trust in German agencies. However, students appeared to shift their judgements in the course of the interview, indicating towards the end that the handling of vulnerabilities is indeed an important problem as well as being critical of state stockpiling. This suggests that students became aware of the potential misuse of vulnerabilities by state agencies only in the course of the interview and lacked that awareness beforehand. Including more interdisciplinary courses and technical peace research in informatics curricula appear a promising avenue for helping students understand socio-political implications of their discipline (Reuter et al., 2022), the dual-use potential of ICT (Riebe & Reuter, 2019), or the possibility of vulnerabilities stockpiling (Reinhold & Reuter, 2019). Other studies suggest that reasons for the limited contact with ethical issues lie in the interest on the part of students (Polmear et al., 2018) or the stronger focus of degree programmes on technical specialisation (Casañ et al., 2020).

Looking at the motivations for volunteering, the study shows that a volunteer force is perceived as valuable and participation would be based on altruistic motives as well as with the aim of personal development. As such, the motivations appear to be similar to those that motivate volunteering in analogue disasters (Kehl et al., 2017). Because the main purpose of participants' engagement is to increase civilian security, finding that vulnerabilities were withheld would greatly demotivate most prospective volunteers. While IT specialists are strongly sought and highly-paid, only one participant mentioned that he was already working in an IT security company and would not forfeit a well-paid job in order to volunteer. Other participants were also sceptical that volunteering would take up a lot of time for the deployments and the training. This suggests that the recruitment of volunteers should take into consideration that prospective volunteers are unaware of the legal frameworks that exist for volunteering in analogue disaster management, which, for example, require employers to exempt employees for the duration of their missions. At the same time, the missions and training should be structured in a manner that limits the time strain on individual volunteers.

The results also show that students are unsure about whether they are sufficiently trained to participate in a cyber volunteer force. On the one hand, this could suggest that such an organisation should follow the Estonian model, with

participation which is limited to IT security experts. This is to a large extent due to its emergence: The Estonian "Cyber Defence League" was established after the Russian attacks on Estonian systems, as a response an ad hoc network of cyber security experts emerged to help confront the incident. Afterwards, the organisation was officially institutionalised as a branch of the Ministry of Defence (Kaska et al., 2013). Still, participation as a volunteer is limited to specialists. In contrast, in Germany, due to its militarised history, reservations towards its military prevail. At the same time, Germany has a culture of volunteering in civilian aid organisations. If the volunteer group would seek to connect to that history, rather than forming an expert network, efforts would have to be made to achieve wide participation. These could include providing training that enables volunteers to perform the tasks and encouraging potential volunteers to participate by stressing the importance of non-specialist roles.

In addition, the results offer further insights for the establishment of a cyber security volunteer force and the successful recruitment of volunteers: (1) *Appeal to a positive cause:* In the promotion of the volunteer force, the main focus should be on highlighting the altruistic aspects – that involvement helps to overcome disasters, helps people and generally does something good for society. The curiosity and inquisitiveness of potential volunteers can also be targeted, by emphasising training and learning opportunities. The organisation should also be designed to foster socialising and a sense of community. Gamification in solving technical challenges and assignments could also be addressed. (2) *Address practical concerns:* The practical obstacle to participation most frequently expressed by interviewees is the lack of technical knowledge. It is important to allay the fears of potential volunteers and to show that all the necessary knowledge can be provided within the framework of membership in the volunteer force. The concerns about time that are also expressed could be countered by communicating the average time required (e.g. in weekly hours) from the outset. (3) *Define principles:* The results show that interviewees prefer a separation between the volunteer force and state security organisations. It is therefore very important that before the volunteer force is established, a clear set of internal rules and principles is drawn up that conclusively define how sensitive information about vulnerabilities and exploits are to be dealt with. (4) Finally, it should be noted that according to a participant, the current Hessian Fire and Disaster Protection Act stipulates that volunteers may only be members of two relief organisations at the same time. Such restrictions are also conceivable in other federal states. When the volunteer force is established, political efforts should be made to ensure that the volunteer force does not fall under such regulations, as involvement in the volunteer force tends to result in fewer conflicts at emergency sites with other aid organisations. (5) *Avoid potential for misuse:* Participants frequently mentioned the potential danger of untrained and non-vetted volunteers. Security vetting, codes of conduct, anti-corruption and anti-failure systems, such as a two-person principle, should be implemented. Commitment to these principles should be proactively communicated to ensure that the organisation is perceived as trustworthy.

6. Limitations and Future Work

While the qualitative approach offers nuance and the emergence of unanticipated aspects, it also limits the scope of participants that could be included. Therefore, the sample of this study is small (N = 11) and not representative. Despite this limitation, the sample allows insights into potential volunteers' attitudes that can be further explored in a larger study and compared to communities that might be potential volunteers, such as hacktivists and IT security professionals. We interviewed students of computer science as proxies for potential future volunteers. Yet, this results in a rather young sample that is not representative of the IT profession. While participants study at different universities, making the study less dependent on a single curriculum, they were recruited in the same geographical area, with 8 persons studying at the same university, giving them access to the same choice of classes. Yet, most interviewees expressed they had only had limited contact with ethics in computer science during their studies. However, we do not expect a great influence of geography. Rather, future work might analyse whether students who study at universities that offer specific courses, e.g. related to ethics or peace and conflict studies (Reuter et al., 2022), differ in their awareness and judgements. In addition, the sample consists of only male participants. With a national average of 21.8 % females in computer science degree programmes (Statistisches Bundesamt, 2022), their representation is growing but still rather low. Despite several rounds of recruiting, this average could not be met. This is particularly regrettable, as the founding of a new volunteer group would be a good opportunity to include female perspectives and demands into the organisational design, to increase their participation, which is typically low in disaster management organisations. Past research indicates that men and women's reasons for volunteering were mostly the same, except that women are particularly motivated to volunteer in civil protection based on the motivation to, e.g., to gain knowledge and skills (Kehl et al., 2017).

Discussing a non-existent volunteer force for cyber security poses many challenges. The topic relates to many socio-political aspects, which requires knowledge of political, legal and social systems. Assumptions were made about framework conditions, partly in the scenarios and partly by

the participants – for example, related to the likelihood of a misused vulnerability. As the results show, these conditions are important for students' judgements. While these insights can shape the design of a volunteer force, its ultimate framework should again be investigated under the socio-economic and political conditions then. In order to fully understand the complexities of managing vulnerabilities, a lot of knowledge about state security organisations and international relations is required. Focus group sessions or expert interviews could add important knowledge. Especially working IT security professionals, who are prime candidates as volunteers due to their professional qualifications, could be another important stakeholder group to be analysed. It can be assumed that IT security professionals are already sensitised to the ethical dilemmas of dealing with vulnerabilities, and can therefore provide another valuable perspective. At the same time, it is important that further research is conducted with a larger sample to ensure greater general validity of the results. Despite the neutral design of the question, the time spent discussing potential uses of exploits to intervene in other systems might have lead participants to try to meet a perceived norm of being critical towards these issues.
The interviews were conducted in the period of October 2021 to January 2022, and therefore took place before the start of the Russian war against Ukraine. This is a limitation that must be considered, as two participants emphasised attributing particular importance to cyber defence. Concerning the connection of the volunteer force to security agencies, these views may have changed significantly in light of the attack.

7. Conclusion

This study used guided interviews with computer science students (N = 11) to explore attitudes regarding the concept of a cyber security volunteer force. Four research questions addressed potential volunteers' awareness regarding ethics in informatics, especially related to the handling of security vulnerabilities (RQ1a), the ethical concerns (RQ1b), preferences with regard to the organisational structure (RQ2), and practical incentives and challenges related to joining a cyber security volunteer force (RQ3). The analysis of the discussions revealed the following key findings:

- Regarding the awareness of ethics in computer science (**RQ1a**) participants are mainly conscious of algorithmic biases and problems regarding algorithmic decisions.
- Participants strongly favour the responsible disclosure of vulnerabilities, while some simultaneously supported hack backs, i.e. defensive measures which rely on the exploitation of vulnerabilities.
- While participants supported the extension of state reactive capabilities, they were also open towards establishing a volunteer force and perceived that many in the IT security community would be prepared to volunteer.
- Regarding ethical concerns (**RQ1b**), volunteers mainly mentioned the risk of abuse of information and secrecy by volunteers (n = 4) and the shifting of responsibility for cyber security onto the volunteers (n = 3).
- Constructing a scenario that suggested that vulnerabilities found by the volunteer force were withheld by state agencies, participants initially suspected benign reasons rather than strategic intentions on the part of state agencies, indicating low levels of mistrust or unawareness of the strategic motives of using vulnerabilities.
- In case of intentionally withholding vulnerabilities, a majority of participants would refrain from engaging in the volunteer force, as this is perceived as running contrary to the main aim of volunteering, which is seen in increasing civilian security.
- When it comes to the preferences regarding the co-operation with state security agen- cies (**RQ2**), participants largely prefer independence of the volunteer force and the responsible disclosure of vulnerabilities rather than a VEP. A minority (n = 2) regard a defence case as an exception in which closer cooperation with state security agencies might be feasible.
- Regarding practical incentives and concerns (**RQ3**), altruistic motives of helping others (n = 7), getting to know new people and gaining experience (n = 4) were mentioned as the main incentives. The biggest practical concerns were seen in the lack of technical knowledge (n = 6) as well as a lack of time (n = 4) and time-intensive education and training (n = 2). Eight out of eleven interviewees stated that they would be interested in joining a volunteer force for large-scale IT security incidents, whereas three interviewees would not volunteer due to a lack of interest in IT security.

While the findings cannot be generalised due to the small unrepresentative sample, the study contributes first insights related to volunteering in cases of IT incidents from potential volunteers' perspectives and interesting aspects that should be explored in larger studies and compared with attitudes of IT security professionals. We derive implications for aspects that could be used to motivate potential volunteers to participate and suggest misconceptions and fears that should be addressed.

Acknowledgements
This research work has been funded by the German Federal Ministry of Education and Research and the Hessian State Ministry for Higher Education, Research and the Arts within their joint

support of the National Research Center for Applied Cybersecurity ATHENE and the LOEWE initiative (Hessen, Germany) within the emergenCITY centre.

References

Ablon, L., Libicki, M., & Abler, A. (2014). *Markets for Cybercrime Tools and Stolen Data: Hackers' Bazaar*. RAND Corporation. DOI:10.7249/RR610.

AG KRITIS. (2020). Das Cyber-Hilfswerk: Konzept zur Steigerung der Bewältigungskapazitäten in Cyber-Großschadenslagen. https://web.archive.org/web/20220928113547/ https://ag.kritis.info/wp-content/uploads/2020/02/chw-konzept_v1.0.pdf

Athey, S. (1993). A comparison of experts' and high tech students' ethical beliefs in computer- related situations. *Journal of Business Ethics*, *12*(5), 359–370. DOI : 10.1007 / BF00882026.

Balaban, S., Boehm, F., Brodowski, D., Dickmann, R., Franzen, F., Goerke, N., Golla, S., Koloßa, S., Kreutzer, M., Krüger, J., Leicht, M., Obermaier, J., Pieper, M., Schink, M., Schreiber, L., Schuster, D., Sorge, C., Tran, H., Vettermann, O., . . . Wagner, M. (2021). Whitepaper zur Rechtslage der IT-Sicherheitsforschung: Reformbedarf aus Sicht der angewandten Sicherheitsforschung. https:// web.archive.org/ web/ 20221014140029/https://sec4research.de/assets/Whitepaper.pdf

Boeke, S. (2018). National cyber crisis management: Different European approaches. *Governance*, *31*(3), 449–464. DOI:10.1111/gove.12309.

Bundesamt für Sicherheit in der Informationstechnik (Ed.). (2022a). Digitaler Ersthelfer: Firstlevel Support für Verbraucherinnen und Verbraucher. https://web.archive.org/ web/20221125133914/ https://www.bsi.bund.de/DE/Themen/Unternehmen-und- Organisationen/Informationen- und- Empfehlungen/Cyber- Sicherheitsnetzwerk/ Qualifizierung/Digitaler_Ersthelfer/Digitaler_Ersthelfer_node.html

Bundesamt für Sicherheit in der Informationstechnik (Ed.). (2022b). Vorfallunterstützung – Mit CERT-Bund und MIRT: CERT-Bund und Mobile Incident Response Teams leisten BSI-Support vor Ort. https://web.archive.org/web/20220502160340/https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Cyber- Sicherheitslage/Reaktion/Vorfallunterstuetzung/MIRT/mirt_node.html

Cardash, S. L., Cilluffo, F. J., & Ottis, R. (2013). Estonia's Cyber Defence League: A Model for the United States? *Studies in Conflict & Terrorism*, *36*(9), 777–787. DOI:10.1080/1057610X.2013.813273.

Casañ, M. J., Alier, M., & Llorens, A. (2020). Teaching Ethics and Sustainability to Informatics Engineering Students, An Almost 30 Years' Experience. *Sustainability*, *12*(14), 5499. DOI:10.3390/su12145499.

Code of Federal Regulations. (2013). 15 C.F.R. § 730.3: "Dual use" and other types of items subject to the EAR. https://web.archive.org/web/20221125134812/https://www.law.cornell.edu/cfr/text/15/730.3

Collier, J. (2017). Strategies of Cyber Crisis Management: Lessons from the Approaches of Estonia and the United Kingdom. In M. Taddeo & L. Glorioso (Eds.), *Ethics and Policies for Cyber Operations* (pp. 187–212). Springer International Publishing. DOI:10.1007/978-3-319-45300-2_11.

Cyberwehr-bw.de (Ed.). (2022). Cyberwehr – Startseite. https:// web.archive. org/ web/ 20221125134130/https://cyberwehr-bw.de/startseite

Ding, A. Y., de Jesus, G. L., & Janssen, M. (2019). Ethical hacking for boosting IoT vulnerability management: A First Look into Bug Bounty Programs and Responsible Disclosure. In A. Lazarov, B. Shishkov, D. Mitrakos, & M. Janssen (Eds.), *Proceedings of the Eighth International Conference on Telecommunications and Remote Sensing - ICTRS '19* (pp. 49–55). ACM Press. DOI:10.1145/3357767.3357774.

Dunn Cavelty, M. (2014). Breaking the cyber-security dilemma: aligning security needs and removing vulnerabilities. *Science and engineering ethics*, *20*(3), 701–715. DOI: 10.1007/s11948-014-9551-y.

Dürig, M., & Fischer, M. (2018). Cybersicherheit in Kritischen Infrastrukturen. *Datenschutz und Datensicherheit - DuD*, *42*(4), 209–213. DOI:10.1007/s11623-018-0909-1.

European Commission. (2018). Dual-use trade controls. https:// web.archive. org/ web/ 20220928114512/https://policy.trade.ec.europa.eu/help-exporters-and-importers/ exporting-dual-use-items_en

Fathi, R., Thom, D., Koch, S., Ertl, T., & Fiedrich, F. (2020). VOST: A case study in voluntary digital participation for collaborative emergency management. *Information Processing & Management*, *57*(4). DOI:10.1016/j.ipm.2019.102174.

Fiedrich, F., & Fathi, R. (2021). Humanitäre Hilfe und Konzepte der digitalen Hilfeleistung. In C. Reuter (Ed.), *Sicherheitskritische Mensch-Computer-Interaktion* (pp. 539–558). Springer Fachmedien Wiesbaden. DOI:10.1007/978-3-658-32795-8_25.

Forge, J. (2010). A Note on the Definition of "Dual Use". *Science and engineering ethics*, *16*(1), 111–118. DOI:10.1007/s11948-009-9159-9.

FragDenStaat.de. (2018). ZV 32-21-5391.04-2/18 – Widerspruchsbescheid zur Anfrage Überprüfung von Produkten der ITÜ. https://web.archive.org/web/20221125134718/ https://fragdenstaat.de/anfrage/uberprufung-von-produkten-der-itu/

Gedris, K., Bowman, K., Neupane, A., Hughes, A., Bonsignore, E., West, R., Balzotti, J., & Hansen, D. (2021). Simulating Municipal Cybersecurity Incidents: Recommen- dations from Expert Interviews. In T. Bui (Ed.), *Proceedings of the 54th Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences. DOI:10.24251/hicss.2021.249.

Haas, J. M. (2020). Ehrenamtliche Führung als Rückgrat des deutschen Bevölkerungsschutzes. In E.-M. Kern, G. Richter, J. C. Müller, & F.-H. Voß (Eds.), *Einsatzorganisationen* (pp. 379–397). Springer Fachmedien Wiesbaden. DOI:10.1007/978-3-658-28921-8_22.

Hashemian, G., & Loui, M. C. (2010). Can instruction in engineering ethics change students' feelings about professional responsibility? *Science and engineering ethics*, *16*(1), 201–215. DOI:10.1007/s11948-010-9195-5.

Helfferich, C. (2011). *Die Qualität qualitativer Daten: Manual für die Durchführung qualitativer Interviews*. VS Verlag für Sozialwissenschaften. DOI:10.1007/978-3-531- 92076-4.

Helfferich, C. (2019). Leitfaden- und Experteninterviews. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp. 669–686). Springer Fachmedien Wiesbaden. DOI:10.1007/978-3-658-21308-4_44.

Herpig, S. (2018). *Governmental Vulnerability Assessment and Management: Weighing Temporary Retention versus Immediate Disclosure of 0-Day Vulnerabilities*. Stiftung Neue Verantwortung. https://web.archive.org/web/20220728003819/https://www. stiftung-nv.de/sites/default/files/vulnerability_management.pdf

Herpig, S., & Rupp, C. (2021). *Deutschlands staatliche Cybersicherheitsarchitektur* (6th ed.). Stiftung Neue Verantwortung. https://web.archive.org/web/20220928114907/https://www.stiftung-nv.de/sites/default/files/deu_impuls- deutschlands_staatliche_ cybersicherheitsarchitektur_6._auflage.pdf

Jaquet-Chiffelle, D.-O., & Loi, M. (2020). Ethical and Unethical Hacking. In M. Christen, B. Gordijn, & M. Loi (Eds.), *The Ethics of Cybersecurity* (pp. 179–204, Vol. 21). Springer International Publishing. DOI:10.1007/978-3-030-29053-5_9.

Jasiūnas, J., Lund, P. D., & Mikkola, J. (2021). Energy system resilience – A review. *Renewable and Sustainable Energy Reviews*, *150*, 111476. DOI:10.1016/j.rser.2021. 111476.

Jazdi, N. (2014). Cyber physical systems in the context of Industry 4.0. *2014 IEEE International Conference on Automation, Quality and Testing, Robotics*, 1–4. DOI: 10.1109/AQTR.2014.6857843.

Kaska, K., Osula, A.-M., & LTC Stinissen, J. (2013). The Cyber Defence Unit of the Estonian Defence League: Legal, Policy and Organisational Analysis. https://ccdcoe.org/ uploads/2018/10/CDU_Analysis.pdf

Kehl, D., Kietzmann, D., & Schmidt, S. (2017). Reasons for Volunteering in the Field of Civil Protection in Germany. *Journal of Homeland Security and Emergency Management*, *14*(1). DOI:10.1515/jhsem-2016-0042.

Kuckartz, U., & Rädiker, S. (2020). *Fokussierte Interviewanalyse mit MAXQDA*. Springer Fachmedien Wiesbaden. DOI:10.1007/978-3-658-31468-2.

Lechner, U. (Ed.). (2018). *Monitor 2.0: IT-Sicherheit Kritischer Infrastrukturen* (1. Auflage). Universitätsbibliothek der Universität der Bundeswehr München. DOI:5341.

Leinhos, L. (2019). Der Organisationsbereich Cyber- und Informationsraum (CIR) als wichtiger Teil einer gesamtstaatlichen Sicherheitsvorsorge. *Ethik und Militär*, (1), 46–50. https : / / web. archive. org/web/ 20220916113228 / https : / / epub. sub. unihamburg.de/epub/volltexte/2019/91700/pdf/Ethik_und_Militaer_2019_1.pdf

Lin, H. (2016). Governance of Information Technology and Cyber Weapons. In E. D. Harris (Ed.), *Governance of Dual-Use Technologies* (pp. 112–157). American Academy of Arts and Sciences.

Ludwig, T., Stein, M., Castelli, N., & Hoffmann, S. (2021). Sicherheitskritische Mensch- Maschine-Interaktion bei Industrie 4.0. In C. Reuter (Ed.), *Sicherheitskritische Mensch-Computer-Interaktion* (pp. 253–276). Springer Fachmedien Wiesbaden. DOI:10.1007/978-3-658-32795-8_12.

Mayring, P., & Fenzl, T. (2019). Qualitative Inhaltsanalyse. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp. 633–648). Springer Fachmedien Wiesbaden. DOI:10.1007/978-3-658-21308-4_42.

Meister, A. (2015). Geheime Kommunikation – BSI programmierte und arbeitete aktiv am Staatstrojaner, streitet aber Zusammenarbeit ab (Netzpolitik.org, Ed.). https://web.archive.org/web/20221125122645/https://netzpolitik.org/2015/geheime- kommunikation-bsi-programmierte-und-arbeitete-aktiv-am-staatstrojaner-streitet- aber-zusammenarbeit-ab/

Meyer, P. (2020). Norms of Responsible State Behaviour in Cyberspace. In M. Christen, B. Gordijn, & M. Loi (Eds.), *The Ethics of Cybersecurity* (pp. 347–360, Vol. 21). Springer International Publishing. DOI:10.1007/978-3-030-29053-5_18.

Michaelsen, M. (2020). The Digital Transnational Repression Toolkit, and Its Silencing Effects: Special Report 2020. https://web.archive.org/web/20221125142223/ https://freedomhouse.org/report/special- report/2020/digital- transnational- repression- toolkit-and-its-silencing-effects

Miller, S. (2018). *Dual Use Science and Technology, Ethics and Weapons of Mass Destruction*. Springer International Publishing. DOI:10.1007/978-3-319-92606-3.

National Cyber Security Centre (Ed.). (2018). Coordinated Vulnerability Disclosure: the Guideline: National Cyber Security Centre of the Netherlands. https://web.archive.org/web/20220928120853/https://english.ncsc.nl/binaries/ncsc-en/documenten/ publications/2019/juni/01/coordinated- vulnerability- disclosure- the-guideline/ WEB_Brochure-NCSC_EN.pdf

Netzpolitik.org. (2021). Bundespolizeigesetz: Große Koalition einigt sich auf Staatstrojaner-Einsatz schon vor Straftaten. https://netzpolitik.org/2021/bundespolizeigesetz-grosse-koalition-einigt-sich-auf-staatstrojaner-einsatz-schon-vor-straftaten/

Polmear, M., Bielefeldt, A., Knight, D., Swan, C., & Canney, N. (2018). Faculty Perceptions of Challenges to Educating Engineering and Computing Students About Ethics and Societal Impacts. *2018 ASEE Annual Conference & Exposition Proceedings*. DOI:10.18260/1-2--30510.

Reinhold, T. (2020). Cyberspace as Military Domain: Monitoring Cyberweapons. In D. Feldner (Ed.), *Redesigning Organizations* (pp. 267–278). Springer International Publishing. DOI:10.1007/978-3-030-27957-8_20.

Reinhold, T., & Reuter, C. (2019). Arms Control and its Applicability to Cyberspace. In C. Reuter (Ed.), *Information Technology for Peace and Security* (pp. 207–231). Springer Fachmedien Wiesbaden. DOI:10.1007/978-3-658-25652-4_10.

Reuter, C., & Kaufhold, M.-A. (2018). Fifteen years of social media in emergencies: A retro- spective review and future directions for crisis Informatics. *Journal of Contingencies and Crisis Management*, *26*(1), 41–57. DOI:10.1111/1468-5973.12196.

Reuter, C., Riebe, T., Haunschild, J., Reinhold, T., & Schmid, S. (2022). Zur Schnittmenge von Informatik mit Friedens- und Sicherheitsforschung: Erfahrungen aus der inter- disziplinären Lehre in der Friedensinformatik. *Zeitschrift für Friedens- und Konfliktforschung*. DOI:10.1007/s42597-022-00078-4.

Riebe, T., Kaufhold, M.-A., & Reuter, C. (2021). The Impact of Organizational Structure and Technology Use on Collaborative Practices in Computer Emergency Response Teams: An Empirical Study. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–30. DOI:10.1145/3479865.

Riebe, T., & Reuter, C. (2019). Dual-Use and Dilemmas for Cybersecurity, Peace and Technology Assessment. In C. Reuter (Ed.), *Information Technology for Peace and Security* (pp. 165–183). Springer Fachmedien Wiesbaden. DOI:10.1007/978-3-658-25652-4_8.

Rundfeldt, J. (2020). CyberHilfsWerk - Konzeption für eine Cyberwehr 2.0: Vortrag auf der Konferenz DefensiveCon am 07. Februar 2020. *DefensiveCon*. https://www. defensivecon.org/dcon2020/talk/77KJJQ/

Schallbruch, M. (2017). IT-Sicherheitsrecht – Schutz kritischer Infrastrukturen und staatlicher IT-Systeme. *Computer und Recht*, *33*(10). DOI:10.9785/cr-2017-1007.

Schallbruch, M. (2020). Der Staat als Hacker. In T. Klenk, F. Nullmeier, & G. Wewer (Eds.), *Handbuch Digitalisierung in Staat und Verwaltung* (pp. 527–537). Springer Fachmedien Wiesbaden. DOI:10.1007/978-3-658-23668-7_49.

Schneider, C. (2013). “Das muss man immer für sich selber abwägen” oder: Das moralische Wissen von Studierenden der Informatik. *Informatik-Spektrum*, *36*(3), 287–292. DOI:10.1007/s00287-013-0695-y.

Schulze, M. (2019). *Governance von 0-Day-Schwachstellen in der deutschen Cyber-Sicherheitspolitik (SWP-Studie 10)*. Stiftung Wissenschaft und Politik – Deutsches Institut für Internationale Politik und Sicherheit. DOI:10.18449/2019S10.

Scott-Railton, Campo, E., Marczak, B., Razzak, B. A., Anstis, S., Böcü, G., Solimano, S., & Deibert, R. (2022). CatalanGate: Extensive Mercenary Spyware Operation against Catalans Using Pegasus and Candiru. https://web.archive.org/web/20221125141357/ https://citizenlab.ca/2022/04/catalangate-extensive-mercenary-spyware-operation- against-catalans-using-pegasus-candiru/

Statistisches Bundesamt (Ed.). (2022). Studierende in Mathematik, Informatik, Naturwis- senschaft (MINT) und Technik-Fächern. https://web.archive.org/web/20220929151056/ https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung- Kultur/Hochschulen/Tabellen/studierende-mint-faechern.html

Technisches Hilfswerk (Ed.). (2022). VOST – Cyberprofis im THW. https://web.archive. org/ web/ 20220519114437 / https:// www.thw.de/ SharedDocs/ Meldungen/ DE/ Veranstaltungen/national/2022/05/meldung_001_vost.html

Terberl, C. (2015). Akteure des Katastrophen- und Bevölkerungsschutzes in Deutschland. In H.-J. Lange & C. Gusy (Eds.), *Kooperation im Katastrophen- und Bevölkerungss- chutz* (pp. 17–63). Springer Fachmedien Wiesbaden. DOI:10.1007/978-3-658- 07151-6_2.

Thomas, G., Low, G., & Burmeister, O. (2018). “Who Was That Masked Man?”: System Penetrations—Friend or Foe? In H. Prunckun (Ed.), *Cyber Weaponry* (pp. 113–124). Springer International Publishing. DOI:10.1007/978-3-319-74107-9_9.

Verizon (Ed.). (2021). DBIR: 2021 Data Breach Investigations Report. Executive Summary. https : / / www.verizon .com / business / resources / reports / 2021 - dbir - executive - brief.pdfx

Wagenknecht, S., & Korn, M. (2016). Hacking as Transgressive Infrastructuring. In D. Gergle, M. R. Morris, P. Bjørn, & J. Konstan (Eds.), *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 1104–1117). ACM. DOI:10.1145/2818048.2820027.

Weber, M. (1919). *Politik als Beruf. In: Geistige Arbeit als Beruf. Vier Vorträge vor dem Freistudentischen Bund. Zweiter Vortrag*. Duncker & Humblot München. http://www.deutschestextarchiv.de/weber_politik_1919

Wenger, F., Jaquet-Chiffelle, D.-O., Kleine, N., Weber, K., Morgan, G., Gordijn, B., Inversini, R., Bangerter, E., & Schlehahn, E. (2017). Canvas White Paper 3 Attitudes and Opinions Regarding Cybersecurity. *SSRN Electronic Journal*. DOI:10.2139/ssrn. 3091920.

Wenzel, D., Beerlage, I., & Springer, S. (2012). Das Ehrenamt im Bevölkerungsschutz Deutschlands. In D. Wenzel, I. Beerlage, & S. Springer (Eds.), *Motivation und Haltekraft im Ehrenamt* (pp. 5–16, Vol. 39). Centaurus Verlag & Media. DOI : 10.1007/978-3-86226-978-5_2.

A Appendix

Table A1. Ethical concerns of participants towards the volunteer concept

Participants' ethical concerns towards the volunteer concept	No. of P.
– Potential risk of abuse by volunteers	**n = 4**
"Yes, I don't know what it's like to rely on volunteers and give them such responsibility. [...] If you give people who are relatively unscreened a lot of permissions in this system, maybe that's not right. [...] " (P06)	
– Shifting responsibility for cyber security to volunteers	**n = 3**
"Then I actually rather see [...] the operator of this hydroelectric power plant as being obligated to see to it that this is fixed again. And for that they should, may, however, [...] hire external consultants. I don't know how I like it when they try to pass this on to people." (P05)	
– Concerns about invasion of privacy during public operations	**n = 1**
"If something is really broken in the smart home or IoT sector [...] then private individuals would have to invade other people's privacy, i.e. their homes. Which is one thing with the police alone, they're not allowed to do that now either. [...] And I can imagine that maybe not everyone wants that [...]. " (P02)	
– No more engagement in case of retained vulnerability	**n = 8**
"I would never see myself supporting something like that. Because if [authorities] want to do that, let them find their own vulnerabilities. [...] Is it my job to find vulnerabilities to attack someone? Or is it my job to help someone who is in need? Those are two different approaches that don't belong together, can't belong together. That's why I would be out of there immediately." (P09)	
– Opacity of ethical implications	**n = 2**
"It's an exciting topic, and I think that [others] wouldn't think of these ethical pitfalls at the first moment. [...] I think you first have to think about it a bit, or consider exactly what is happening now with the vulner- abilities. Or, at the first moment, assuming that a scenario actually occurs where security vulnerabilities are withheld, I don't think that [a vuln. having been withheld is] the first thing you think about." (P03)	
– Uncertainty about consequences of one's actions (at risk of withholding vulnerabilities)	**n = 2**
"If [the volunteer force] was so tightly meshed with security agencies or intelligence services, then I would have reservations. Because you simply can't understand any more, 'What effect do my decisions have?' [...] If you work for a weapons company, you are aware that 'what I do may – or most likely will – kill people'. So there you have security. But with something like this, you have the constant uncertainty, 'What will happen as a result of my actions?'" (P03)	

Table A2. Participants' attitudes towards co-operation with other state security agencies.

Reactions to the principle on the use of the volunteer force	No. of P.
Favourite alternative:	
– Favourite is altern. 1, which excludes support of state agencies	**n = 8**
"Law enforcement, security, preparation, support or execution of cyber attacks and hack back scenarios, these are all things that should obviously be put in the hands of experts [of the authorities responsible for them]. [...] It really is sad if a state doesn't have the expertise to carry out these tasks and has to rely on volunteers to carry them out." (P01)	
– Favourite is altern. 2, where support of state agencies may be considered	**n = 1**
"I exclude alternative 1 because we have defined hack backs as not harming the target but stopping the attack. [...] The third [alternative], is aimed at the fact that I then have quasi special forces to harm another country. And I think that, as long as you're not in a war situation, that doesn't fit into the situation at all. That's why I think two is best." (P09)	
– Favourite undecided between alternative 1 and 2	**n = 2**
"[...] this can always be interpreted in different ways. I think that if hack backs are carried out to carry out forensic activities, [...] for example to find out the origin and other weaknesses or vulnerabilities, what they might know about you, in other words [activities] that really serve your own defence, that's perfectly fine. But especially the [second] scenario with the nuclear power plant that was attacked, that [...] does not necessarily serve one's own defence. [...] That's why I think that the second and third alternatives leave a lot of room to manoeuvre. " (P06)	
Participation in volunteer force:	
– Participation in alternative 2 only with restrictions	**n = 3**
"So the worst that can happen, I'll be kicked out of [the volunteer force] because I kind of refused. That's why I think I would continue to be there, just to do what I would like to do, and that is to help somehow.	
But I would definitely campaign for this paragraph to be changed." (P02)	**n = 2**
– No participation in alternative 2	**n = 4**
"In my opinion, hacker attacks are active warfare. As a civilian, I don't want to be involved in that." (P05)	
– No participation in altern. 3, where support of state is expected	**n = 11**
"I wouldn't be on board with the second and definitely not with the third alternative. [...] I could not ethically justify that people who are affected by such a security vulnerability have their systems hacked and suffer damage as a result. And that's why I wouldn't want to be part of it." (P11)	

Table A3. Participants' preferences concerning the handling of identified vulnerabilities through responsible disclosure or VEP

Principles for handling vulnerabilities	No. of P.
Favourite alternative:	
– Favourite is altern. 1, in which found vuln. are responsibly disclosed	**n = 11**
"I like [...] that one really publishes [vulnerabilities] within the framework of responsible disclosure, without the gaps existing and being withheld, that is also what I would want to exclude." (P07)	
– Defence case constitutes special situation	**n = 2**
"So [Alternative 1] simply does not define a procedure for the defence case. And the case of defence is actually a very extreme case, namely one is exposed to an attack that threatens the internal security of the state. And this was not necessarily the case in the previous scenarios. [...] In the case of defence, I am not opposed in principle to the existence of corresponding clauses that release or mobilise certain information for the case of defence. Or at least I could discuss that." (P01)	
Participation in volunteer force:	
– Participation in alternative 2 only with restrictions	**n = 1**
"[...] there might then be an in-between between the first and the second [alternative]. That there is the possibility [of handing a vulnerability to state a state agency], but that there are very strict requirements to clarify this internally first. Then [the vuln.] is either published or not. The government or the authorities [should] not have any direct influence or knowledge of it to influence people to want to do it[...]." (P08)	
– No participation in altern. 2, which involves participation in a VEP	**n = 9**
"I don't think I would be on board with the middle alternative. The Vulnerabilities Equities Process is not IT security as it should be." (P05)	
– No participation in altern. 3, which involves retention of vuln. in case of defence	**n = 9**
"Because it also goes in the direction of me being a pro-active war participant, which I don't want to be." (P10)	

Table A4. Readiness to engage in volunteer cyber force

Engagement in volunteer force	No. of P.
Want to get involved or can imagine doing so	**n = 4**
"I can imagine it. I can't guarantee it yet [...] but it would definitely appeal to me. Even if IT security isn't really my area of expertise now, it would still be something that would definitely appeal to me." (P10)	
Do not want to get involved	**n = 3**
"No [...]. I don't know what my career will be like, but IT security is not my hobby, as I said." (P05)	
Would only get involved under certain conditions:	**n = 4**
– Depending on proximity to local association	**n = 1**
"That's why if I'm in some village and I'm the only one who has to decide things in case of doubt, I wouldn't feel much like it. But all in all, yes." (P02)	
– Depending on judgement of social environment or expert groups	**n = 1**
"I could imagine it. But I would also make it very dependent on what my environment says about it. [...] And talk about it with people or try to talk about it with people who are also part of it." (P11)	
– Volunteer force must be independent from security authorities	**n = 1**
"I think that's a good thing, as long as the [principles constituting a separate set-up of the volunteer force from military and security authorities] come to pass." (P03)	
– People in volunteer force must be likeable	**n = 1**
"If the people who work there or are involved there are nice to me and it fits into my schedule." (P03)	
– Legal frameworks (such as HBKG) must allow for	**n = 1**
"It depends on the legal framework, so I wouldn't give up the volunteer fire fighters for that." (P07)	

DOI: 10.2478/9788366675896-007 • Technikpsychologie • 82-94

Proceedings of the International Symposium on Technikpsychologie

Exoskeleton developments at the Technical University of Darmstadt

Martin Grimmer†, Maximilian Stasica†, Guoping Zhao†

Technical University of Darmstadt Institute for Sport Science Magdalenenstraße 2764289, Darmstadt

Abstract

Exoskeletons are a potential solution to assist individuals with mobility needs. In this review we provide an overview of the major exoskeleton-related developments at the Technical University of Darmstadt. In addition to two upper limb exoskeletons that assist with load holding and carrying as well as drilling, we discuss exoskeleton developments that assist the lower limb during walking. Some of these developments were performed in collaboration with other national and international universities. To achieve the high-level target of mobility assistance, research focused on hardware and control developments and an understanding of human motor control. Most developed exoskeletons have used active designs that included motors for actuation. Stiff and soft (exosuits) materials have been used for the user-machine interface and a variety of concepts and sensors have been used for exoskeleton control. These developments have guided the exploration of solutions aimed at advancing the exoskeleton state of the art. This work has also led to several students being qualified in the interdisciplinary field of human-machine interaction.

Keywords

exoskeleton • mobility • assistance • human-machine interaction

1. Introduction

The number of elderly individuals above the age of 65 years old will double by the year 2050 due to increased lifespan and demographic changes [DESA, 2012, World Health Organization, 2011]. The quality of life of these individuals is greatly influenced by health and mobility, which are two factors that strongly influence each other. For instance, a higher number of steps per day is associated with positive health outcomes [Moreau et al., 2001, Swartz et al., 2003, Bravata et al., 2007, Murtagh et al., 2010, Erlichman et al., 2002]. However, with aging, the number of steps per day decreases by 75% [Tudor-Locke et al., 2013]. Moreover, with increasing age, there are decreases in walking speed (38%), lower extremity maximum muscle strength (33%), maximum oxygen consumption (40%), and muscle power (49%) [Grimmer et al., 2019b] Therefore, improving walking ability through some means of assistance is desirable to promote mobility and overall health.

Aside from elderly people, more than 500 million people suffer from reduced physical and functional capacity due to diseases that affect the respiratory, circulatory, musculoskeletal, or neurological system. While some neurological diseases or spinal cord injuries result in complete loss of walking ability, many afflicted with neurological conditions do have the ability to walk to some degree (e.g. Parkinson's disease, cerebral palsy, chronic obstructive pulmonary disease, peripheral vascular disease, osteoarthritis) [Grimmer et al., 2019b]. For those with walking ability, a limitation in one physiological system leads to declines in other physiological systems, and this starts a vicious cycle that is also known as the cycle of frailty [Fried et al., 2003]. For example, reduced lower extremity muscle force was found in people with respiratory, cardiovascular [McDermott et al., 2004], musculoskeletal [Tawil et al., 1994], osteoarthritis [Alnahdi et al., 2012], and neurological [Wiley and Damiano, 1998] diseases. However, the effects of the most limiting factor to the cycle of frailty can be broken through assisting or training the capacity that most limits mobility [Fried et al., 2003].

†*Corresponding author: Martin Grimmer; Maximilian Stasica; Guoping Zhao*
E-mail: grimmer@sport.tu-darmstadt.de; maximilian_alexander.stasica@tudarmstadt.de; guoping.zhao@tu-darmstadt.de

1.1 Exoskeletons

When summarizing the physical changes of elderly and people with mobility impairments due to diseases, it is obvious that there is a need for solutions to mitigate or prevent the loss of walking ability. Thus far, crutches, walkers, or wheelchairs are commonly used to support lower limb mobility impairments. Crutches and walkers take pressure off the joints to prevent pain and can help improve balance when walking, but these aids require significant use of the upper body and are therefore not suitable for every condition. Wheelchairs, on the other hand, do not provide an equivalent substitute for the ability to walk as many facilities are not designed for wheelchair accessibility.

Exoskeletons are a new generation of technical aids that have the potential to overcome these limitations as they address both the restoration and augmentation of physical and functional deficits. We believe that not only those with limited mobility could benefit from exoskeleton assistance, but also people with increased mobility needs such as nursing aides, orderlies, craft workers, machine operators, agricultural workers, and construction workers. By decreasing user effort for a specific movement task, exoskeletons could lower the risk of long-term physical damage. By reducing fatigue, it is possible to reduce the risk of injuries and increase worker effectiveness. However, a systematic review showed that there is little evidence thus far for such impacts [Steinhilber et al., 2020].

Passive orthoses (exoskeletons) to support human mobility may have existed for centuries, and designs of powered exoskeletons for assisting human mobility have been documented in patents since the early 20th century. Two examples include the Pedomotor and General Electric's Hardiman exoskeleton. The Pedomotor was a steam-powered exoskeleton proposed to assist with running [Kelley, 1919]. General Electric's Hardiman was one of the first exoskeletons to assist the upper and the lower limbs. This exoskeleton was designed to allowed users to lift objects of up to 680 kg and to provide walking assistance up to 0.8 m/s using 30 hydraulically-powered, servo-controlled joints [Mosher, 1968, Makinson, 1971]. However, due to several technological challenges, including the human-machine interaction and the huge system weight of 680 kg, user benefits were never demonstrated.

1.1.1 Lower limb

After stationary rehabilitation systems such as the Lokomat [Jezernik et al., 2003] showed reasonable performance in lower limb rehabilitation [Nam et al., 2017], a worldwide effort began to develop autonomous lower limb exoskeletons early in the 21st century. These exoskeletons included powered assistance for at least the hip and the knee joints and were connected to the human by a rigid structure that lies in parallel with the lower limb. These systems were developed for rehabilitation and to assist with walking in daily life [Aach et al., 2014, Strickland, 2012, Kilicarslan et al., 2013, Esquenazi et al., 2012, Quintero et al., 2011]. In addition, these exoskeletons were targeted to assist with lifting heavy objects and with sustaining fatigue-free load carrying for military, manufacturing, or medical applications [Kazerooni et al., 2005, Berkeley Bionics, 2014, Sankai, 2010, Yoshimitsu and Yamamoto, 2004]. A review of 27 published studies on neurorehabilitation concluded that it is easy to learn and to use these classical, rigid, and complete leg exoskeletons. Such exoskeletons can increase mobility, improve function, and reduce the risk of secondary injury by reinstating a more normal gait pattern [Federici et al., 2015]. They are also able to provide and improve locomotion capability for those impaired with neurological diseases.

However, almost none of these systems have been used to assist in daily life, nor have they demonstrated that they can provide benefit to the much larger population with reduced walking capabilities and limited physical and functional capacities [Grimmer et al., 2019b]. One major reason is due to their system mass of about 13-48 kg [Quintero et al., 2011, Kilicarslan et al., 2013] that significantly increases lower limb inertia and carrying effort, which essentially negates the walking and body weight support assistance. For example, 1 kg of mass placed on each foot will increase the walking effort, as determined by measuring metabolic cost, by 6% [Browning et al., 2007].

To avoid such contrary effects, device masses at the lower limb should be minimal and placed proximally to minimize inertial effects. Following these principles, recent minimalistic lower limb exosuits [Quinlivan et al., 2017, Schmidt et al., 2017] and exoskeletons have been developed to assist populations with walking capability but with limited mobility.

Based on analyses with monoarticular and multiarticular tethered lower limb exoskeletons during walking [Quinlivan et al., 2017, Schumacher et al., 2018, Grimmer et al., 2019a, Zeiss et al., 2020, Bryan et al., 2021], metabolic cost reductions of up to 50% [Franks et al., 2021] have been seen with assistance at the hip, knee, and ankle when only carrying the exoskeleton human-machine interface. However, minimalistic and autonomous lower limb exoskeletons weighing between 2 kg and 5 kg have only been able to demonstrate reductions of 20% thus far [Sawicki et al., 2020].

While most of these experiments were conducted with subjects that were free of mobility impairments, some studies have also focused on the target populations. When assisting stroke patients at a single joint (e.g. hip or ankle) with an exoskeleton, it was found that it is possible to reduce walking asymmetries [Bae et al., 2015, Buesing et al., 2015]. Improved balance recovery after slippage was also demonstrated in elderly individuals with a hip exoskeleton [Monaco et al., 2017].

Further, metabolic cost reductions of 4-7% were found when assisting elderly individuals with a hip [Lee et al., 2017, Martini et al., 2019] or an ankle [Galle et al., 2017] exoskeleton. For a single subject with an incomplete spinal cord injury, a walking speed increase of 30% and a metabolic cost decrease of 9% [Xiloyannis et al., 2020] were demonstrated when using the Myosuit [Schmidt et al., 2017] on a sloped mountain path.

1.1.2 Upper limb

Upper limb exoskeletons target lifting, lowering, holding, or carrying objects as well as rehabilitation [Rehmat et al., 2018, Thalman et al., 2018]. Similar to lower limb exoskeletons, several rigid, powered upper limb exoskeletons have been developed [Yoshimitsu and Yamamoto, 2004, Ferris, 2009, Sankai, 2010]. However, user benefits have mainly been achieved for passive devices [de Vries and de Looze, 2019, McFarland and Fischer, 2019]. These passive devices (e.g. PAEXO by OttoBock, [Maurice et al., 2019]) primarily use elastic elements in combination with rigid mechanical structures to unload the shoulder and arm. As passive exoskeletons have no actuators, electronics, or batteries, they can be lightweight and autonomous, have little safety and failure risk, and require little maintenance.

Studies have demonstrated that passive upper limb exoskeletons can reduce muscular effort during occupational tasks such as overhead drilling, assembly, lifting, and stacking [McFarland and Fischer, 2019, de Vries and de Looze, 2019]. However, antagonistic muscular effort can increase [de Vries and de Looze, 2019], and the impact on kinematics and fatigue have not yet been thoroughly investigated [McFarland and Fischer, 2019].

Several semi-passive, clutch-based systems [Naito et al., 2007, Diller et al., 2016, Ramachandran et al., , Grazi et al., 2020] and numerous powered systems [Aida et al., 2009, Muramatsu et al., 2011, Gopura et al., 2016, Otten et al., 2018] have been developed in recent years. However, so far there is limited evidence to support the use of active upper limb exoskeletons.

Semi-passive and active exoskeletons are generally heavier than their passive counterparts [Maurice et al., 2019]. Similar to lower limb exoskeletons, the system weight and inertia will increase overall human effort, which works against the goal of effort reduction. Therefore, lightweight exosuits have been investigated [Gaponov et al., 2017, Xiloyannis et al., 2017, Wei et al., 2018, Kim et al., 2018, Lessard et al., 2018] and have demonstrated reductions in muscle activity, biological elbow moment, and the onset of fatigue [Xiloyannis et al., 2017, Xiloyannis et al., 2019].

While exosuits for the lower limb typically use motors for actuation, pneumatic actuation has also been investigated to assist the upper limb [Gopura et al., 2016, Thalman and Artemiadis, 2020]. Pneumatic actuation allows for a soft exoskeleton design with soft textile- based pneumatic actuators [Thalman and Artemiadis, 2020, Nassour et al., 2020] and a soft human-machine interface. Such designs were developed to assist the shoulder [O'Neill et al., 2017, Simpson et al., 2020], elbow [Koh et al., 2017, Thalman et al., 2018, Irshaidat et al., 2019], wrist [Sasaki et al., 2005, Al-Fahaam et al., 2016], and fingers [Meng et al., 2017, Shahid et al., 2018], and reductions in elbow flexor activity have been demonstrated for one subject [Thalman et al., 2018].

1.2 Exoskeleton control

A variety of control concepts have been explored for wearable robotics to assist users in diverse terrains [Tucker et al., 2015, Pinto-Fernandez et al., 2020, Baud et al., 2021]. Control concepts consist of high-level, mid-level, and low-level control. High-level control determines the general behavior of the exoskeleton and typically involves multiple operating modes that cover desired activities and environments (e.g. walking on flat terrain, climbing stairs, and sit-to-stand transitions). Mid-level control defines the target torque or position at each timestep of the main control loop based on the assistance strategy. The assistance strategy determines the physical output of the exoskeleton (e.g. amplitude and timing) to achieve the desired assistance. Mid-level control is also used to detect the gait phase or state to provide the desired output, and it therefore has a key role in shaping the interaction between the device and the user. Low-level control aims to accurately track a reference input while remaining stable [Baud et al., 2021]. With high-level control the general behavior of the system is specific to the activity of interest. To determine the desired activity, researchers rely on machine learning-based approaches [Labarrière et al., 2020] or artificial intelligence [Baud et al., 2021]. While the easiest approach would be to rely on manual user input, this less intuitive approach is generally not desirable. To avoid this, simple movement recognition algorithms have been developed that use heuristic-based classifiers or pattern recognition strategies to detect the relevant gait mode [Tucker et al., 2015]. Other systems rely on terrain recognition. These algorithms are often fed by worn sensors, such as surface EMG, force and position sensors [Tucker et al., 2015] or brain-computer interfaces [Baud et al., 2021].

Mid-level control allows for continuous gait phase estimation by using phase plane approaches [Holgate et al., 2009, Quintero et al., 2017] and machine learning regression [Seo et al., 2019, Kang et al., 2019, Weigand et al., 2020]. Many approaches distinguish between a detection sublayer and an action sublayer. The detection sublayer provides the current gait state while the action sublayer provides a motor command [Baud et al., 2021]. In [Seo et al., 2019] recurrent neural networks (RNNs) with long short-term memory nodes (LSTMs) were used in combination with a shank-mounted

inertial measurement unit (IMU) to estimate the continuous gait phase for level walking. In [Kang et al., 2019] artificial neural networks (ANNs) were used to estimate walking speed and the continuous gait phase during level walking with a robotic hip exoskeleton. For generating the motor command, the action sublayer often uses position or torque profiles, impedance control, direct joint torque estimation, balance models, or neuromuscular models [Baud et al., 2021].
Low-level control has used a variety of torque-based, position-based, and hybrid approaches [Zhang et al., 2015, Baud et al., 2021, Tucker et al., 2015]. Position-based control is preferably used to provide full lower limb mobilization whereas torque-based control is preferably used for partial mobilization [Baud et al., 2021].
Often, the timings and amplitudes of the assistance strategy have been inspired by human biomechanics, such as joint moments and joint power trajectories [Quinlivan et al., 2017, Schmidt et al., 2017, Grimmer et al., 2019a]. While worldwide efforts have led to significant improvements in walking assistance, nearly all exoskeletons still only provide minor benefits for the user [Sawicki et al., 2020]. One major reason for such limited performance benefit is that conventional assistance control concepts are based on a certain generalized logic, and even when hand-tuned by an expert, most assistance strategies will have sub-optimal assistance timings and amplitudes. Additionally, even if the assistance strategy is optimal for one user, other users will respond differently as seen by the large variation in assistance when using the same strategy for several subjects. While a mean reduction of about 23% in net metabolic cost during walking has been shown, minimum and maximum reductions were 15% and 36%, respectively, when using the Harvard exosuit compared to unassisted walking [Quinlivan et al., 2017]. Similar results have been seen with other exoskeletons and assistance strategies as well [Young et al., 2017]. Hence, there exists a need for individualized optimization of the control strategies to increase the level of support for each user. Human- in-the-loop (HITL) optimization is a potential solution to achieve individualized assistance [Koller et al., 2016, Zhang and Arakelian, 2018, Franks et al., 2021]. This optimization approach uses performance measures such as user effort or gait symmetry to iteratively optimize parameters in a feedback loop that specify the type of assistance throughout the stride. Thus, optimal assistance can be provided for each individual user independent of the biological (e.g. muscle and tendon properties) and neurological (e.g. existing motor patterns) conditions.

1.3 Development needs

The state of the art of exoskeleton technology demonstrates that minimalist and lightweight concepts are most promising for providing assistance to people with at least partial movement capabilities of the lower and upper limbs. While several lower limb exoskeletons have demonstrated user benefits, benefits from autonomous exoskeletons are still limited and far from producing results that have been achieved with experimental tethered setups. For the upper limb, passive exoskeletons have proven to provide user benefits for overhead work. However, active upper limb exoskeleton concepts struggle to be lightweight, comfortable, free from movement restrictions, and to demonstrate local muscle and overall (e.g. metabolic cost) benefits for lifting, holding, and carrying objects. To increase benefits for users with mobility needs for specific movement tasks, the exoskeleton hardware and control could be improved and individualized.

2. Exoskeletons at TU Darmstadt

This work reviews the efforts at the Technical University of Darmstadt (TU Darmstadt) through 2022 in regards to the field of lower and upper limb exoskeletons. As the extent and quality of the different developments are diverse, this review focuses on summarizing each larger development in terms of the motivation, the approach, and the study outcomes and insights. Funding and literature sources are provided if available. Compared to classical reviews where literature is scanned based on keywords and inclusion and exclusion criteria are used to select relevant works, this review solely relies on internal information from the wearable robotics community at TU Darmstadt for selecting specific works.

2.1 Lower limb

2.1.1 Powered Ankle Knee Orthosis (PAKO) to explore ankle stiffness

When attempting to mimic or assist human gait with lower limb wearable robotics, one limiting factor is the power source. Exoskeletons and prostheses primarily use motors to assist the user. Human joint movements in daily activities require high accelerations, velocities, and moments [Grimmer et al., 2020a]. However, motors struggle to achieve movement characteristics similar to our biological actuation. Biological actuation results from an interaction between muscles and tendons, where motors and springs, respectively, could be seen as mechanical counterparts. Tendons are used to store energy during the stance phase to release it during the push-off. To minimize motor requirements in specific gait modes, human gait biomechanics were used to analyze the theoretical benefits of springs when used in elastic actuators. In this project funded by the German Research Foundation DFG, the Powered Ankle Knee Orthosis (PAKO) (see figure 1, a, [Eslamy, 2014]) and a powered prosthetic foot

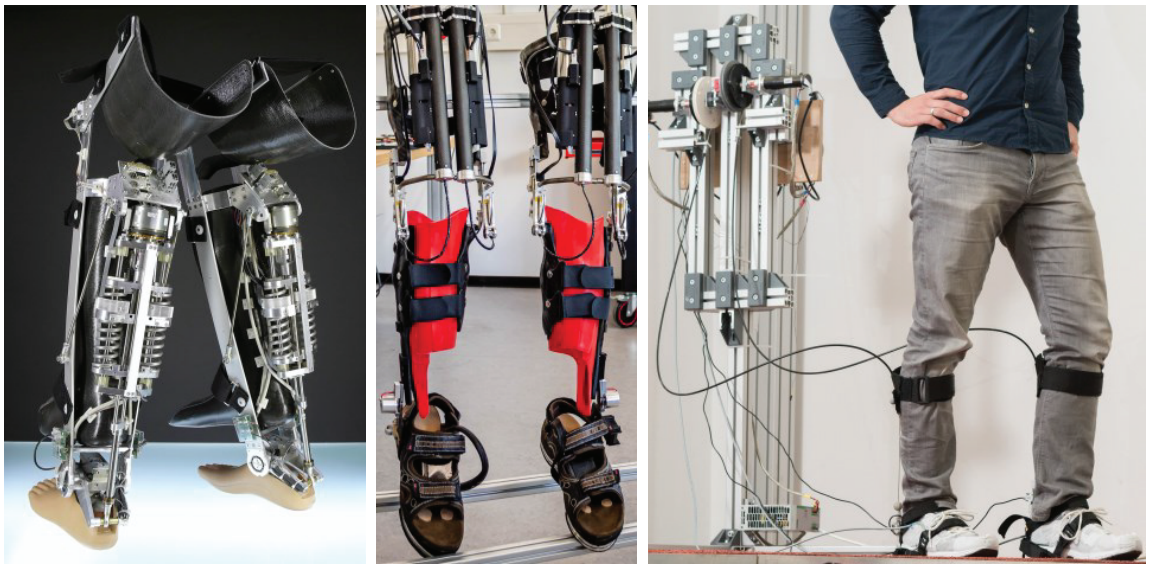
(Walk-Run Ankle) were used for practical evaluation of the elastic actuation benefits. PAKO is a hardware-based simulator that replaces both the user's ankles and feet with mechanical components. PAKO allows for changing the spring stiffness of a series elastic actuator, which mimics the function of the calf muscles and the Achilles tendon of both legs by changing the free length of coil springs. An inertial measurement unit is used to provide information to control the movement of both artificial feet during gait. Based on theoretical results [Grimmer et al., 2014], evaluations of PAKO [Eslamy, 2014], and evaluations of the Walk-Run Ankle [Grimmer et al., 2016] it was found that when mimicking the ankle during walking and running, the required actuator peak power and energy can be greatly reduced when assisted with elasticity.

2.1.2 Rigid exoskeleton to assist the hip and the knee against gravity

During aging, the maximal strength of the knee extensor muscles decreases. Thus, daily tasks such as sit-to-stand transitions or stair ambulation could benefit from knee extensor assistance provided by a knee exoskeleton. To provide appropriate support, a control algorithm is necessary to scale assistance based on the load on the knee joint. In collaboration with members of Heidelberg University, a powered knee exoskeleton (see figure 1, b) was developed (DFG-funded project) that provides assistance for knee extension (up to 30 Nm) to relieve biomechanical loads and to reduce muscular effort during sit-to-stand transitions, which require about 60 Nm for a 75 kg person [Grimmer et al., 2020a]. In order to scale assistance, foot plates with force sensors were developed and used to measure ground reaction forces, which were used in combination with leg kinematics to control knee torque through an inverse dynamics model [Pott et al., 2017]. This knee exoskeleton was tested in the sit-to-stand transition and provided an assistive knee torque of 23% of the total required knee moment. With such assistance, the muscle activity of the rectus femoris, a knee extensor, decreased by 26% [Pott et al., 2017].

2.1.3 Soft exoskeleton to assist and perturb the ankle

Artificially-induced perturbations during human movements (or standing) are used to study human motor control [Tokur et al., 2020]. Perturbations are primarily applied by changing the environment (e.g. terrain slope or obstacles) or by pulling or pushing subjects via external forces. Exoskeletons allow for both direct assistance and perturbation of human joints. A variety of passive and active tethered soft exoskeletons were developed (unfunded work) to study human-exoskeleton interaction, and this includes an exoskeleton to assist the ankle (see figure 1, c, [Schumacher et al., 2018]). This exoskeleton uses two RE–30 60 W DC motors to apply forces of up to 0.1 N/BW via Bowden cables to assist with plantarflexion. Forces were applied to anchor points located above the calf muscles and the foot. The ankle exoskeleton aimed to explore the human response and recovery strategies following temporary perturbations at the ankle joint [Schumacher et al., 2018]. With the help of the tethered ankle exoskeleton, the ankle was perturbed with a plantarflexion torque during the second half of stance during hopping. The performance of the control approach and the impact on the user



(a) (b) (c)

Figure 1. (a) Tethered Powered Ankle Knee Orthosis (PAKO). (b) Tethered knee exoskeleton. (c) Tethered ankle exosuit.

were evaluated. Even with a small perturbation a significant increase in ankle plantarflexion was found during take-off of the perturbed hop. Following an imposed perturbation, joint kinematics were compensated for during the flight phase and joint kinetics were compensated for during the following stance phase. In general, this first powered design also showed that higher-powered actuators are required to increase the potential for manipulating human movements.

2.2 Soft exoskeleton to assist hip flexion and extension

Following the developments of the ankle exoskeleton, a tethered soft exosuit was developed to assist or perturb users at the hip (unfunded work, see figure 2, a, [Bobzien, 2018]). While the assistance is useful for people with mobility needs, perturbations can be used for training a variety of physical capacities or to explore human motor control. Four brushless direct current (BLDC) motors (HYMOTOR E8318 120KV) were used to assist hip flexion and extension via Bowden cables at both limbs. Forces were applied to anchor points located at the waist and at the lower shank. Force sensors were used to measure applied forces for each of the four actuators. Inertial measurement units were used to measure the user's absolute limb orientations to control the timing of assistance based on a virtual leg approach [Grimmer et al., 2019c]. Additionally, the exoskeleton electronics were designed to measure heart rate and muscle activity of up to 16 muscles in real time. Bench testing and level walking experiments were conducted to evaluate exoskeleton performance and control with a predefined assistance trajectory based on biological hip moments. Results showed that each tethered actuator could deliver an effective power of at least 50 W and a maximum applied force of 350 N (35 Nm using a 0.1 m lever). For comparison, a 75 kg person requires about 90 W and 90 Nm at the hip for walking at 1.6 m/s [Grimmer et al., 2020b].
The research team (Zhao, Grimmer, and students) realized a couple of disadvantages of this design that could help guide future developments. Donning and adapting the exoskeleton to the user were found to be quite time-consuming. As well, Bowden cables and the human- machine interface were found to interfere with EMG sensor fixation.

2.3 Exoskeleton to assist hip flexion and extension

With internal funding from TU Darmstadt (Athene Young Investigator grant of Grimmer), a hip exoskeleton (see figure 2, b) was developed to minimize donning and doffing effort and to keep the human-machine interface as minimalistic as possible in order to affix EMG sensors to different locations of the limb. The EMG sensors are intended to be used as feedback to investigate the individualization of exoskeleton assistance by using human- in-the-loop optimization to assist users with mobility needs [Grimmer et al., 2022]. The hip exoskeleton has a weight of 3.1 kg and is able to provide up to 50 Nm and 1.1 kW at each hip joint to assist or perturb hip flexion and extension. The initial unpublished results, which were acquired in the interdisciplinary teaching project 'Analysis and synthesis of human movements' (ANSYMB), demonstrated that the exoskeleton is able to assist and perturb human walking with different torque amplitudes and timings. With one of the applied assistance torque patterns, on average, a group of subjects showed net metabolic reductions of about

(a) (b) (c)

Figure 2. (a) Tethered hip exosuit. (b) Autonomous hip exoskeleton. (c) Orthoprosthesis to mimic the ankle with mono- and biarticular actuation.

10%. Also, when providing hip torque that counteracts human movements, it is possible to increase walking effort beyond subject limits. Further, in a second series of experiments on individualized assistance based on EMG, we were able to identify torque patterns with similar net metabolic reductions as found in ANSYMB automatically. Future studies could use this hip exoskeleton to perform experiments on assistance, perturbations, and training of user groups with related needs.

2.4 Exoskeleton-prosthesis combination to mimic the ankle

Lower limb prosthetic feet should enable standing and locomotion. However, passive prosthetic feet lack the human range of motion and power output [Grimmer and Seyfarth, 2014]. Active prosthetic feet were developed to overcome these deficiencies [Grimmer et al., 2016] and demonstrated that they can mimic healthy human sagittal ankle kinematics and kinetics. However, walking asymmetries and compensatory strategies of users remain [Ferris et al., 2012, Grimmer et al., 2017]. Researchers assume one reason to be the inability of users to adapt their lower limb motor pattern. Another reason could be the absence of mechanical structures that exist in the biological limb. For example, the gastrocnemius muscle is currently not represented in active prosthetic feet though it could be mimicked by a biarticular actuator. To investigate the benefits of mono- and biarticular actuators, the functionality of a powered prosthetic ankle with a monoarticular actuator was extended in a DFG-funded project to include a biarticular tethered exoskeleton-like structure to assist with ankle plantarflexion (see figure 2, c). Such concepts are also referred to as orthoprosthesis. An initial walking experiment demonstrated the ability of the orthoprosthesis to mimic human ankle torque and angle using a variable distribution of torque between the mono- and biarticular actuator. Ankle torque provided by the biarticular actuator was able to reach up to 55 Nm, which is about 50% of the total ankle torque provided by the orthoprosthesis when walking at 1 m/s [Zeiss et al., 2020].

2.5 Multiarticular exoskeleton to assist the lower limb

Passive and active exoskeleton designs have used a combination of biologically-representative biarticular and compliant interconnections [Sharbafi, 2020, Barazesh and Sharbafi, 2020]. The potential of these designs for locomotion assistance has been investigated through works that were partially funded by the DFG and the Iran National Science Foundation (INSF). Experiments with passive exoskeletons using biarticular thigh springs that cross the hip and the knee joint at each leg have demonstrated reduced metabolic effort during walking [Barazesh and Sharbafi, 2020]. In addition to the biologically-inspired design, the development of multiarticular active exoskeletons is also aimed at biologically-inspired control concepts that include locomotor subfunctions [Sharbafi et al., 2018] and the force- modulated compliant hip (FMCH, [Sharbafi et al., 2017]. In FMCH, the ground reaction force (GRF) is utilized to adjust hip compliance. This GRF-based control concept was also implemented in the exoskeleton LOPES II in a collaboration with the University of Twente [Zhao et al., 2017, Zhao et al., 2019]. Experiments with LOPES II revealed reductions in the user's metabolic cost of walking and in muscular activity. Further, data-driven (experiment- based) simulations using neuromuscular (OpenSim) modeling were performed, including a model of the active biarticular thigh exosuit BATEX, to support walking in the sagittal and frontal planes [Firouzi et al., 2022]. Based on the simulation outcomes, preliminary real-world tests with the BATEX exoskeleton support the applicability of this control concept [Firouzi et al., 2021]. Currently, a second generation of the BATEX exosuit is being developed that uses more advanced sensors and actuators and includes the ability to easily change the actuation from monoarticular to biarticular and from an active to a passive mode (see figure 3).

3. Upper limb

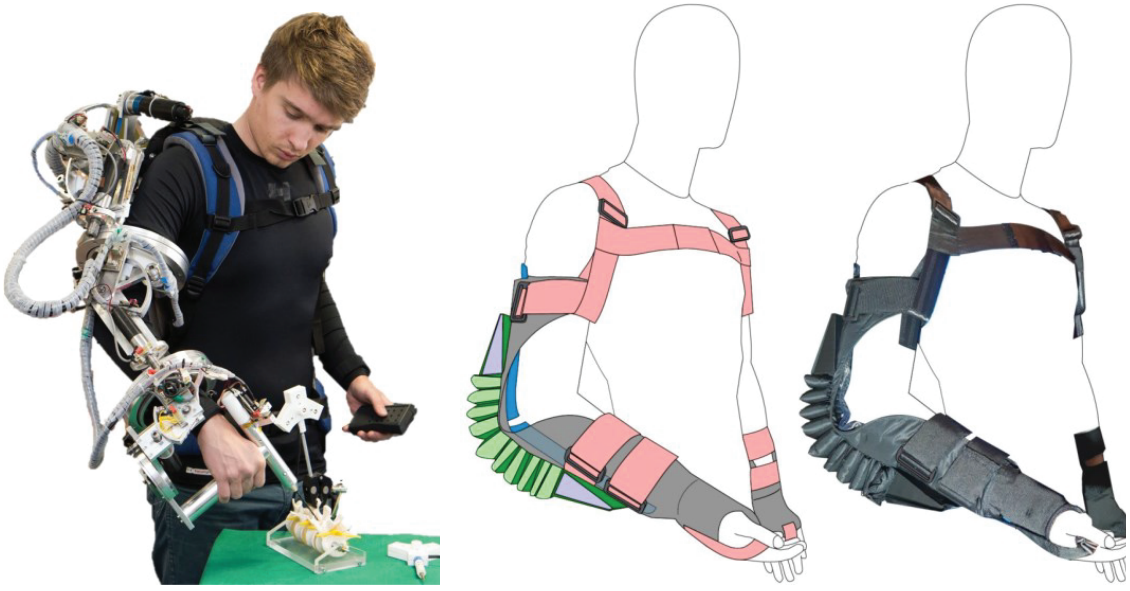
Exoskeleton to assist drilling

To enhance the accuracy of the target position and constant thrust force during drilling tasks [Hessinger et al., 2017], an upper limb exoskeleton that provides haptic guidance was developed [Hessinger et al., 2018] with funding from the German Federal Ministry of Education and Research (see figure 4). This exoskeleton is able to move in seven degrees of freedom and allows for active assistance at the shoulder, elbow, and wrist using BLDC motors. Torque sensors were used to detect user intention. Preliminary evaluations investigated the use of an implicit, hybrid force-position controller to enable position-controlled drilling without a user. A motion capture system was used to control the absolute position of the end effector in operational space. Results showed that the exoskeleton can drill 20 mm deep holes in 4 s with a constant thrust force of 4 N and with a maximum position error of 1.3 mm.

3.1.1 Soft exoskeleton to assist the elbow while holding and carrying weights

The ability to hold and carry objects is limited due to one's maximum muscle force and accumulating fatigue. In addition, fatigue increases the risk of accidents, joint degeneration, injuries, and pain. To sustain workloads and to reduce the risk of all of these potentially detrimental effects, the exoskeleton Carry was developed [Nassour et al., 2021] in cooperation with TU Chemnitz. Funding was partially provided by the Athene Young Investigator grant of Grimmer. Carry has a soft

Figure 3. Biarticular thigh exosuit BATEX.



(a) (b)

Figure 4. (a) Shoulder-Elbow-Wrist exoskeleton to assist drilling. (b) Elbow exoskeleton Carry that assists with carrying and holding of loads.

human-machine interface and uses soft pneumatic actuators to assist with elbow flexion in lifting, holding, or carrying scenarios. Initial experiments evaluated holding and carrying tasks with a static elbow position of about 90°. Using an elbow flexion torque of 7.2 Nm, Carry reduced the muscle activity of arm and shoulder muscles by up to 50% (biceps brachii, brachioradialis, flexor carpi radialis, and trapezius) for a group of individuals holding a 5 kg weight. Additionally, Carry was able to decrease the net metabolic rate for holding and carrying a 5 kg weight by 61% and 32%, respectively. Fatigue, as determined by maximum voluntary contractions before and after the loaded sessions, decreased by 99% for holding the weight and by 80% for carrying the weight. It is expected that increased assistance torque will further increase user benefits. While the system that was evaluated was tethered and only evaluated in static scenarios, the analyses of operation dynamics and autonomous use demonstrated the utility of Carry for common and dynamic scenarios. Results show that Carry can reduce systemic aerobic load and local muscle fatigue, which potentially reduces a variety of risks associated with lifting, holding, or carrying.

4. Summary and conclusions

This review demonstrates that researchers at TU Darmstadt have made significant contributions to the design and control of upper and lower limb exoskeletons. While some projects were accomplished solely by members of TU Darmstadt, collaborative efforts include members of the Heidelberg University, the TU Chemnitz, the University of Twente, and the University of Tehran. Several Bachelor, Master, and PhD students were trained on the related topics in the interdisciplinary field of wearable robotics. A variety of technological concepts have been explored. Stiff and soft materials have been used for the user-machine interface. Motors and pneumatic actuation have been used to provide the required torque and power. A variety of sensors have been used for control, including motor and joint encoders, force sensors, inertial measurement units, motion capture, EMG, and metabolic measurements. High-, mid-, and low-level control have also been investigated. Assistance tasks of interest for the lower limb have included walking, hopping, and sit-to-stand transitions, and tasks of interest for the upper limb have included holding and carrying of loads and drilling.
In summary, several exoskeleton-related developments have been performed at TU Darmstadt. The research aims of the projects described align with the global efforts to improve the effectiveness of upper and lower limb exoskeletons for a wide range of target populations with mobility needs. Some of the research directions investigated have been quite unique and some have demonstrated significant user benefits. A number of works have proven to be fundamental and are the foundation to future exoskeleton research and development at TU Darmstadt.

References

Aach, M., Cruciger, O., Sczesny-Kaiser, M., Höffken, O., Meindl, R. C., Tegenthoff, M., Schwenkreis, P., Sankai, Y., and Schildhauer, T. A. (2014). Voluntary driven exoskeleton as a new tool for rehabilitation in chronic spinal cord injury–a pilot study. *The Spine Journal*.

Aida, T., Nozaki, H., and Kobayashi, H. (2009). Development of muscle suit and application to factory laborers. In *2009 International Conference on Mechatronics and Automation (ICMA)*, pages 1027–1032. IEEE.

Al-Fahaam, H., Davis, S., and Nefti-Meziani, S. (2016). Wrist rehabilitation exoskeleton robot based on pneumatic soft actuators. In *2016 International Conference for Students on Applied Engineering (ICSAE)*, pages 491–496. IEEE.

Alnahdi, A. H., Zeni, J. A., and Snyder-Mackler, L. (2012). Muscle impairments in patients with knee osteoarthritis. *Sports Health*, 4(4):284–292.

Bae, J., De Rossi, S. M. M., O'Donnell, K., et al. (2015). A soft exosuit for patients with stroke: Feasibility study with a mobile off-board actuation unit. In *International Conference on Rehabilitation Robotics (ICORR)*, pages 131–138. IEEE.

Barazesh, H. and Sharbafi, M. A. (2020). A biarticular passive exosuit to support balance control can reduce metabolic cost of walking. *Bioinspi- ration & Biomimetics*, 15(3):036009.

Baud, R., Manzoori, A. R., Ijspeert, A., and Bouri, M. (2021). Review of control strategies for lower-limb exoskeletons to assist gait. *Journal of NeuroEngineering and Rehabilitation*, 18(1):1–34.

Berkeley Bionics (2014). Hulc. http://bleex.me.berkeley.edu/research/exoskeleton/hulc/. June,.

Bobzien, L. (2018). Design and implementation of soft exosuit for hip assistance. Master's thesis, Technical University Darmstadt.

Bravata, D. M., Smith-Spangler, C., Sundaram, V., Gienger, A. L., Lin, N., Lewis, R., Stave, C. D., Olkin, I., and Sirard, J. R. (2007). Using pedometers to increase physical activity and improve health: a systematic review. *JAMA*, 298(19):2296–2304.

Browning, R. C., Modica, J. R., Kram, R., and Goswami, A. (2007). The effects of adding mass to the legs on the energetics and biomechanics of walking. *Med Sci Sports Exerc*, 39(3):515–525.

Bryan, G. M., Franks, P. W., Klein, S. C., Peuchen, R. J., and Collins, S. H. (2021). A hip–knee–ankle exoskeleton emulator for studying gait assistance. *The International Journal of Robotics Research*, 40 (4-5):722–746.

Buesing, C., Fisch, G., ODonnell, M., Shahidi, I., et al. (2015). Effects of a wearable exoskeleton stride management assist system (sma®) on spatiotemporal gait characteristics in individuals after stroke: a randomized controlled trial. *J NeuroEng Rehabil*, 12(1):69.

de Vries, A. and de Looze, M. (2019). The effect of arm support exoskeletons in realistic work activities: A review study. *J Ergonomics*, 9:255.

DESA, U. (2012). World population prospects: the 2012 revision. *New York: Department for Economic and Social Affairs*.

Diller, S., Majidi, C., and Collins, S. H. (2016). A lightweight, lowpower electroadhesive clutch and spring for exoskeleton actuation. In *2016 International Conference on Robotics and Automation (ICRA)*, pages 682–689. IEEE.

Erlichman, J., Kerbey, A., and James, W. (2002). Physical activity and its impact on health outcomes. paper 1: the impact of physical activity on cardiovas- cular disease and all-cause mortality: an historical perspective. *Obes Rev*, 3(4): 257–271.

Eslamy, M. (2014). *Emulation of ankle function for different gaits through active foot prosthesis: Actuation concepts, control and experiments*. PhD thesis, Universitäts-und Landesbibliothek Darmstadt.

Esquenazi, A., Talaty, M., Packel, A., and Saulino, M. (2012). The rewalk powered exoskeleton to restore ambulatory function to individuals with thoracic- level motor-complete spinal cord injury. *Am J Phys Med Rehabil*, 91(11):911–921.

Federici, S., Meloni, F., Bracalenti, M., and De Filippis, M. L. (2015). The effectiveness of powered, active lower limb exoskeletons in neurorehabilitation: A systematic review. *NeuroRehabilitation*, 37(3):321–340.

Ferris, A. E., Aldridge, J. M., Rábago, C. A., and Wilken, J. M. (2012). Evaluation of a powered ankle-foot prosthetic system during walking. *Archives of Physical Medicine and Rehabilitation*, 93(11):1911–1918.

Ferris, D. P. (2009). The exoskeletons are here. *J Neuroeng Rehabil*, 6(1):17.

Firouzi, V., Davoodi, A., Bahrami, F., and Sharbafi, M. A. (2021). From a biological template model to gait assistance with an exosuit. *Bioinspiration & Biomimetics*, 16(6):066024.

Firouzi, V., O, M., and Sharbafi, M. A. (2022). Model-based control for gait assistance in the frontal plane. *RAS/EMBS IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)*.

Franks, P. W., Bryan, G. M., Martin, R. M., Reyes, R., Lakmazaheri, A. C., and Collins, S. H. (2021). Comparing optimized exoskeleton assistance of the hip, knee, and ankle in single and multi-joint configurations. *Wearable Technologies*, 2.

Fried, L. P., Darer, J., and Walston, J. (2003). Frailty. In *Geriatric Medicine*, pages 1067–1076. Springer.

Galle, S., Derave, W., Bossuyt, F., Calders, P., Malcolm, P., and De Clercq, D. (2017). Exoskeleton plantarflexion assistance for elderly. *Gait & Posture*, 52:183–188.

Gaponov, I., Popov, D., Lee, S. J., and Ryu, J.-H. (2017). Auxilio: a portable cable-driven exosuit for upper extremity assistance. *Int J Control Autom Syst*, 15(1):73–84.

Gopura, R., Bandara, D., Kiguchi, K., and Mann, G. K. (2016). Developments in hardware systems of active upperlimb exoskeleton robots: A review. *Rob Auton Syst*, 75:203–220.

Grazi, L., Trigili, E., Proface, G., Giovacchini, F., Crea, S., and Vitiello, N. (2020). Design and experimental evaluation of a semi-passive upper-limb exoskeleton for workers with motorized tuning of assistance. *IEEE Trans Neural Syst Rehabil Eng*.

Grimmer, M., Elshamanhory, A. A., and Beckerle, P. (2020a). Human lower limb joint biomechanics in daily life activities: a literature based requirement analysis for anthropomorphic robot design. *Front Robot AI*.

Grimmer, M., Eslamy, M., and Seyfarth, A. (2014). Energetic and peak power advantages of series elastic actuators in an actuated prosthetic leg for walking and running. *Actuators*, 3(1):1–19.

Grimmer, M., Holgate, M., Holgate, R., Boehler, A., Ward, J., Hollander, K., Sugar, T., and Seyfarth, A. (2016). A powered prosthetic ankle joint for walking and running. *Biomedical engineering online*, 15(3):37–52.

Grimmer, M., Holgate, M., Ward, J., Boehler, A., and Seyfarth, A. (2017). Feasibility study of transtibial amputee walking using a powered prosthetic foot. In *2017 International Conference on Rehabilitation Robotics (ICORR)*, pages 1118–1123. IEEE.

Grimmer, M., Quinlivan, B. T., Lee, S., Malcolm, P., Rossi, D. M., Siviy, C., and Walsh, C. J. (2019a). Comparison of the human-exosuit interaction using ankle moment and ankle positive power inspired walking assistance. *J Biomech*, 83:76–84.

Grimmer, M., Riener, R., Walsh, C. J., and Seyfarth, A. (2019b). Mobility related physical and functional losses due to aging and disease-a motivation for lower limb exoskeletons. *J NeuroEng Rehabil*, 16(1):2.

Grimmer, M., Schmidt, K., Duarte, J. E., Neuner, L., Koginov, G., and Riener, R. (2019c). Stance and swing detection based on the angular velocity of lower limb segments during walking. *Frontiers in Neurorobotics*, page 57.

Grimmer, M. and Seyfarth, A. (2014). Mimicking human- like leg function in prosthetic limbs. In *Neuro-Robotics*, pages 105–155. Springer.

Grimmer, M., Zeiss, J., Weigand, F., and Zhao, G. (2022). Exploring surface electromyography (emg) as a feedback variable for the human-in-the-loop optimization of lower limb wearable robotics. *Frontiers in Neurorobotics*, 16:948093.

Grimmer, M., Zeiss, J., Weigand, F., Zhao, G., Lamm, S., Steil, M., and Heller, A. (2020b). Lower limb joint biomechanics-based identification of gait transitions in between level walking and stair ambulation. *PLoS ONE*, 15(9):e0239148.

Hessinger, M., Christmann, E., Werthschützky, R., and Kupnik, M. (2018). Messung von Nutzerinteraktion mit einem Exoskelett durch EMG und Gelenk- Drehmomente. *tm-Technisches Messen*, 85(7-8):487–495.

Hessinger, M., Pingsmann, M., Perry, J. C., Werthschützky, R., and Kupnik, M. (2017). Hybrid position/force control of an upper-limb

exoskeleton for assisted drilling. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1824–1829. IEEE.

Holgate, M. A., Sugar, T. G., and Bohler, A. W. (2009). A novel control algorithm for wearable robotics using phase plane invariants. In *2009 International Conference on Robotics and Automation*, pages 3845–3850. IEEE.

Irshaidat, M., Soufian, M., Al-Ibadi, A., and Nefti-Meziani, S. (2019). A novel elbow pneumatic muscle actuator for exoskeleton arm in post-stroke rehabilitation. In *2019 2nd International Conference on Soft Robotics (RoboSoft)*, pages 630–635. IEEE.

Jezernik, S., Colombo, G., Keller, T., Frueh, H., and Morari, M. (2003). Robotic orthosis lokomat: A rehabilitation and research tool. *Neuromodulation: Technology at the neural interface*, 6(2):108–115.

Kang, I., Kunapuli, P., and Young, A. J. (2019). Real-time neural network-based gait phase estimation using a robotic hip exoskeleton. *IEEE Transactions on Medical Robotics and Bionics*, 2(1):28–37.

Kazerooni, H., Racine, J.-L., Huang, L., and Steger, R. (2005). On the control of the berkeley lower extremity exoskeleton (bleex). In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 4353–4360. IEEE.

Kelley, L. C. (U.S. Patent US1308675A, Jul. 1919). Pedomotor.

Kilicarslan, A., Prasad, S., Grossman, R. G., and Contreras-Vidal, J. L. (2013). High accuracy decoding of user intentions using eeg to control a lower-body exoskeleton. In *35th EMBC*, pages 5606–5609. IEEE.

Kim, Y. G., Xiloyannis, M., Accoto, D., and Masia, L. (2018). Development of a soft exosuit for industrial applications. In *2018 7th International Conference on Biomedical Robotics and Biomechatronics (Biorob)*, pages 324–329. IEEE.

Koh, T. H., Cheng, N., Yap, H. K., and Yeow, C.-H. (2017). Design of a soft robotic elbow sleeve with passive and intent-controlled actuation. *Front Neurosci*, 11:597.

Koller, J. R., Gates, D. H., Ferris, D. P., and Remy, C. D. (2016). 'body-in-the-loop'optimization of assistive robotic devices: A validation study. In *Robotics: Science and Systems*, volume 2016, pages 1–10.

Labarrière, F., Thomas, E., Calistri, L., Optasanu, V., Gueugnon, M., Ornetti, P., and Laroche, D. (2020). Machine learning approaches for activity recognition and/or activity prediction in locomotion assistive devices—a systematic review. *Sensors*, 20(21):6345.

Lee, J., Seo, K., Lim, B., Jang, J., Kim, K., and Choi, H. (2017). Effects of assistance timing on metabolic cost, assistance power, and gait parameters for a hip-type exoskeleton. In *2017 International Conference on Rehabilitation Robotics (ICORR)*, pages 498–504. IEEE.

Lessard, S., Pansodtee, P., Robbins, A., Trombadore, J. M., Kurniawan, S., and Teodorescu, M. (2018). A soft exosuit for flexible upper-extremity rehabilitation. *IEEE Trans Neural Syst Rehabil Eng*, 26(8):1604–1617.

Makinson, B. J. (1971). Research and development prototype for machine augmentation of human strength and endurance. hardiman project. Technical report, General Electric CO Schenectady NY Speciality Materials Handling Products . . . .

Martini, E., Crea, S., Parri, A., Bastiani, L., Faraguna, U., McKinney, Z., Molino-Lova, R., Pratali, L., and Vitiello, N. (2019). Gait training using a robotic hip exoskeleton improves metabolic gait efficiency in the elderly. *Scientific reports*, 9(1):1–12.

Maurice, P., Cˇ amernik, J., Gorjan, D., Schirrmeister, B., Bornmann, J., Tagliapietra, L., Latella, C., Pucci, D., Fritzsche, L., Ivaldi, S., et al. (2019). Objective and subjective effects of a passive exoskeleton on overhead work. *IEEE Trans Neural Syst Rehabil Eng*.

McDermott, M. M., Criqui, M. H., Greenland, P., Guralnik, J. M., Liu, K., Pearce, W. H., Taylor, L., Chan, C., Celic, L., Woolley, C., O'Brien, M. P., and Schneiderothers, J. R. (2004). Leg strength in peripheral arterial disease: associations with disease severity and lower-extremity performance. *J Vasc Surg*, 39(3):523–530.

McFarland, T. and Fischer, S. (2019). Considerations for industrial use: A systematic review of the impact of active and passive upper limb exoskeletons on physical exposures. *IISE Trans Occup Ergon Hum Factors*, pages 1–26.

Meng, Q., Xiang, S., and Yu, H. (2017). Soft robotic hand exoskeleton systems: Review and challenges surrounding the technology. In *2017 2nd International Conference on Electrical, Automation and Mechanical Engineering (EAME)*. Atlantis Press.

Monaco, V., Tropea, P., Aprigliano, F., Martelli, D., Parri, A., Cortese, M., Molino-Lova, R., Vitiello, N., and Micera, S. (2017). An ecologically-controlled exoskeleton can improve balance recovery after slippage. *Scientific Reports*, 7:46721.

Moreau, K. L., Degarmo, R., Langley, J., McMahon, C., Howley, E. T., Bassett, D. R., and Thompson, D. L. (2001). Increasing daily walking lowers blood pressure in postmenopausal women. *Med Sci Sports Exerc*, 33(11):1825–1831.

Mosher, R. S. (1968). Handyman to hardiman. *Sae Transactions*, pages 588–597.

Muramatsu, Y., Kobayashi, H., Sato, Y., Jiaou, H., Hashimoto, T., and Kobayashi, H. (2011). Quantitative performance analysis of exoskeleton augmenting devices-muscle suit-for manual worker. *Int J Autom Technol*, 5(4):559–567.

Murtagh, E. M., Murphy, M. H., and Boone-Heinonen, J. (2010). Walking–the first steps in cardiovascular disease prevention. *Curr Opin Cardiol*, 25(5):490.

Naito, J., Obinata, G., Nakayama, A., and Hase, K. (2007). Development of a wearable robot for assisting carpentry workers. *Int J Adv Robot Syst*, 4(4):48.

Nam, K. Y., Kim, H. J., Kwon, B. S., Park, J.-W., Lee, H. J., and Yoo, A. (2017). Robot-assisted gait training (lokomat) improves walking function and activity in people with spinal cord injury: a systematic review. *Journal of Neuroengineering and Rehabilitation*, 14(1):1–13.

Nassour, J., Hamker, F. H., and Cheng, G. (2020). High-performance perpendicularly-enfolded-textile actuators for soft wearable ro-

bots: Design and realization. *IEEE Trans Med Robot Bionics*, 2(3): 309–319.

Nassour, J., Zhao, G., and Grimmer, M. (2021). Soft pneumatic elbow exoskeleton reduces the muscle activity, metabolic cost and fatigue during holding and carrying of loads. *Sci Rep*, 11(1):1–14.

O'Neill, C. T., Phipps, N. S., Cappello, L., Paganoni, S., and Walsh, C. J. (2017). A soft wearable robot for the shoulder: Design, characterization, and preliminary testing. In *2017 International Conference on Rehabilitation Robotics (ICORR)*, pages 1672–1678. IEEE.

Otten, B. M., Weidner, R., and Argubi-Wollesen, A. (2018). Evaluation of a novel active exoskeleton for tasks at or above head level. *IEEE Robot Autom Lett*, 3(3):2408–2415.

Pinto-Fernandez, D., Torricelli, D., del Carmen Sanchez-Villamanan, M., Aller, F., Mombaur, K., Conti, R., Vitiello, N., Moreno, J. C., and Pons, J. L. (2020). Performance evaluation of lower limb exoskeletons: a systematic review.

IEEE Trans Neural Syst Rehabilitation Eng., 28(7):1573–1583.

Pott, P. P., Wolf, S. I., Block, J., van Drongelen, S., Gruen, M., Heitzmann, D. W., Hielscher, J., Horn, A., Mueller, R., Rettig, O., et al. (2017). Knee-ankle-foot orthosis with powered knee for support in the elderly. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 231(8):-715–727.

Quinlivan, B. T., Lee, S., Malcolm, P., Rossi, D. M., Grimmer, M., Siviy, C., Karavas, N., Wagner, D., Asbeck, A., Galiana, I., and Walsh, C. (2017). Assistance magnitude versus metabolic cost reductions for a tethered multiarticular soft exosuit. *Science robotics*, 2(2):eaah4416.

Quintero, D., Lambert, D. J., Villarreal, D. J., and Gregg, R. D. (2017). Real-time continuous gait phase and speed estimation from a single sensor. In *2017 Conference on Control Technology and Applications (CCTA)*, pages 847–852. IEEE.

Quintero, H. A., Farris, R. J., and Goldfarb, M. (2011). Control and implementation of a powered lower limb orthosis to aid walking in paraplegic individuals. In *International Conference of Rehabilitation Robotics (ICORR)*, pages 1–6. IEEE.

Ramachandran, V., Shintake, J., and Floreano, D. All-fabric wearable electroadhesive clutch. *Adv Mater Technol., volume=4, number=2, pages=1800313, year=2019, doi=https://doi.org/10.1002/admt.201800313, publisher=Wiley Online Library*.

Rehmat, N., Zuo, J., Meng, W., Liu, Q., Xie, S. Q., and Liang, H. (2018). Upper limb rehabilitation using robotic exoskeleton systems: A systematic review. *Int J Intel Robot Appl*, 2(3):283–295.

Sankai, Y. (2010). Hal: Hybrid assistive limb based on cybernics. In *Robotics research*, pages 25–34. Springer.

Sasaki, D., Noritsugu, T., and Takaiwa, M. (2005). Development of active support splint driven by pneumatic soft actuator (assist). In *2005 International Conference on Robotics and Automation (ICRA)*, pages 520–525. IEEE.

Sawicki, G. S., Beck, O. N., Kang, I., and Young, A. J. (2020). The exoskeleton expansion: improving walking and running economy. *J Neuroeng Rehabil*, 17(1):1–9.

Schmidt, K., Duarte, J. E., Grimmer, M., Sancho-Puchades, A., Wei, H., Easthope, C. S., and Riener, R. (2017). The myosuit: Bi-articular anti-gravity exosuit that reduces hip extensor activity in sitting transfers. *Frontiers in Neurorobotics*, 11:57.

Schumacher, C., Grimmer, M., Scherf, A., Zhao, G., Beckerle, P., and Seyfarth, A. (2018). A movement manipulator to introduce temporary and local perturbations in human hopping. In *2018 7th International Conference on Biomedical Robotics and Biomechatronics (Biorob)*, pages 940–947. IEEE.

Seo, K., Park, Y. J., Lee, J., Hyung, S., Lee, M., Kim, J., Choi, H., and Shim, Y. (2019). Rnn-based on-line continuous gait phase estimation from shank-mounted imus to control ankle exoskeletons. In *2019 16th International Conference on Rehabilitation Robotics (ICORR)*, pages 809–815. IEEE.

Shahid, T., Gouwanda, D., Nurzaman, S. G., et al. (2018). Moving toward soft robotics: A decade review of the design of hand exoskeletons. *Biomimetics*, 3(3):17.

Sharbafi, M. A. (2020). The key elements in the design of passive assistive devices. In *International Symposium on Wearable Robotics*, pages 19–25. Springer.

Sharbafi, M. A., Barazesh, H., Iranikhah, M., and Seyfarth, A. (2018). Leg force control through biarticular muscles for human walking assistance. *Frontiers in Neurorobotics*, 12:39.

Sharbafi, M. A., Seyfarth, A., and Zhao, G. (2017). Locomotor subfunctions for control of assistive wearable robots. *Frontiers in Neurorobotics*, 11:44.

Simpson, C., Huerta, B., Sketch, S., Lansberg, M., Hawkes, E., and Okamura, A. (2020). Upper extremity exomuscle for shoulder abduction support. *IEEE Trans Med Robot Bionics*, 2(3): 474–484.

Steinhilber, B., Luger, T., Schwenkreis, P., Middeldorf, S., Bork, H., Mann, B., von Glinski, A., Schildhauer, T. A., Weiler, S., Schmauder, M., et al. (2020). Einsatz von Exoskeletten im beruflichen Kontext zur Primär-, Sekundär-, und Tertiärprävention von arbeitsassoziierten muskuloskelettalen Beschwerden.

Strickland, E. (2012). Good-bye, wheelchair. *Spectrum, IEEE*, 49(1):30–32.

Swartz, A. M., Strath, S. J., Bassett, D. R., Moore, J. B., Redwine, B. A., Groër, M., and Thompson, D. L. (2003). Increasing daily walking improves glucose tolerance in overweight women. *Prev Med*, 37(4):356–362.

Tawil, R., McDermott, M., Mendell, J. R., Kissel, J., and Griggs, R. (1994). Facioscapulohumeral muscular dystrophy (fshd) design of natural history study and results of baseline testing. *Neurology*, 44(3 Part 1):442–442.

Thalman, C. and Artemiadis, P. (2020). A review of soft wearable robots that provide active assistance: Trends, common actuation methods, fabrication, and applications. *Wearable Technologies*, 1.

Thalman, C. M., Lam, Q. P., Nguyen, P. H., Sridar, S., and Polygerinos, P. (2018). A novel soft elbow exosuit to supplement bicep lifting capacity. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6965–6971. IEEE.

Tokur, D., Grimmer, M., and Seyfarth, A. (2020). Review of balance recovery in response to external perturbations during daily activities. *Human movement science*, 69:102546.

Tucker, M. R., Olivier, J., Pagel, A., Bleuler, H., Bouri, M., Lambercy, O., del R Millán, J., Riener, R., Vallery, H., and Gassert, R. (2015). Control strategies for active lower extremity prosthetics and orthotics: a review. *J Neuroeng Rehabil*, 12(1):1–30.

Tudor-Locke, C., Schuna, J. M., Barreira, T. V., Mire, E. F., Broyles, S. T., Katzmarzyk, P. T., and Johnson, W. D. (2013). Normative steps/day values for older adults: Nhanes 2005–2006. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 68(11):1426–1432.

Wei, W., Qu, Z., Wang, W., Zhang, P., and Hao, F. (2018). Design on the bowden cable-driven upper limb soft exoskeleton. *Appl Bionics Biomech*, 2018.

Weigand, F., Zeiss, J., Grimmer, M., and Konigorski, U. (2020). A novel approach for gait phase estimation for different locomotion modes using kinematic shank information. *IFAC PapersOnLine*, 53(2):8697–8703.

Wiley, M. E. and Damiano, D. L. (1998). Lower-extremity strength profiles in spastic cerebral palsy. *Dev Med Child Neurol*, 40(2):100–107.

World Health Organization (2011). Global health and ageing. Technical report, World Health Organization.

Xiloyannis, M., Cappello, L., Binh, K. D., Antuvan, C. W., and Masia, L. (2017). Preliminary design and control of a soft exosuit for assisting elbow movements and hand grasping in activities of daily living. *J Rehabil Assist Technol Eng*, 4:2055668316680315.

Xiloyannis, M., Chiaradia, D., Frisoli, A., and Masia, L. (2019). Physiological and kinematic effects of a soft exosuit on arm movements. *J Neuroeng Rehabil*, 16(1):29.

Xiloyannis, M., Haufe, F. L., Duarte, J. E., Schmidt, K., Wolf, P., and Riener, R. (2020). Physical therapy and outdoor assistance with the myosuit: Preliminary results. In *International Symposium on Wearable Robotics*, pages 257–261. Springer.

Yoshimitsu, T. and Yamamoto, K. (2004). Development of a power assist suit for nursing work. In *SICE 2004 Annual Conference*, volume 1, pages 577–580. IEEE.

Young, A. J., Foss, J., Gannon, H., and Ferris, D. P. (2017). Influence of power delivery timing on the energetics and biomechanics of humans wearing a hip exoskeleton. *Frontiers in bioengineering and biotechnology*, 5:4.

Zeiss, J., Weigand, F., Grimmer, M., and Konigorski, U. (2020). Control of a transtibial prosthesis with monoarticular and biarticular actuators. *IFAC PapersOnLine*, 53(2):8689–8696.

Zhang, J., Cheah, C. C., and Collins, S. H. (2015). Experimental comparison of torque control methods on an ankle exoskeleton during human walking. In *2015 International Conference on Robotics and Automation (ICRA)*, pages 5584–5589. IEEE.

[Zhang and Arakelian, 2018] Zhang, Y. and Arakelian, V. (2018). Design of a passive robotic exosuit for carrying heavy loads. In *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, pages 860–865. IEEE.

Zhao, G., Sharbafi, M., Vlutters, M., Van Asseldonk, E., and Seyfarth, A. (2017). Template model inspired leg force feedback based control can assist human walking. In *2017 International Conference on Rehabilitation Robotics (ICORR)*, pages 473–478. IEEE.

Zhao, G., Sharbafi, M. A., Vlutters, M., van Asseldonk, E., and Seyfarth, A. (2019). Bio-inspired balance control assistance can reduce metabolic energy consumption in human walking. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(9):1760–1769.

DOI: 10.2478/9788366675896-008 • Technikpsychologie • 95-103

Proceedings of the International Symposium on Technikpsychologie

Competence retention for non-routine activities in digitized working environments (CONDITION) - studies based on the professions of chemical technician and pharmaceutical technician

Stephanie Conein†, Thomas Felkl†

Federal Institute for Vocational Education and Training

Abstract

Automation[1] related skill decay was already being researched in high-risk industries such as aviation or nuclear-power plants, but no empirical studies concerning process-oriented industries such as chemical or pharmaceutical production have been conducted so far. The study CONDITION aims to close this gap by investigating whether the problem of automation induced skill decay also exists in the workplaces of the chemical and pharmaceutical production. Furthermore, we wanted to find out in what way this decay applies to the occupational activities of chemical technician and pharmaceutical technician (which competences are affected, how often does this problem occur, which factors influence the skill decay and what are the consequences of it). The research questions of the project were addressed using a mixed methods approach, started with a systematic literature analysis, followed by 21 qualitative interviews and an online survey with 210 participants.

We were able to show that there is a problem of automation related skill decay affecting the workplaces of the chemical and pharmaceutical production. Moreover we identified some essential competences to master these situations successfully. Experience is crucial to acquire these competences and this implies that the problem of automation related skill decay certainly will become bigger because opportunities for young professionals to experience the actual plant get less and less since from the very first day they mostly work in a highly digitized and highly automated environment.

1 Talking about digitization and automation it is necessary to make clear that these are different topics which are nevertheless closely related. A comprehensive distinction is made by Schumacher, A., Sihn, W., & Erol, S. (2016) which analyze the terms digitization and automation trying to understand about differences and common ground. They describe digitization as "(…) the conversion of continuous analog, noisy and smoothly varying information into clear bits of 1s and 0s," and stated that automation "Describes the implementation of technology, software and programs to accomplish a procedural outcome with little or no human interference". Moreover they make clear, that "(…) one cannot exist without the other as any kind of automation nowadays requires digital elements to work without human interference and any kind of digitization requires elements to automatically handle and display information. Therefore, research focusing on advanced or smart manufacturing has to include both concepts to allow for a comprehensive analysis." This is what we did in our investigation. When talking about automation we always imply digitization as the basis for it.

Keywords

automation • skill decay • competences • production

1. Introduction

In the course of the ongoing digitization of workplaces, numerous studies in recent years have addressed the question which new competence requirements arise for skilled workers from the technological developments. The focus was partly cross-industry (acatech, 2016; Hammermann & Stettes, 2016; Schmidt et al., 2016), partly sector-specific (Spöttl et al., 2016) or occupation-specific (Conein, 2019). The aim of these studies was to anticipate new qualification requirements at an early stage and to align education and training accordingly.

Less attention, however, was paid to the issue of what consequences digitization has for existing vocational

†*Corresponding author: Stephanie Conein; Thomas Felkl*

E-mail: conein@bibb.de; thomas.felkl@bibb.de

competences. There seems to be agreement that they will not generally lose their relevance. Tenberg & Pittich (2017), for example, state in this context that although there are considerable changes on the horizon that will have an impact on most training occupations, no previously significant competences will become superfluous.

One reason given for this is that workplaces are currently still very differently permeated by digital technologies. There are major differences between occupations, but also differences within occupations that are specific to company size and sector (Zinke, 2019). Finally, there is also a varying degree of digitization within individual companies, depending on the workplace. Thus, many workplaces are not yet even affected by a change in competency requirements, and many skilled workers still need the same range of skills and abilities.

But even in the case of jobs that are largely digitized and therefore automated, there are arguments for maintaining the competences that have been relevant up to now. These arguments are based primarily on non-routine work situations that occur again and again, because of the so called "Ironies of Automation" postulated by Lisanne Bainbridge (1983). Bainbridge states that the attempt to eliminate the operator as a human source of error, which is mostly undertaken by system designers in the course of automation, is counteracted by the fact that system designers are also humans and thus transfer their own errors into the system, so that malfunctions will occur even in highly automated systems.

Operators then have to use ad hoc skills and knowledge that they have not needed for some time. They have to switch from routine operation, in which they primarily act as mere supervisors of the systems, to non-routine operation, whose demands on their skills and knowledge are incomparably greater. Weyer (1997, p. 245) states for example that in these non-routine-situations it is necessary to interpret deviating values, to diagnose the cause of the malfunction in the shortest possible time and to take manual countermeasures in order to avoid a crisis-like escalation.

Due to the infrequent occurrence of non-routine situations, the operator rarely applies the relevant skills and knowledge to handle the situation. Knowledge and skills that have been acquired once but are infrequently applied over longer periods of non-use may be prone to decay. Skill decay[1] is defined as the inability to retrieve formerly trained and acquired knowledge and skills after periods of non-use with a consequence of decreased performance (Arthur Jr et al., 1998). As a result, operators experience difficulties in acting adequately when a non-routine situation arises. This phenomenon is described extensively by Bjork & Bjork (2006, p. 114) „ (…) no matter how well learned items are at some point in time they eventually become non-recallable given a long enough period of disuse".

Automation-related skill decay as a problem of mastering non-routine-situations has so far been addressed primarily in the context of aviation (Wiener & Curry, 1980), military (O'Hara, 1990), police (Angel et al., 2012) and critical infrastructure (Webb & Angel, 2018). With few exceptions, the research focuses on the extent and potential solutions to workplaces in the sectors mentioned.

The fact that the problem of automation-related loss of skills also exists in production has already been recognized. Frank and Kluge (Frank & Kluge, 2018, p. 215) explicitly point out that in the course of the automation of production, the issue of skill loss also represents a challenge: "But many skills are only required infrequently, for instance due to a high level of automation in production (...), demands of worker flexibility and a high task variety (...), or long periods of non-use during daily operations. Accordingly, skill retention becomes a challenge (...)". Specifically with regard to chemical production, Baumhauer et al. (2019) state in their working paper on skilled production work in the chemical industry that retention of knowledge and skills for possible incidents represents a major challenge in digital transformation for companies.

However, empirical studies on the exact quality and quantity of the problem are still lacking, as Kluge (2014, p. 175) states: "In process automation, skill decay has not been investigated systematically", although the consequences are likely to be at least potentially as serious as in the already well-studied areas.

The research project "Competence retention for non-routine activities in digitized working environments (CONDITION) - studies based on the professions of chemical technician and pharmaceutical technician" carried out from 2020 to 2023 at the Federal Institute for Vocational Education and Training aims to fill this gap. In this project, the problem of the automation-related loss of competence at the workplaces of chemical and pharmaceutical technicians was first examined qualitatively and then in a second step quantitatively.

The following research questions were to be answered:

1. Does the problem of automation-related loss of competence also exists at the workplaces of the chemical and pharmaceutical production?
2. Which competences are affected?
3. Are there factors that influence the loss of competences?

2 When talking about skill decay in the present article the word skill is used as a synonym for competence, because in the English-speaking literature the words are often used synonymously in this context. Nevertheless, in our research we distinguish between skills and competences, defining the latter referring to the German Qualification Framework as the sum of all knowledge, skills and willingness necessary to perform a task.

4. What are the current and potential consequences?
5. Are there already measures to prevent or compensate for the loss of competence?

In this paper due to the limited space we only deal with the first two questions.

2. Methodology

The research questions of the project were addressed using a mixed methods approach. First, we carried out a systematic literature analysis. The aim of this analysis was on the one hand, to check to what extent the research question of the project has already been dealt with in exactly the same or in a slightly modified way. On the other hand, the current state of research in the field of skill decay should be surveyed in order to identify the main influencing factors, which should then also be taken into account in our own surveys.
The systematic literature analysis was followed by the empirical collection of qualitative data. Originally there was the intention to carry out work process studies consisting of work observation and action-oriented professional interviews (Becker 2018). However, due to the corona pandemic, access to the workplaces was not possible, so only telephone interviews could be conducted, which now form the basis for the qualitative data. The interviews were semi-structured. There were seven main questions but related to the respective answers additional questions were asked.
The main questions were:

1. Could you please briefly describe your work area and the associated production processes?
2. On a scale of 1-5, how automated would you classify the production process you describe?
3. Are there any situations that deviate from the normal daily routine?
4. What knowledge and skills help you (or your colleagues) in these situations?
5. How does this knowledge and skill differ from what is needed in routine situations?
6. Where did you (or your colleagues) get this skill or knowledge from?
7. Is there anything you would like to have in order to be able to cope better with the non-routine situations described?

A total of 21 telephone interviews were carried out. The average time of the conversations was 30 minutes. Of the interviewees 16 worked as professionals, four of them in the pharmaceutical sector and twelve in the chemical sector. Two persons worked in quality management, one of them in the area of chemistry and one in the area of pharmaceuticals. One person works mainly in training (chemistry), two others in human resources (pharma) and one as a consultant (both sectors).
The interviews were conducted by telephone, recorded, transcribed and evaluated using the content analysis according to Mayring (2022). The codes used were formed in a first step on the basis of the outcomes of the literature analyses which led to relevant factors concerning the loss of competence and were supplemented in a second step by codes inductively generated from the source material.
When selecting the interview partners, the aim was to interview professionals from both the pharmaceutical and chemical production sectors. In addition, individuals should have several years of professional experience (preferably more than ten). It was also intended to interview both professionals from corporate groups and professionals from small and medium-sized enterprises which unfortunately could not be realized. Only two of the interviewees work for small companies, all others work in larger corporations.
Based on the data from the qualitative interviews, an online questionnaire was developed in a further step. The online survey was carried out at the beginning of 2022 and was aimed at skilled workers (usually chemical and pharmaceutical technicians) and supervisors and managers (hereafter referred to as managers) in chemical and pharmaceutical production. The aim was to investigate the extent to which the findings of the interviews could also be confirmed within a larger group of skilled workers.
The general themes of the questionnaire were:

1. General data of the participants and their workplace
   - professional position
   - size of the enterprise/company
   - sector
   - plants serviced
   - work experience
   - degrees
   - age
2. Data about occurrence of (problematic) situations
   - non-routine situations (NRS)
   - reaction to the occurrence of a NRS
   - relevance and recall of knowledge/skills when a NRS occurs
   - relevance of different personal characteristics when a NRS occurs
   - possible consequences of the occurrence of NRS
3. Existing and wished support in (problematic) situations
   - general possibilities of support
   - training in particular

The original target of 700 participants was not reached. Here, too, the lack of and difficulty in establishing personal contacts

due to the pandemic certainly played a role. The questionnaire was completed 210 times, with 50 professionals and 160 managers participating, with over 80% of the latter having been trained in chemistry or pharmacy.

3. Results

3.1 The problem of automation-related loss of competence at the workplaces of the chemical and pharmaceutical production

A prerequisite for the existence of problematic automation-related loss of competence is not only a partially or fully automated production, but also the occurrence of non-routine situations, because only in these situations the lack of competences lost due to automated routine becomes evident. Therefore, in a first step we asked whether non-routine situations (NRS) occur at the workplaces of the respective chemical or pharmaceutical plants. On the basis of the data from the qualitative interviews, we could clearly answer this question in the affirmative. Moreover, we were able to distinguish three types of NRS: the frequent but unplanned NRS, the rare but planned NRS and the rare and unplanned NRS. Out of the three only the latter appeared relevant for our specific interest since it seems less likely that skill decay plays a role in frequent or planned NRS.

The results from the online survey confirmed the data from the interviews. We asked professionals and managers whether one (or more) of the three types of NRS occur in their everyday work. Only 3% (Managers) and 2% (Professionals) of the respondents respectively stated that they did not experience any non-routine situations in their everyday work. 50% were aware of rare unplanned NRS and more than 60% confirmed that they experienced the other two types.

In a second question, we asked whether in in these situations it ever happened that they or their colleagues did not immediately know what to do. This was also reported to us during the interviews and confirmed. 67% of the managers and 65% of the professionals (Figure 1), who had previously reported that rare, unplanned non-routine situations occur in their daily work, stated that they or their colleagues or employees have already faced this problem.

The last thing to clarify is whether the reason for the loss of competence is to be found due to a largely automated environment. Again, the data from the interviews contain statements that support this hypothesis:

“I think it would certainly be good if we were to say in some places, take the process by hand and run it. I don’t think the employees could do that anymore. I don’t think they would be able to cope with all the temperatures, (...) they are basically dependent on the system, because the automated system can do that. And I’m convinced that it will be difficult if things actually get out of hand, and experience shows that they don’t do the right thing” (Manager).

For this purpose, the respondents who stated that they had experienced loss of competences for non-routine situations or saw this in their colleagues or co-workers were asked about the reasons for this in the online-questionnaire. The results show that forgetting knowledge due to the automation of the systems is named as the main reason for the difficulties in recalling the knowledge and skills relevant in the NRS (see Figure 2) with more than 75% agreeing strongly or rather strong to the respective statement.

In conclusion it can be said that (as already postulated theoretically) there is empirical evidence of an experienced automation-related loss of competence in the workplaces of chemical and pharmaceutical production, which prevents or hinders the retrieval of the relevant competences in rare and unplanned non-routine situations.

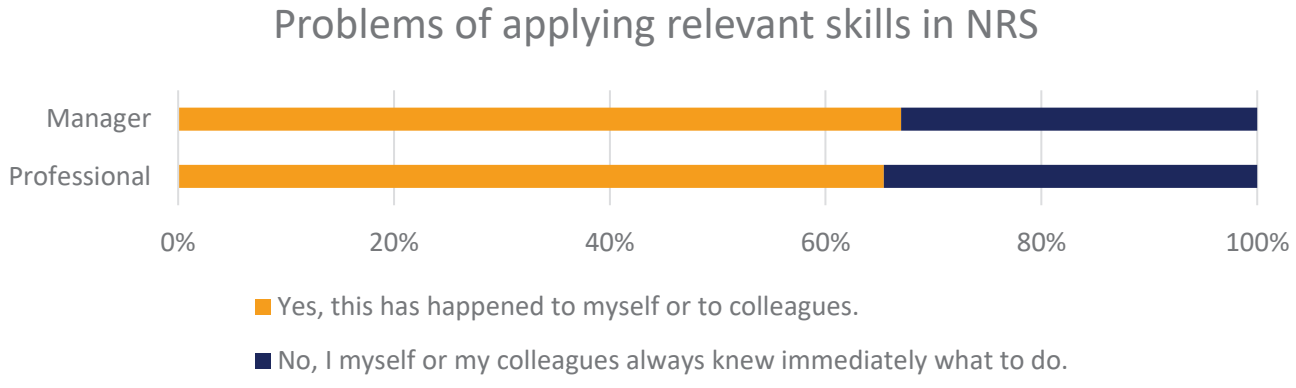


Figure 1. Results from the question: „In the past, have you or a colleague ever been in a rare non-routine situation where you or a colleague did not immediately know what to do?“ Managers n=80, Professionals n=26.

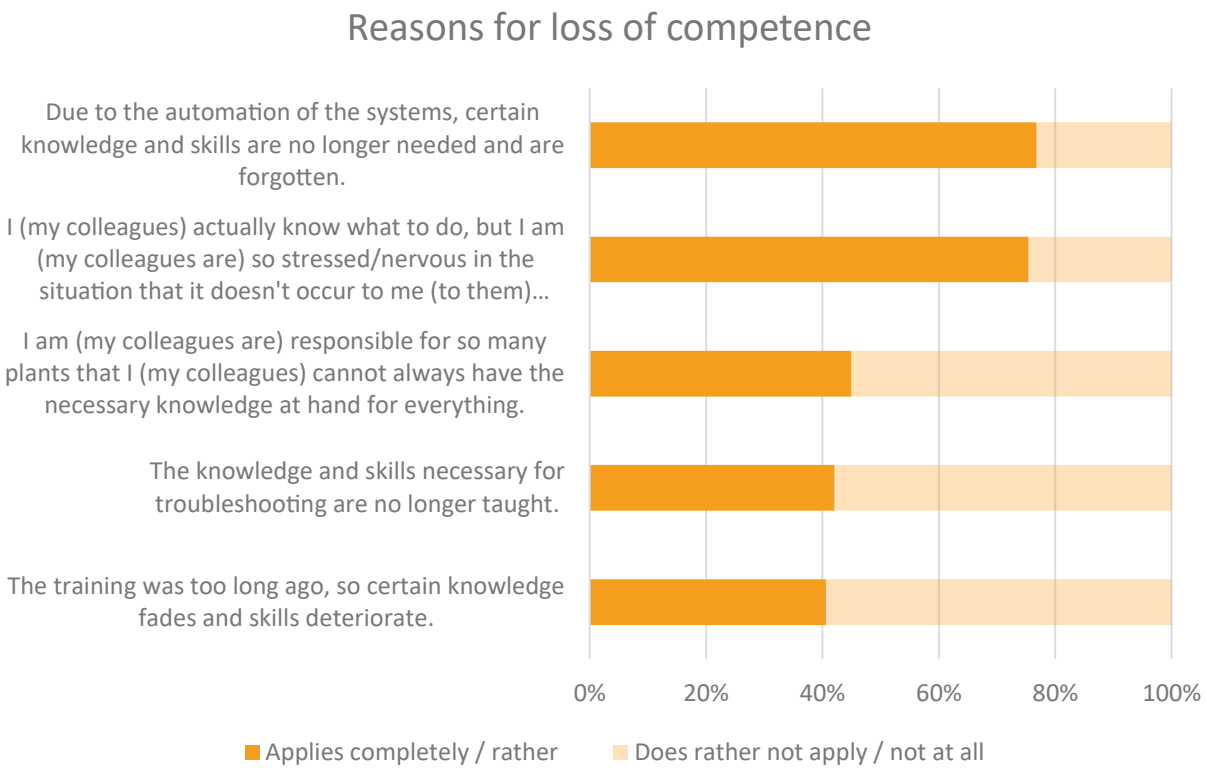


Figure 2. Results from the question: „You have just stated that you or your colleagues sometimes find it difficult to recall certain knowledge or apply skills in non-routine situations. What reasons do you see for this? Please indicate to what extent the following reasons apply.“ n = 69.

3.2 Which competences are affected?

As stated above we adopted the competence definition of the German Qualification Framework (DQR) (Arbeitskreis Deutscher Qualifikationsrahmen), and define competence therefore as the sum of all knowledge, skills and willingness necessary to fulfil a task. This threefold division also overlaps with many other definitions of competence, as for example in Westera's remarks (2001, p. 87) on the concept of competence, who states at the end of a detailed analysis of various competence concepts: "Likely we could have come to this conclusion before the analysis: when all is said and done, the only determinants of human abilities are possessing (knowledge), feeling (attitudes) and doing (skills)!"

Analogous to the DQR, knowledge refers to the totality of facts, principles, theories and practice in a field of learning or work that come a result of learning and understanding. Knowledge can be in an explicable form or as implicit knowledge (2015, p. 99), i.e. not completely or adequately verbalizable, objectifiable and formalizable. It can also be declarative (knowing what), i.e. referring to facts, or procedural (knowing how), forming the prerequisite for actions.

Skills, also analogous to the DQR, refer to the ability to apply knowledge in order to carry out tasks and solve problems. Skills are divided into cognitive skills (logical, intuitive and creative thinking) and practical skills (dexterity and use of methods, materials, tools and instruments).

Willingness is understood in the context of competence as an attitude and therefore coincides with the term "attitude" used in English.

Due to the limited space of this paper we couldn't present all our findings concerning the relevance of certain knowledge, skills and attitudes. We limit ourselves to the most important elements of competences that were reported to be especially helpful in handling NRS in the chemical- and pharmaceutical production.

On the basis of the outcomes of the qualitative interviews we asked for a special set of knowledge and skills in the online-questionnaire. Three of them were reported as especially important (see Figure 3).

To sum it up, concerning knowledge and skills, the retrieval of knowledge about the actual plant, its dimensions and processes that take place as well as the skill to operate the plant by hand (e.g. with no or limited support from the process control system) are reported as most important to handle NRS. More than 90% of the respondents rated these three as very important or rather important. These findings were supported by the interviews. The relevance of knowledge about the actual plant "outside" and their representation in the

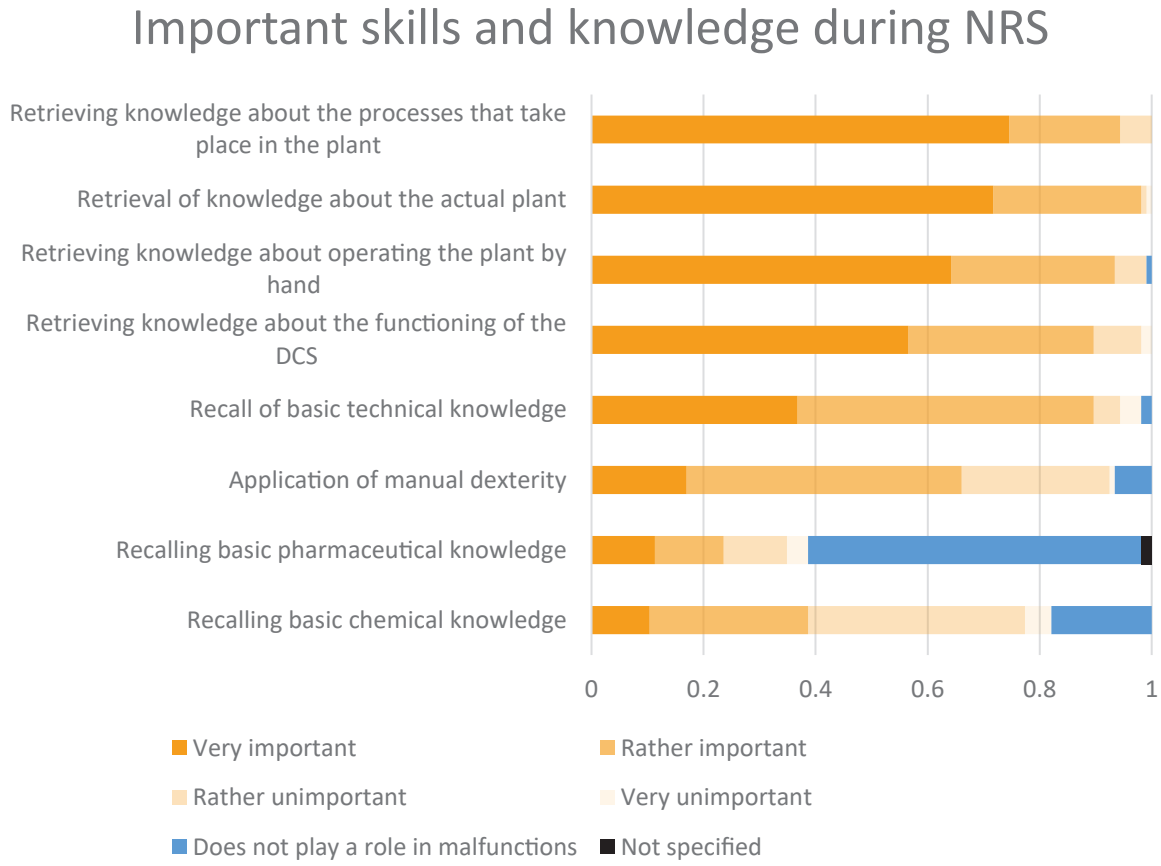


Figure 3. Results from the question: „Imagine that you or your colleagues are confronted with such a non-routine situation at your current workplace. What is important or unimportant and what is easy or difficult for you or your colleagues in this situation?“ n=106.

process control system were mentioned several times: “And that is, I think, actually elementarily important, that people have an orientation of their plant, they have to know their plant. And you can’t just do that in the control room at the process control system. I should say, yes, that is absolutely not enough. That is the experience we have made, that it is extremely important that people manage this link with outside and inside” (Manager).

The so-called process-knowledge was seen as particularly relevant as well. Process-knowledge refers to the chemical processes taking place in the plants, their modes of operation and interdependencies. It comprises the chemical components involved, the order in which the components are added, but also knowledge about the reasoning behind the order and the skills to intervene successfully:

“Well, of course we then look to see what the reason was in this particular battery that the values were so bad. And then, I’ll say, various adjustments are made to experiment with the water, the filter, the vacuum, the composition of the first reaction stage. And yes, there are many, many things that can be adjusted to improve the result” (Manager).

In addition to the question which competences are crucial to handle NRS we asked, how difficult it is for the professionals to apply the respective competences in the NRS. The assessment of the difficulty of retrieving the eight aspects was recorded using a 4-point scale - here from “very easy” to “very difficult” (see Figure 4).

The distributions for the individual aspects are all very similar; no particular aspect of recalling knowledge stands out. A majority of the respondents find it at least rather easy to recall certain knowledge in a non-routine situation: However, around 30% each reported difficulties in recalling specific knowledge. This also affected the knowledge about the actual plant and its processes and the knowledge of operating the plant by hand which were emphasized as particularly important in NRS. To sum it up, recalling the relevant knowledge and skills in the NRS doesn’t currently seem to be a big problem but it isn’t completely negligible either.

As mentioned above competence comprises not only knowledge and skills but also attitudes. The latter was also seen as particularly relevant in the interviews. On the one hand, it is relevant in the non-routine situations themselves.

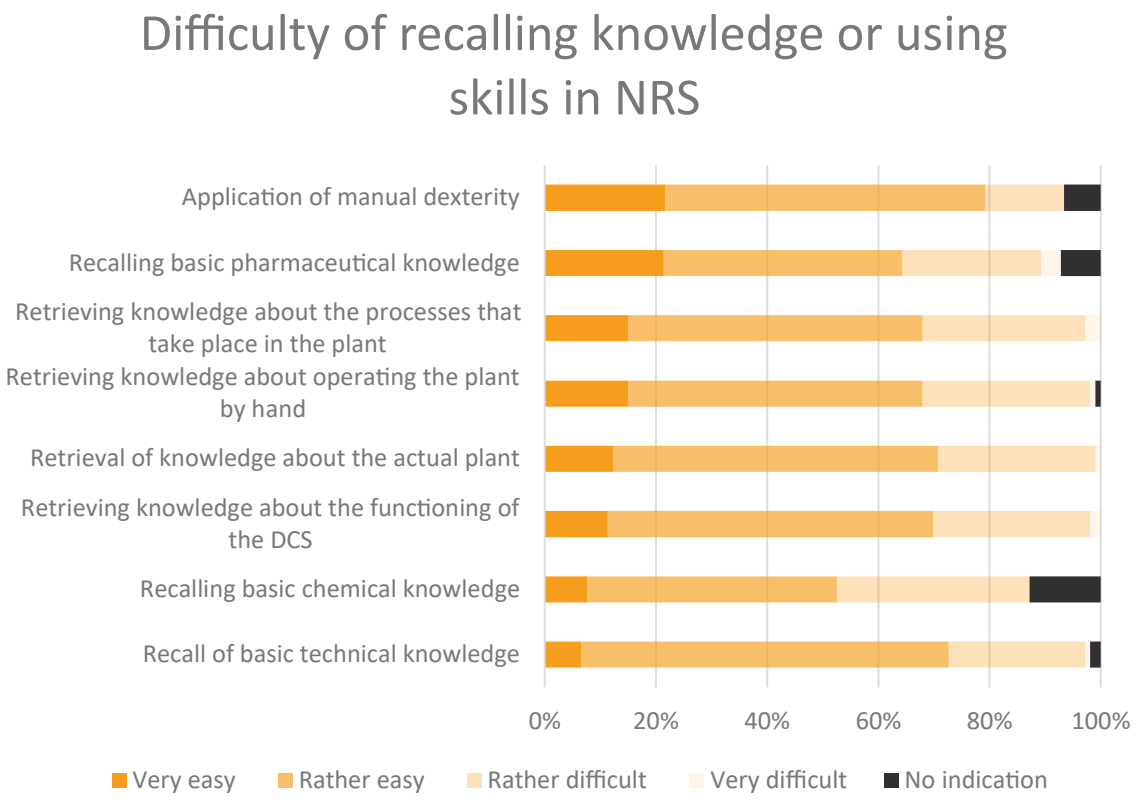


Figure 4. Results from the question: „Imagine that you or your colleagues are confronted with such a non-routine situation at your current workplace. What is important or unimportant and what is easy or difficult for you or your colleagues in this situation?“ n = 106.

Attitudes named as beneficial are calmness, composure but also courage: “But when you are in such a facility and you hear this monotonous operating noise, suddenly you hear a “BRRRRrrrrrrrr” and it gets quiet. That’s a worst-case scenario where I say: “Boah. Olli Kahn once said: You need balls for that” (Manager).

On the other hand, in the run-up to the non-routine situation, an attitude of curiosity and interest in the plant and the processes taking place in it are necessary in order to develop a deeper understanding of the process. Here, the attitude conditions and supports the acquisition of competence.

“Yes, that’s how it is. You have to be interested in the system, you also have to be interested in the failure. That’s what happens when you come back on shift, oh, they’re running again, how did you manage that? Or not, thank goodness, they’re running again” (Manager). Again, the findings of the interviews were supported by the outcomes of the survey. As Figure 5 shows, calmness and composure were rated as very helpful to handle NRS. Also, intelligent colleagues are particularly trusted to deal with these situations which might support the importance of curiosity and interest but further and systematical investigation into this matter would be required to make more definite statements.

For acquiring these important competences (knowledge about the actual plant, process-knowledge which comprises also skills to control and adjust the processes and the attitude of calmness and composure) one key factor is essential: experience. To gain for example a comprehensive understanding of the processes taking place in a factory plant it is necessary to gain experience with the actual plant and to not just stay in the control room. This is clearly articulated in the interviews. Gaining experience involves going outside, inspecting the plant, working manually and using all senses. “In the past, when the plant was started up, I went outside and it was running visually well inside, and then I walked around outside, and it wasn’t just a matter of walking around and looking to see if something was splashing around or something, but you go into the plant and it’s like a concert. Every piece of equipment makes a sound and the whole system is like a concert, so everyone plays their part. And if someone plays the wrong note, you have to hear it” (Manager). And: “Yes, you can’t run the control room if you don’t know what’s going on outside. That is already an important point. You have to understand the procedure and know what is important, what you need to do outside and what you are not allowed to do.” (Manager).

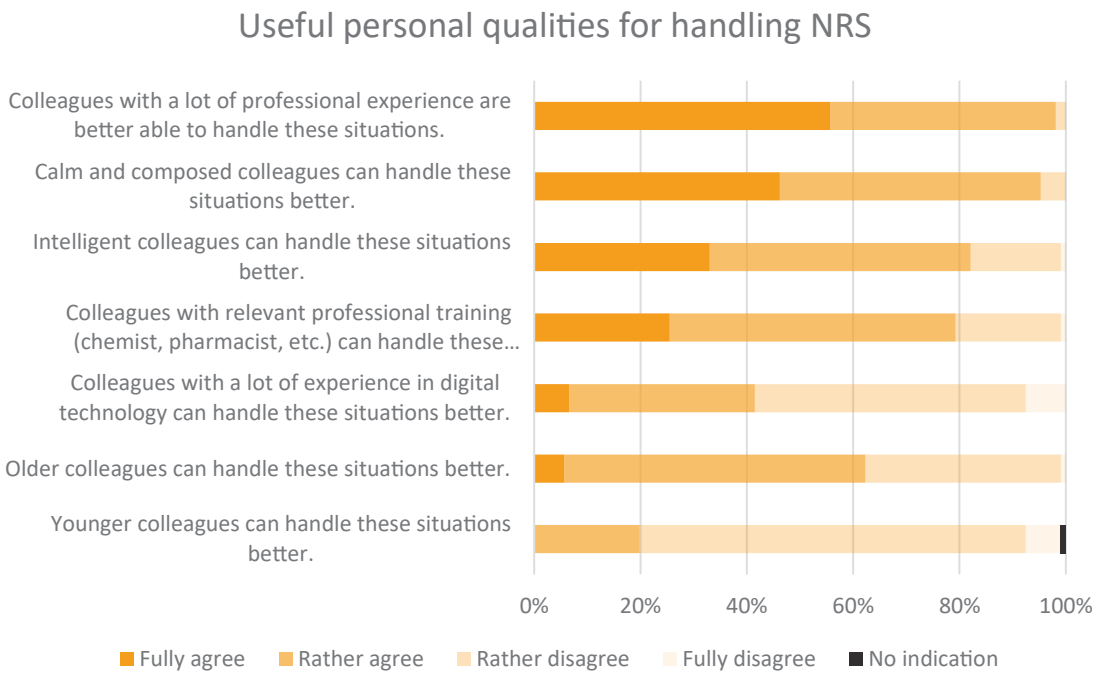


Figure 5. Results from the question: „People deal with such non-routine situations in very different ways. Based on your own experience, how do you agree with the following statements?“ n=106.

Experience also seems to be one important factor to develop an attitude of calmness and composure: “You have to say that the older ones know the processes better, maybe they already know a lot of processes that have been running for 30 years. If something happens, they’ve experienced everything, I must say, from mistakes or problems, where they really know exactly what you have to do or how you have to react. And they also stay calmer, I must say.” (Manager).

The importance of experience is also obvious in the findings of the survey. As shown in Figure 5 more than 96% (the highest approval rate) of the respondents agree with the statement, that experienced colleagues are better at handling NRS.

The fact that the knowledge, skill and attitudes most important for handling NRS could only be acquired by experience, especially by the experience of the actual plant and its components and processes leads to the problem, that because of the ongoing digitization and the associated automation the opportunities to gain this experience diminish. Therefore, it becomes more and more difficult for the younger professionals to acquire the relevant competences for NRS. Thus, the individual skill decay, deriving from working in highly automated environments and applying the respective competences rarely, is accompanied with a generation-related skill decay. This problem is mentioned in the interviews: “Then, of course, the employees got to know the plant in a less digitized state, (...) Younger employees lack exactly this knowledge, so of course, people who have been there longer and know the old state can react better to such disturbances” (Professional). This goes together with other findings from the survey. At the end of the survey we asked the participants how well they generally feel prepared for NRS. The main finding was that the more professional experience the professionals have, the better prepared the respondents generally feel for disruptions. Only people with at least 11 years of experience in their profession stated that they felt very well prepared. In contrast, the respondents with less than 5 years of professional experience, feel (usually rather) less prepared.

4. Summary

Automation related skill decay occurs at the workplaces of chemical and pharmaceutical production. A significant proportion of our respondents reports that they or their colleagues have faced NRS in which they did not immediately knew what to do and the majority of them indicated that the automation is the reason for this.

Moreover, some competences could be identified which were crucial in handling NRS. Knowledge of the actual plant, process-knowledge and an attitude of calmness and

composure are essential to act adequately in the case of an unexpected disturbance. The key factor to acquire them is experience and this leads to the problem that although currently the automation related skill decay doesn't seem to be a problem in the chemical processing industry, it certainly will become bigger because opportunities for young professionals to experience the actual plant (with all senses) get less and less since from the very first day they mostly work in a highly automated environment.

References

acatech (Ed.). (2016). *Kompetenzentwicklungsstudie Industrie 4.0 – Erste Ergebnisse und Schlussfolgerungen*. München.

Angel, H., Adams, B. D., Brown, A., Flear, C., Mangan, B., Morten, A., Ste-Croix, C., & Gruson, G. (2012). *Review of the Skills Perishability of Police "Use of Force" Skills*. Humansystems Incorporated.

Arbeitskreis Deutscher Qualifikationsrahmen. (2011). *Deutscher Qualifikationsrahmen für lebenslanges Lernen.* Arbeitskreis Deutscher Qualifikationsrahmen.

Arthur Jr, W., Bennett Jr, W., Stanush, P. L., & McNelly, T. L. (1998). Factors That Influence Skill Decay and Retention: A Quantitative Review and Analysis. *Human Performance*, *11*(1), 57–101. https://doi.org/10.1207/s15327043hup1101_3

Bainbridge, L. (1983). Ironies of automation. In *Analysis, design and evaluation of man–machine systems* (pp. 129–135). Elsevier. https://doi.org/10.1016/0005-1098(83)90046-8

Baumhauer, M., Beutnagel, B., Meyer, R., & Rempel, K. (2019). *Produktionsfacharbeit in der chemischen Industrie: Auswirkungen der Digitalisierung aus Expertensicht*. Düsseldorf. Forschungsförderung, Working Paper.

Bjork, R. A., & Bjork, E. L. (2006). Optimizing treatment and instruction: Implications of a new theory of disuse. In L.-G. G. Nilsson & N. Ohta (Eds.), *Memory and society: Psychological perspectives.* (pp. 109–134). Psychology Press.

Conein, S. (2019). Berufsbildung 4.0 – Fachkräftequalifikationen und Kompetenzen für die digitalisierte Arbeit von morgen: Der Ausbildungsberuf „Verfahrensmechaniker/-in für Kunststoff- und Kautschuktechnik" im Screening. *Wissenschaftliche Diskussionspapiere*.

Frank, B., & Kluge, A [Annette] (2018). Is there one best way to support skill retention? Putting practice, testing and symbolic rehearsal to the test. *Zeitschrift Für Arbeitswissenschaft*, *73*(2), 214–228.

Hammermann, A., & Stettes, O. (2016). Qualifikationsbedarf und Qualifizierung: Anforderungen im Zeichen der Digitalisierung. *IW Policy Paper*(3/2016).

Kluge, A [Anette], & Frank, B. (2014). Counteracting skill decay: four refresher interventions and their effect on skill and knowledge retention in a simulated process control task. *Ergonomics*, *57*(2), 175–190. https://doi.org/10.1080/00140139.2013.869357

Mayring, P. (2022). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (13. Neuausgabe). Julius Beltz GmbH & Co. KG. http://nbn-resolving.org/urn:nbn:de:bsz:31-epflicht-2019387

Neuweg, G. H. (2015). *Das Schweigen der Könner: Gesammelte Schriften zum impliziten Wissen*. Waxmann Verlag.

O'Hara, J. M. (1990). The retention of skills acquired through simulator-based training. *Ergonomics*, *33*(9), 1143–1153. https://doi.org/10.1080/00140139008925319

Schmidt, K., Winkler, B., & Gruber, B. (2016). Skills for the future: zukünftiger Qualifizierungsbedarf aufgrund erwarteter Megatrends. *Ibw-Forschungsbericht*(187).

Schumacher, A., Sihn, W., & Erol, S. (2016). Automation, digitization and digitalization and their implications for manufacturing processes: In: Innovation and sustainability conference Bukarest (pp. 1-5)., *2016*, 1–5.

Spöttl, G., Gorldt, C., Windelband, L., Grantz, T., & Richter, T. (2016). Industrie 4.0 – Auswirkungen auf Aus- und Weiterbildung in der M+E Industrie.

Tenberg, R., & Pittich, D. (2017). Ausbildung 4.0 oder nur 1.2? Analyse eines technisch-betrieblichen Wandels und dessen Implikationen für die technische Berufsausbildung. *Journal of Technical Education (JOTED)*, *5*(1), 27–46.

Webb, B., & Angel, H. (2018). Maintaining skills and knowledge at work. *Applied Ergonomics and Human Factors*, 22–23.

Westera, W. (2001). Competences in education: a confusion of tongues. *Journal of Curriculum Studies*, *33*(1), 75–88.

Weyer, J. (1997). Die Risiken der Automationsarbeit: Mensch-Maschine-Interaktion und Störfallmanagement in hochautomatisierten Verkehrsflugzeugen. *Zeitschrift Für Soziologie*, *26*(4), 239–257. https://doi.org/10.1515/zfsoz-1997-0401

Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: promises and problems. *Ergonomics*, *23*(10), 995–1011. https://doi.org/10.1080/00140138008924809

Zinke, G. (2019). *Berufsbildung 4.0 - Fachkräftequalifikationen und Kompetenzen für die digitalisierte Arbeit von morgen: Branchen- und Berufescreening : vergleichende Gesamtstudie* (1. Auflage). *Wissenschaftliche Diskussionspapiere: Heft 213*. Barbara Budrich. http://nbn-resolving.de/urn:nbn:de:0035-0807-9