Computers & Security 148 (2025) 104025

Contents lists available at ScienceDirect

Computers & Security

journal homepage: www.elsevier.com/locate/cose

Cybersecurity behavior change: A conceptualization of ethical principles for behavioral interventions

Konstantinos Mersinas [a,*], Maria Bada [b], Steven Furnell [c]

[a] Department of Information Security, School of Engineering, Physical and Mathematical Sciences, Royal Holloway University of London, UK
[b] Department of Psychology, School of Biological and Behavioural Sciences, Queen Mary University of London, UK
[c] Faculty of Science, School of Computer Science, University of Nottingham, UK

ARTICLE INFO

Keywords:
Cybersecurity behavior change
Behavioral interventions
Ethics
Autonomy
Justice
Nonmaleficence
Beneficence
Transparency
Privacy

ABSTRACT

The importance of changing behaviors is gradually being acknowledged in cybersecurity, and the reason is the realization that a notable portion of security incidents have a human-related component. Thus, enhancing behaviors at individual level, can bring a significant reduction in security breaches overall. Behavior change refers to any modification of human behavior through some type of intervention. Interventions from behavioral economics and psychology are being increasingly introduced in the field, however, the ethics surrounding such interventions are largely neglected. In this paper, we raise the ethical issues associated with behavioral intervention approaches. We draw on the traditionally more mature field of biomedical ethics and propose six clusters of ethical principles suitable for cybersecurity behavior change. We conducted a survey ($N = 141$) to identify individuals' perceptions on the proposed ethical principles and validate their perceived usefulness. We analyze an existing intervention in the light of our six-principle conceptualization to showcase how it can be used as a practical apparatus. Our set of ethical principles are aimed for cybersecurity professionals, policy makers, and behavioral intervention designers, and can serve as a starting point for best-practice development in cybersecurity behavior change ethics.

1. Introduction

The term cybersecurity behavior change (CBC) refers to any modification in the behaviors of individuals which is related to cybersecurity. Behavior change has been utilized in fields such as psychology and health sciences as a means to shift individuals towards favorable or preferable actions, e.g., individual-targeted campaigns to change smoking habits (Hastings et al., 2004). Most CBC approaches are based on theories deriving from the fields of cognitive psychology and behavioral economics, and particularly nudge theory (Thaler and Sunstein, 2008), which posits that behavioral interventions can be designed into systems to influence human choice upon interaction (e.g., e-commerce and mobile health apps; Briggs et al., 2017). After their popularization by Thaler and Sunstein (2008), and Kahneman (2011), behavioral interventions have been adopted widely by policy-makers.

Cybersecurity breaches are repeatedly reported as being linked to people's behaviors (Alnifie and Kim, 2023), whereas, there are preferable behaviors, namely, behaviors which minimize risk and promote security hygiene. Thus, it is critical to prioritize the strengthening of 'human defenses'. We refer to *security hygiene* as a synonym for secure and/or preferable behaviors, on the criterion that they reduce the contextual attack surface. We note that behavioral interventions aim in minimizing the *human attack surface*, i.e., the set of all possible points of entry or, attack vectors, which can be exploited by attackers for gaining unauthorized access to information assets, or compromise systems, applications, and networks.

In more detail, security hygiene, i.e., the 'knowledge and behaviors which protect social, financial, and personal information risk' (Neigel et al., 2020) has been empirically conceptualized into the dimensions of device hygiene, data storage and transmission, social media, authentication, and email and messaging hygiene (Vishwanath et al., 2020). Thus, individual behaviors related to appropriate handling of phishing emails, the use of antivirus, firewalls, good password practices (e.g., in terms of strength and non-reuse), encryption, the acceptance of certificates, online shopping and online banking habits, information sharing, the utilization of privacy and security settings in apps, which apps are trusted and why, information sharing, and many more, all include preferable and secure behaviors which can minimize the risk exposure of

* Corresponding author.
E-mail address: konstantinos.mersinas@rhul.ac.uk (K. Mersinas).

https://doi.org/10.1016/j.cose.2024.104025
Received 25 February 2024; Received in revised form 16 July 2024; Accepted 25 July 2024
Available online 14 August 2024

individuals and their data. Such behaviors, however, need to be adopted and followed by users, and CBC provides the means to security professionals and policy-makers to achieve such an adoption (Mersinas and Bada, 2023).

CBC interventions range from simple reminders, to the provision of additional information, impact scenarios, examples of how to secure assets, or can be based on choice architecture, i.e., the design of how choices are presented (Münscher et al., 2016), e.g., the power of default options has been confirmed in several settings, indicating that the majority of people do not deviate from the default due to loss aversion, inattention, or the associated transaction costs (Dhingra et al., 2012). The acceptance of behavioral interventions is increasing in cybersecurity, since the so-called 'human aspects' of security are embraced in the field nowadays. But it is important to highlight that, since CBC interventions target human beings, they face the danger of becoming inconsiderate (e.g. by ignoring users' discomfort, frustration or harm), authoritative (e.g., through mandatory and strict policies), deceptive (e.g., by intentionally misleading users), coercive (e.g., by utilizing sanctions) or manipulative (e.g., without users' informed consent and awareness) in the name of achieving security.

However, the way by which such interventions can be utilized, e.g., in organizational settings, or more broadly for the public, have not been considered, to the best of our knowledge. Indicatively, at the time of writing, Google Scholar does not return any behavior change related paper via the search: *'((cybersecurity) OR (cyber security) OR (information security)) AND ((behavior) OR (behaviour)) AND (ethics)'*.

Behavioral interventions can be transparent or non-transparent, and, they can encourage reflective or automatic decisions. Non-transparent and/or automatic interventions can be considered manipulative (Caraban et al., 2019), but both categories can raise ethical concerns. In this paper, we build on previous work which identifies the strengths, weaknesses, and ethical issues of behavioral interventions in cybersecurity (Mersinas and Bada, 2023), and we propose a conceptualization of ethical principles to be used by security professionals and practitioners when evaluating or implementing CBC interventions.

The paper has the following structure. Section 2 explains the need for CBC ethics. In Section 3, we briefly describe three ethical traditions. Section 4, then, provides six representative ethical principles which constitute the proposed conceptualization. In Section 5, we present the analysis of the survey and highlight the key findings, demonstrating the perceived need for ethics in CBC interventions. Section 6 provides a discussion along with an example of how these ethical principles can be applied in CBC. Section 7 indicates limitations and future work, and the last section concludes the paper.

2. A view into cybersecurity behavior change

Behavior change has been studied in the context of cybersecurity, and, in particular, in relation to online security behaviors (Briggs et al., 2017) and approaches have been developed to assist in building relevant interventions in security (Coventry et al., 2014). With regards to ethics, more specifically, research focuses on penetration testing practices, DDoS attacks, ransomware, and system administration (Formosa et al., 2021), whereas other ethical frameworks focus on research in cybersecurity and right-based principles with legal applications (Loi and Christen, 2020). Specific types of behavioral interventions, namely fear appeals, have been examined in a security context (Renaud and Dupuis, 2019). However, there is a lack of scholarship on the broad ethical issues related to behavior change interventions in cybersecurity.

The importance of CBC ethics is evidenced by security reports. The Ponemon Institute (2019) estimates that 24 % of breaches are caused by human error. Thus, promoting secure behaviors via changing existing individual behaviors can reduce cybersecurity risk substantially.

A traditional dominant view in cybersecurity has been that of humans as the weakest link (indicatively, Anon., Cisco, 2017). At the same time, security training effectiveness is at least debatable (Pruemmer et al., 2023). Additionally, the new, at the time, view that users are not the enemy (Adams and Sasse, 1999) is becoming prevalent nowadays. The diversified cybersecurity working environments and the fast-paced nature of the field, and shortages of knowledge and resources, e.g., time pressure, may lead to anxiety, frustration, and risk-taking, making users more susceptible to attacks (Chowdhury et al., 2020).

Cyber-attacks can occur by insiders' negligence, lack of knowledge or malicious intent (Georgiadou et al., 2022). Lack of knowledge and understanding, specifically, can be the result of suboptimal information dissemination within an organization (Simon, 1991). Although, some models of behavior, such as the Theory of Reasoned Action (Fishbein and Ajzen, 1975) and the Theory of Planned Behavior (Ajzen, 1980) assume that humans make rational, informed, and predictable decisions, it has been repeatedly observed that individuals tend to make irrational, sub-optimal, non-utility-maximizing decisions, often in a predictable fashion (Ariely, 2008; Camerer, 2003, 2004; Kahneman, 2011). Such findings have been experimentally measured in cybersecurity too, e.g., security professionals are not found to minimize expected losses, are risk and ambiguity averse, and are susceptible to framing (Mersinas et al., 2016; Safi et al., 2021), and risk perceptions are found to be influenced by affect (Van Schaik et al., 2020).

Further, the limitations of training and education are linked to a neglect of human decision-making processes, and a failure to consider the various rationality types under which humans make choices (Mersinas et al., 2019). An indicative type of rationality which portrays such limitations is *bounded rationality*, according to which human decision-makers are not fully rational since they are bounded by limited time, cognitive capacity, and only have partial access to information for any given problem (Simon, 1972).

3. Ethical traditions

Various codes of ethics exist in cybersecurity (e.g., Anon., BCS, Anon., ISC^2, Anon., CREST), and these are, naturally, oriented towards the specific goals within the environments they are to be utilized in. Indicatively, the code of ethics for Certified Information Systems Security Professionals (Anon., ISC^2, 2024) encourages professionals to 'tell the truth' and 'make stakeholders aware of their actions on a timely basis', due to the organizational setting of their application.

There are various pathways to achieve behavior change (Mersinas and Bada, 2023). For the endeavor of building an ethical conceptualization for CBC, we draw on the three main ethical traditions (Bednar and Spiekermann-Hoff, 2020) as building blocks to base these potential pathways on; namely, utilitarian ethics, deontological ethics, and virtue ethics, which we present here, briefly.

- **Utilitarian** ethics are built around the notion of *utility*, a measure of anything valuable, from money to well-being. Individuals seek to maximize utility, often via cost-benefit analyses (Bentham, 1876; Mill, 1859). However, the focus is not on individualized benefits. Instead, a utilitarianist 'always prioritizes society' or the benefit of the many, and considers individuals as having secondary importance. The orientation towards universal benefit also presupposes a harm-avoidance orientation.
- Then, **deontological** ethics are closely related to a sense of duty and the idea of following universalizable rules of conduct (Kant, 1998 ed.). Here, the individual has the ultimate freedom of choice. This ethical tradition advocates that individuals ought to act in a fashion that they would want the whole society to follow, as a universal rule. This notion is depicted in Kant's Categorical Imperative, namely in the decision rule: 'act only according to that maxim by which you can at the same time will that it should become a universal law' (Kant, 1998, p. 422).
- Finally, **virtue** ethics posit that individuals make 'good' decisions for their own sake, not for further goals (Aristotle et al., 1980). A characteristic is that virtue ethics are applied in specific

environments, in contrast to, e.g., the abstract and universal deontological ethics. Context-reliance is coupled with a portrayed importance of individual voluntary action. In virtue ethics, these two factors indicate the key role of individual *responsibility* (Van Staveren, 2007).

The applicability and relevance of the three ethical traditions with cybersecurity is not straightforward, with various components across the traditions being related to characteristics within different security contexts (Mersinas and Bada, 2023). For example, the individual, can be both a target and an attack vector in cyber-attacks. Then, successful attacks directly impact both systems and additional users. Thus, it would be beneficial to build a line of defense at the individual level. The potential impact spread to other users and systems is related to virtue ethics, which focus on collective social responsibility and the common good through individual responsibility.

4. Ethics in cybersecurity and other fields

The literature on ethics in cybersecurity is largely dichotomized into ethics for research involving human subjects, such as the Menlo report (Dittrich and Kenneally, 2012), and ethics focused on rights (Loi and Christen, 2020). To the best of our knowledge, there is no equivalent approach for behavior change in cybersecurity.

In order to identify a set of principles, we draw upon previous research on ethics in cybersecurity and other fields. Multiple frameworks exist for ethics in artificial intelligence (AI) and machine learning (ML) (Floridi et al., 2018) utilizing the principles of biomedical ethics: autonomy, beneficence, nonmaleficence, justice, along with seven others, namely, transparency, responsibility, privacy, trust, sustainability, dignity, and solidarity (Jobin et al., 2019). AI ethics, in particular, have attracted significant interest resulting in several sets of proposed principles (Hagendorff, 2020). While acknowledging ethics deriving from other fields, there is a need to select the most suitable principles for CBC on the basis that, although there is the aforementioned pool of potential principles, a practical approach for behavioral security interventions needs to capture the main and necessary principles.

In cybersecurity, the principles of autonomy, nonmaleficence/beneficence, fairness (justice), privacy, trust, and equality have been identified, from a business perspective (Yaghmaei et al., 2017). The systematic literature review of (Morgan and Gordijn, 2020) indicates as the most cited principle, followed by data protection, trust, control, accessibility privacy, confidentiality, responsibility of business, data integrity, consent, transparency, availability, accountability, autonomy, ownership, and usability, ranked according to the number of sources addressing these principles. However, the aforementioned principles are focused on threats to businesses, thus, have a different scope than CBC.

Other researchers have explored the ethics of Internet-of-Things (IoT) security, by utilizing the principles of autonomy and privacy, but without explicitly taking a principilism approach (Atlam and Wills, 2020). Ethics related to cybersecurity in healthcare (Weber and Kleine, 2020) are more relevant to our goals, since new healthcare technologies share: a targeted audience, the need for privacy preservation, and a component of persuading individuals to utilize these technologies.

In 1979 (first publication), Beauchamp and Childress proposed the *Principles of Biomedical Ethics,* a set of four principles for the, at the time, new field of biomedical ethics and healthcare practice. We select these principles because they constitute foundational work in ethics and *principlism*, i.e., ethical decision-making based entirely on a set of specific principles, which simplify decisions (Leikas et al., 2019). Moreover, these four principles, are continuously influencing fields with practical ethics, from nursing, to medical AI, disaster and emergency management, and forensic activities, and have become a widely accepted ethics basis (indicatively, Cuthbertson and Penney, 2023; Jahn, 2011). Additionally, the principles are interwoven with human rights (Brännmark, 2017), an attribute which we utilize for the principle of *privacy*. The selected principles are focused on individuals, which, in contrast to some of the aforementioned fields, are tailored to a unilateral communication of interventions from policy-makers and intervention-designers to users. In healthcare cybersecurity, principles have been suggested as pairs (Weber and Kleine, 2020); namely, the dual principles of privacy-trust, freedom-consent, dignity-solidarity, and fairness-equality are proposed as additional principles to the Beauchamp and Childress principles, indicating links between them.

In the next section we utilize the four ethical principles of Beauchamp and Childress (1989) because, unlike the literature focus on business, AI and other areas, such as warfare-oriented research, this is the closest framework to behavior change. We refer to these as the *core* principles of our approach, explain the need for them and how they are applied to CBC interventions, and accompany them with two additional principles. We explain the meaning and the potential usage of the principles, and showcase their applicability in cybersecurity via an application. The principles are not a tool for complex decision-making, but should be considered as the abstraction of a framework to provide fundamentally and instrumentally important considerations (Canca, 2020). We do not intend the six principles as a fixed or uniquely defined set, but more as representations of *clusters* of principles and notions, as we explain in the following sections, and as a means for security professionals to design, implement, and analyze CBC interventions to influence individual security behaviors.

4.1. An extended set of ethical principles for cybersecurity behavior change interventions

We propose the use of the following six ethical principles for CBC interventions: *autonomy, beneficence, nonmaleficence, justice, transparency,* and *privacy*. Our intention is not to promote principlism, but to provide a practical framework for security professionals to implement the principles in, e.g., organizational environments, and utilize them to resolve behavioral intervention dilemmas. The proposed set of principles could be extended with more principles, but we propose what we consider the minimum number of principles. Namely, we do not consider responsibility, trust, sustainability, dignity, and solidarity for the following reasons. In particular, *responsibility* can be the outcome of the way security processes are implemented (Durojaiye et al., 2021); thus it is broader than CBC. Responsibility is also linked to trust, since responsible users are trusted, but also, users take responsibilities in trustworthy environments. Then, *dignity* is partially captured under nonmaleficence and beneficence, and partially under justice. We select the latter principle, *justice,* as a broader notion which captures angles of responsibility, trust, and dignity, and one that can engender solidarity. *Trust*, or *digital trust*, although relevant, is not included as a principle, because, in the context of CBC, it refers to more advanced interventional stages; namely, it involves the degree of confidence in technologies, interactions, the environment and the designers themselves (Shipp et al., 2023). The same argument holds for *sustainability*, i.e., it goes beyond initial interventional steps, since it relates to further, long-term effects of behavioral interventions and their overall user acceptance and habituation.

The four core principles, along with *explicability,* have been used in cybersecurity, with explicability capturing the intelligibleness, comprehensibility, and transparency (Formosa et al., 2021). However, since the scope of our research is not cybersecurity broadly, but behavioral interventions, thus, we select *transparency* as opposed to explicability, because of the importance of perceived hidden intentions behind the interventions; a prohibiting condition for establishing digital trust (Shipp et al., 2023). The aforementioned principles of Formosa et al. (2021) have been examined by Fenech et al. (2024), but from the perspective of individuals without security knowledge.

The following principles are to be evaluated within their context of implementation. However, we advocate that the principles can be

utilized as an apparatus to assist security professionals and practitioners with identifying considerations and resolving ethical dilemmas when attempting to influence user behavior. Although the first four core principles are in name identical with those of Beauchamp and Childress (1989), they are repurposed and adjusted for CBC as per the following sections.

4.1.1. Principle 1: Autonomy

The first principle is respect for autonomy. Autonomy is the *freedom of individuals to determine whether an action is good or bad for them* (Varelius, 2006). The focus here is the extent to which behavior change affects an individual's ability to make (conscious) choices. Under autonomy, individuals are able to make decisions intentionally, without any controlling influences. A prerequisite for *intentional* decisions is that individuals have an understanding of the matters they decide on (Jahn, 2011). Thus, a necessary component is that users have access to the information needed in order to make a decision. If, e.g., a communicated solution encourages stronger passwords (assuming that the password policy requirements are met), then individuals need to know the 'why' and the 'how' to implement this solution. Subsequently, users can decide on whether to follow the advice or not.

Autonomy is related to deontological ethics. Kant states that we should treat others as ends and not as means under deontological ethics (Gillon, 1994). In other words, the goal of behavioral interventions is to persuade users in getting involved and ultimately aim in their long-term adoption of secure behaviors. Thus, autonomy requires honest communication from policy makers and security professionals to users, in an attempt to inform and persuade users to participate in the behavioral intervention. Autonomy under deontology has been identified as a required factor for perceiving cybersecurity as a *responsibility* amongst organizational employees (Posey and Folger, 2020).

Compliance and security policies, especially if they are mandatory, might create a sense of autonomy abolishment amongst users. This feeling is reinforced by the idea that, e.g., senior management or the IT department know what is best for the users and they try to impose it against the users' will. Autonomy protects users from manipulative interventions, if accompanied by informed consent and transparency. Under autonomy, users need to decide themselves and thus, be in a position to judge and decide whether to follow an intervention or not.

4.1.2. Principles 2 and 3: Beneficence and Nonmaleficence

The second and third principles are beneficence and nonmaleficence, respectively, and are often used together. The principles state that *benefits should be gained and harms should be avoided for the recipient of an action.* In the case of medical ethics there can be complications if an individual has to undergo a degree of 'harm' (e.g., a surgery) in order that a greater benefit is achieved (Beauchamp and Childress, 1989). In security, we suggest that the two principles are used together, especially in cases where beneficence bears the risk of causing harm to intervention recipients. For example, if a privacy risk is potentially introduced through an intervention, this should be done in a nonmaleficent, harm-avoiding fashion, and thus, in line with utilitarian ethics. In simple scenarios, beneficence equates with the protection of individual rights for users who undergo behavior change.

The two principles extend to *cultural* and *background* aspects. Users should have equal access to resources to understand the reasoning behind behavioral interventions (O'neil, 2016), i.e., benefits should be distributed appropriately. The extent to which an intervention is *just* in this sense is covered by the next principle. In organizational settings, access to resources is usually not the most challenging issue, but understanding diversified needs of users based on cultural and background differences might be more challenging. For example, different cultures have diversified tolerance for negative feedback or open disagreement at the workplace (Meyer, 2014).

A characteristic of nonmaleficence is that intervention designers should avoid offensive interventions. Consider, environments of 'blame and shame' in which employees who fail to comply with policies or, e.g., fail to recognize a company-orchestrated phishing campaign. In such environments, it can be argued that there is direct harm to individuals, psychological and/or in the form of sanctions. The principles, thus, ensure that interventions are respectful and considerate for individuals' well-being, devoid of distress and irritation.

The counter-argument, is that the overall benefit for the company (i.e., beneficence) outweighs individual harm. However, violating nonmaleficence for the sake of beneficence, is often problematic, amongst other reasons, due to the immeasurability of indirect intervention effectiveness, i.e., the overall achieved beneficence of a solution. Importantly, failure of security awareness campaigns is associated with a failure to adjust to the targeted individuals' culture (Bada et al., 2015), thus, an individual orientation is intervention-effective.

4.1.3. Principle 4: Justice

The fourth principle, justice, is conflated with the concept of fairness (Rawls, 2001). Justice does not mean equality, because interventions need to be individualized based on user characteristics (e.g., knowledge and skills), an attribute necessary for effective behavioral interventions. As an example, consider a phishing exercise and awareness campaign in a multi-national organization. If the exact same planning, messages and methods are used in countries with different *cultural, societal* and workplace *norms*, the campaign might be unsuccessful in some of these countries. The reasons might depend on the differences in locality, traditions, the role of authority, risk aversion, and other variables in each culture. To ensure that the principle of justice is applied successfully, security practitioners need to consider the specific characteristics of each environment. Similarly, individuality should be taken into account, so that each user is supported according to their 'merit' (Jahn, 2011; Meyer, 2014). In a cybersecurity context, the notion of 'merit' can represent the user's level of *digital literacy* and unique *skillset*.

Therefore, security professionals need to recognize, first, that not all users have the same background, needs, ability or motivation to act securely. In other words, behavioral interventions need to pertain individuality and also be culture-sensitive, as perceptions about the individual and the collective are shown to be significantly different via, e.g., eastern-western world dichotomies (Nisbett, 2004). And second, a sense of fairness needs to be conveyed to users. In a university setting, e.g., users can be part of senior management, academics, students or other staff. In the example scenario of phishing exercises, all users should be potential recipients of the phishing email, independently of their position, otherwise misperceptions and mistrust can emerge (Durojaiye et al., 2021). Since users are equally susceptible to attacks, hierarchy should not dictate the exclusion of individuals from the interventions (Levy, 2017). In that sense, CBC interventions can serve in the building of trust amongst groups, units and departments.

The principle of justice is also linked to an argument against the perception of humans as the weakest link in security. A just culture is described as a 'culture of trust, learning and accountability' which creates safe spaces for individuals (Dekker, 2018). In such a setting, individuals are free to admit personal errors and the organization transforms individual errors into learning opportunities for everyone. In many cases, it is not human error per se, but bad process design which renders human error and security breaches inevitable (Craggs, 2019; Craggs and Rashid, 2017). Justice expands from the design of interventions to handling intervention outcomes, e.g., instead of blaming individuals for security breaches, the underlying reasons can be investigated. Then, security professionals, such as CISOs, can revisit the individualized approach to train individuals into changing vulnerability-causing behaviors consciously, form habits, and eventually shape a stronger security culture.

4.1.4. Principle 5: Transparency

As a fifth principle we propose transparency. The principle relates to autonomy, because respect for autonomy might automatically imply

that users possess a good level of knowledge of an intervention's process. However, in many cases autonomy only refers to users' freedom of choice, whereas with transparency the 'recipients of behavioral interventions have knowledge of the intentions behind these interventions' (Elia, 2009). Thus, transparency concerns intervention designers more than users. For example, supervisory teams can be in place to disseminate the reasons behind a policy-compliance intervention.

Transparency does not hold for a number of behavioral interventions, in particular the ones which attempt to manipulate user choice and behavior (e.g., as in subliminal advertising), instead of encouraging reflective choice and persuasion. The fifth principle goes beyond users' understanding of intentions. It implies a *responsibility* for intervention designers and policy-makers to inform the targeted individuals about the intervention's goals. Such a requirement is not only ethics-related, but increases the chances of intervention success. In fact, one of the main reasons for resistance against behavioral interventions is the suspicion towards the hidden motivations behind the interventions and not against the interventions per se (Sunstein, 2016).

Thus, in order to achieve transparency, users need to become aware of the aims behind the policy and the intervention. These aims, or motivations, of the designers have been identified as a building component of *digital trust* (Shipp et al., 2023), i.e., technology-related trust, which is the case for the interventions at hand. Transparency portraying ethical leadership (at administrative and personal level), e.g., of designers, is also shown to instill compliance to employees (Xue et al., 2021). Additionally, transparency through clarity of intentions ensures the avoidance of deceptive and manipulative behavioral architectures, e.g., in organizational settings (Mersinas and Bada, 2023). And, finally, transparency allows users to provide feedback on security interventions. Many security professionals will agree that engaging in such feedback with users is an achievement in itself, as it can serve as a first step towards security awareness and behavior change.

4.1.5. Principle 6: Privacy

The final principle is the preservation of user privacy. Nowadays, virtually all user activities are potentially captured as behavioral data, even if in the form of metadata. There are concerns that companies and governments are able to build user profiles, identify preferences and trends, and promote either products or ideas, based on this metadata (Schneier, 2015; Zuboff, 2019). From a regulatory perspective, in the EU, data collection and processing fall under the General Data Protection Regulation (Anon., GDPR, 2018) and there are similar regulations in other regions. In the spirit of the regulation, individuals should, ultimately, have control of any personally identifiable information, be assured that they cannot be directly or indirectly identified from data, and be able to exercise their 'right to be forgotten', i.e., to have their data erased upon request. It should be noted that he aforementioned rights have emerged from a European standpoint on privacy, which is not necessarily shared globally. Under the privacy principle, we explain why security professionals need to be increasingly careful with handling behavioral data, beyond the legal requirements.

Observation of cybersecurity behaviors and identification of behavioral trends rely on data collection and processing. Moreover, these interventions need to be customized based on individual characteristics. Indicatively, behavioral variables, such as, self-efficacy, perceived level of risk, and impact, require user-specific information. It is beyond the purposes of this paper to propose solutions for data processing. However, we present the idea that anonymized user data can be *grouped* based on the value ranges of the variables of interest. Interventions can be then applied to groups instead of individuals and, moreover, we propose that individuals could *self-select* the appropriate group, thus, selecting a solution or intervention for themselves, assuming they are provided with the relevant information. An example of a target group for a CBC intervention can be technically not-savvy individuals, who would place themselves in a relevant intervention group once encouraged to identify their savviness level.

The parameter of *trust* mentioned in the principle of justice is important here too, as users or employees should not perceive data collection as a potential 'ubiquitous surveillance' (Anon., POSTnote, 2006) by employers or security professionals. Informed consent is a legal requirement,[d] providing users with the necessary information to render them able to accept an intervention or not. But, even if there is no legal mandate, informed consent, along with the aforementioned *transparency* of designer intentions, can assist with trust-building with users.

5. A survey on the perceptions of the ethical principles

To explore perceptions of our proposed set of principles, we designed a survey and improved question clarity and evaluated questions with input from a pilot conducted with security professionals. We utilized convenience sampling by circulating the survey via the LinkedIn platform from July 8th 2022 to February 19th 2023. This method was chosen because the authors have a joint available pool of >6,000 connections on this platform, most of whom are related to cybersecurity. The survey asked participants for feedback on the need for ethical frameworks in cybersecurity, in relation with CBC. Multiple-choice answers have been selected from previous surveys; indicatively, by the European Union Agency for Cybersecurity (ENISA), the UK Department of Culture, Media, and Sports in the UK, Ipsos Mori, and other respectable survey designers. All such questions include the text-box option 'Other' allowing participants to elaborate beyond the provided choices.

5.1. Respondent demographics and role

Regarding participants' demographics, there were 141 recorded non-empty responses; 114 males, 24 females, and 3 who refused to indicate their gender. Out of 141 respondents, 112 were directly or indirectly related to security (80 %), and 16 (11 %) diversified themselves, due to, e.g., working in the defense sector. Finally, 13 (9 %) participants stated they are not related to security; we did not exclude any participants from the sample. The majority of participants (68 %) were roughly split amongst the 31–40 (33 %) and 41–50 (35 %) age groups, and the average security experience was about 14.5 years (M = 14.41, SD=9.37). The professional roles of participants are depicted in Fig. 1.

5.2. Key results

The vast majority of participants (82 %) stated that more ethical frameworks are somewhat needed (42 % participants) or needed (40 %) in cybersecurity behavior change. In the question *'In which area(s) of cybersecurity are ethical principles currently not being considered?'* answers were distributed across all categories (from 12 % to 17 %, with multiple answers allowed), which indicates a view that overall, ethical principles are missing from the field. Namely, the categories are security awareness campaigns, privacy and personal data, communication between senior management and other teams, security risk management, confidentiality aspects, security policies and procedures, and security strategy, all used in previous surveys.

When asked to specify in which area(s) of cybersecurity ethics are mostly needed, participants indicated privacy and personal data (65 % of participants) as the top area, followed by security awareness campaigns (45 %), security policies and procedures (45 %), and confidentiality aspects (40 %); the area with the least need was indicated as security risk management (31 %).

In the question *'What would you like to see in security behavior change and ethics?'*, 23 % of participants responded 'Senior management getting involved with behavior change and/or ethics' and the second most

[d] Although, there are special cases in GDPR (e.g. Article 6) for employer-employee contractual relationships.

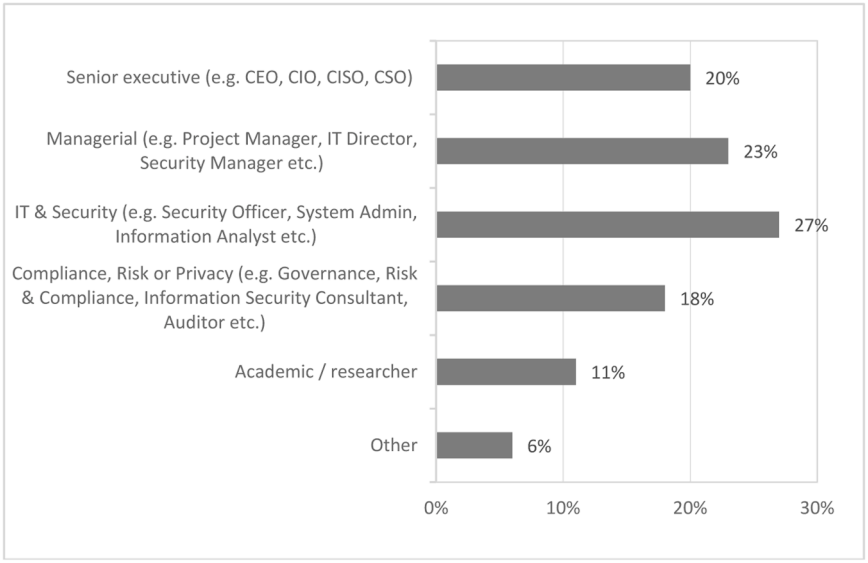

Fig. 1. Participants' professional roles.

popular answer (22 %) was 'End users getting involved with behavior change and/or ethics' (multiple selection across eight answers was possible, including an open answer 'Other').

In accordance, 'end-users having other priorities' and 'senior management having other priorities', are the two major concerns expressed by participants with regards to behavior change and ethics (27 % and 26 % of responses, respectively). The concern expressed the least is that 'the security field is not mature enough' (8 %), whereas only 2 % of participants reported no concerns for behavior change and/or ethics.

Participants were provided with the following contextual information and definitions:

'Consider reminders, prompts or examples that an ***organisation communicates*** *to users about* ***good security practices****, for example, strong passwords, data protection, physical security, or any other practice on security hygiene.*

'Behaviour change' refers to any modification of human behavior through an intervention designed, e.g., *by security professionals or policy makers; an example is communicating reminders to users regarding strong passwords.*

Cybersecurity behaviour changes can increase security hygiene, and reduce the attack surface and human errors.

We propose a six-principle ethical framework for cyber behaviour change interventions:

Principle 1 - Autonomy*: individuals are free to accept or reject the intervention*
Principles 2 & 3 - Beneficence and Nonmaleficence*: interventions benefit and do not harm individuals*
Principle 4 - Justice*: individuals are supported according to their culture, digital literacy and skillset*
Principle 5 - Transparency*: recipients of behavioral interventions know the intentions behind these interventions*
Principle 6 - Privacy*: individuals have control of personally identifiable information and/or cannot be identified from data.'*

The question that followed was *'Which of the six principles are useful / needed in cybersecurity behavior change ethics in your opinion?'*, for which, interestingly, Principle 1 (Autonomy) was considered as the least useful/needed one, namely, 52 % stated it as not useful/needed, compared to 6 %, 4 %, 2 %, and 4 % for principles 2 and 3, 4, 5, and 6, respectively. We also asked participants to rank the principles from the most important (numerical value 1) to the least important (value 5). The most important principle (45 %) was Principle 5 Transparency (μ=2.36, SD=1.13), followed by Principles 2 and 3 Beneficence and Nonmaleficence (μ=2.61, SD=1.32), and then Principle 6 Privacy (μ=2.7, SD=1.34). Principle 4 Justice followed, and the least important principle was reported as Principle 1 Autonomy.

This noteworthy perception on the non-importance of user autonomy was confirmed by the level of agreement to the statement *'Users should be free to accept or reject the security practices'*, where 51 % responded 'I disagree' and 29 % stated 'I somewhat disagree' on a 5-Likert scale. In contrast, the vast majority of participants agreed or somewhat agreed with the other 5 principles with the stronger disagreement being that of a 6 % on the Justice principle, i.e., as a response to the statement *'Users need to be supported to follow security practices according to their culture, digital literacy and skill set'*. It is also noteworthy that no participants expressed any level on disagreement against the Transparency principle described by the statement *'Users should know the intentions behind security practices'* (Fig. 2).

Overall, 82 % of participants stated that ethical frameworks are needed (40 %) or somewhat needed (42 %) in CBC.

Finally, we asked participants about their suggestions, namely, *'Is there any other ethical principle or consideration that you would like to propose or mention?'*. We did not conduct qualitative analysis on the responses, but some of them are worth mentioning as indicative examples. One participant stated:

'*Better training at all levels, helping all end users attain the education and emotional maturity to use security frameworks with confidence. Unfortunately, a lot of cybersecurity training these days is inaccessible to the people who need it most because of the barriers to accessing education, the format of the education, and the fact that the education is rarely tailored to their needs. In areas with poor digital access, literacy, and resources, users are the most vulnerable to cyber-attacks.*'

Others expressed a need for alignment of the corporate mission with personal goals, and the need to include aspects of biometrics and artificial intelligence.

6. Utilizing and discussing the ethical principles

In this section, we employ the widely used application of password strength checkers, which are found to influence behavior (Zimmerman and Renaud, 2021), to showcase an analysis through the six clusters of ethical principles. We also provide a discussion on the principles' relationships and interpretations of key survey results.

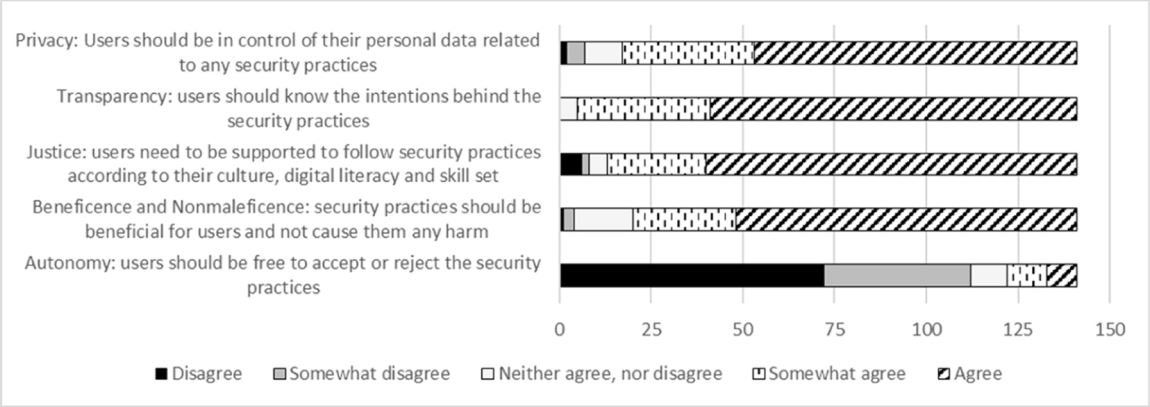

Fig. 2. Agreement with the principles. *'Please state the extent to which you agree or disagree with the following statements.'*.

6.1. An example of using the ethical principles

A simple yet representative example of a behavioral intervention is a password strength checker, or meter, which conveys visual color-coded or numerical information to users in real-time (Fig. 3). As users type a (new) password, the meter indicates the password's strength in terms of entropy, a measure of (pseudo)randomness and unpredictability, and subsequently provides recommendations on increasing strength, e.g., via the inclusion of additional symbols, lower/upper case letters, numbers, and/or increasing the password length. The meter provides a score according to the estimated difficulty in brute-forcing the password (i.e., its strength), which is positively correlated with the level of its entropy. The intervention is assumed to remain informational and not enforced to users.

The above intervention can be analyzed across the six principles. First, the user is notified that a password might be weak, but is free to decide whether to choose it or not, thus, the intervention maintains user autonomy. It is not unusual to provide additional warnings when users take insecure actions; unless user choices are obstructed, *autonomy* is maintained. The principle of *nonmaleficence* is also satisfied, as the intervention does not cause harm. The intervention design promotes *beneficence*, because the user is advised with the best course of action for their security, i.e., to choose a strong(er) password. The principle of justice is satisfied if the application is free and accessible to all users of the relevant service. The principle of transparency is more complex, however. *Transparency* requires that users understand the portrayed *intention* behind the intervention, i.e., *why* stronger passwords are needed. *Explainability* needs to be captured through clarity of the underlying mechanisms, e.g., a level of understanding on the strength-evaluation, and the meaning of the color codes, the numerical values, any additional information provided. Then, the how, when, and where the intervention is provided needs to instill *trust* to users. Finally, to maintain *privacy*, the intervention should not collect user data (or metadata), and this message would need to be communicated to users.

Fig. 3. A behavioral intervention in the form of a visual nudge, e.g., with a password strength meter ranging from 'weak' to 'strong', for increasing password strength.

6.2. Discussion

Arguably, the principles' boundaries are not clear-cut. For example, respect for *autonomy* is linked with *transparency* and *privacy*, since in order to exercise autonomy users need to be informed, assured about their rights, and provide their consent. Universally applied *non-maleficence* and *beneficence* are prerequisites for fairness, and thus, *justice*. Then, *justice* and *transparency* can lead to further constructs, such as building *trust*, and through the exercising of autonomy, promoting *responsibility*. In an organizational environment, it is important that the reasoning behind these principles is understood by and communicated to users/employees, for building (digital) trust.

Additionally, a combination of the aforementioned principles, i.e., *perceived* autonomy, nonmaleficence, beneficence, justice, privacy, intention, explainability, and transparency, all assist in shaping *trust* from users to intervention designers. If, e.g., users or employees suspect covert data collection, or intent for blaming or sanctioning, they might be reluctant to use a proposed application.

Interventions, on the other hand, which lack transparency might be occasionally unsuccessful because users (at least within individualistic cultures) tend to distrust mechanisms which undermine their autonomy. More specifically, people are observed to be more responsive to interventions which address deliberation and autonomy (Levy, 2017). Additionally, if user choices are not reflective and intentional, the effects of an intervention might be short-term, disallowing habitual behavior formation. This is sub-optimal in terms of shaping a security culture or making security more sustainable.

The intended conceptualization behind proposing these principles is highlighted as a catalyst for fostering critical thinking and encouraging contextual analysis regarding the ramifications of violating the ethical principles. Thus, the principles are not to be downgraded into a 'checklist' but can indicate context-dependent solutions, e.g., intervention designers might consider introducing credibility and reputation to eliminate distrust, concepts which are not included in the principles, but can be derived from them.

One of the key findings of the survey is that, although the five principles of privacy, transparency, beneficence, nonmaleficence, and justice are all perceived as useful or needed in cybersecurity by >93 % of participants, the principle of autonomy, defined as 'the freedom of individuals to determine whether an action is good or bad for them' is perceived differently. In particular, 86 % of participants do not agree that users need to maintain the freedom to follow security practices. This stance might indicate a traditional view in the field, and might be related to the argument that autonomous and intentional decision-making requires users' understanding of the context (Jahn, 2011); therefore, the identified view might indicate disbelief that users are knowledgeable

and able to make decisions. In that case, it would be education and training which should be enhanced. Subsequently, this identified view might indicate that participants directly diminish the need for autonomy due to a lack of confidence in existing security awareness training approaches. It is noteworthy that cultural and personal characteristics are largely absent from such approaches, a component which, if considered, might also shift the expressed view of participants. Finally, the finding might reflect a realization by security-related individuals that user security behavior can be sub-optimal and insecure, due to cognitive limitations or the hurdles that users face (Furnell, 2010). But, it should be stated that security professionals themselves manifest similar sub-optimality (Mersinas et al., 2015).

CBC ethical principles might be *prima facie*, depending on the context, i.e., a principle can be binding unless two principles are conflicting, in which case the one with the most importance is followed. We have already mentioned a conflict between overall beneficence and individual nonmaleficence. Another case is the potential conflict between beneficence and privacy. First, the preservation of individual privacy might be considered as equivalent to nonmaleficence. Then, in real-world situations, decisions which violate privacy might need to be made, exceptionally, but for the benefit of the individual. In such a scenario, if the individual benefit (say, the individuals health and safety) dominates privacy preservation, then exceptional violation of a principle might be acceptable. Notably, such a scenario can be in line with virtue ethics – as a contextual interpretation of the decision – and most certainly, with utilitarianism, if the decision maximizes utility. The latter point reinforces our position that the proposed ethical principles are used according to the judgement of security professionals, considering the situational and contextual parameters of an intervention.

7. Limitations and future work

The sample size of our survey ($N = 141$) is a potential limitation. We used convenience sampling and aimed in maximizing the reliability of our findings by narrowing the targeted audience mainly to cybersecurity-related individuals.

We have not conducted a systematic literature review on the topic of ethical principles, since, to the best of our knowledge, there is a lack of frameworks for CBC. Instead, we selected the most appropriate ethical principles from biomedical sciences, and validated their use through similar, but not directly relevant, research in cybersecurity. Although the approach was selected to serve the research goals, it might pose limitations in the generalizability of the principles.

We also did not attempt a formalization of the principles and their interdependencies, as this was beyond the scope of our research. However, we cannot ignore the precision, additional generalizability, and potential predictive power of such formal models.

For future research, we plan to explore more representative samples, including the expansion to non-security-related individuals to allow for comparisons. Moreover, we would like to further explore the degree to which ethical principles are culture-dependent, so we would like to test the acceptability and relative importance of the ethical principles across samples from, e.g., from Northern/Central Europe, Mediterranean countries, US, and Japan, given that security behaviors have been statistically associated with cultural variables (de Bruin and Mersinas, 2024). Empirical research, albeit with student samples and mostly from Oceania and Southeast Asia, provides supportive evidence for this approach (Sadeghi et al., 2023). We also plan to link our parallel work on the perceptions of digital trust with the perceived importance of ethical principles, since digital trust is a complex and desirable construct connecting and transcending ethical principles.

8. Conclusion

In this paper we presented the ethical issues surrounding cybersecurity behavior change (CBC) and we suggest a set of representative ethical principles for behavioral interventions linked to three ethical traditions. Our approach places weight on the individual, as the vessel for further 'common good'. The aim is that through an ethical foundation for behavioral interventions we can tackle issues related to user mistrust and create stronger security cultures, by maintaining user autonomy and privacy, ensuring the transparency of behavioral interventions, respecting and utilizing individual differences, avoiding negative impact, and promoting the benefit of users. We evaluated six clusters of ethical principles via a survey targeted mainly at security-related professionals and practitioners. We identified a consensus on the importance of all ethical principles, except for autonomy. Our ultimate goal is that the proposed conceptualization is utilized as an apparatus for designing and evaluating behavioral interventions in cybersecurity.

CRediT authorship contribution statement

Konstantinos Mersinas: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Maria Bada:** Writing – review & editing, Methodology. **Steven Furnell:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Adams, A., Sasse, M.A., 1999. *Users are not the enemy*. Communications of the ACM 42 (12), 40–46.

Ajzen, I., 1980. Understanding attitudes and predictiing social behavior. *Englewood cliffs*.

Alnifie, K.M., Kim, C., 2023. Appraising the manifestation of optimism bias and its impact on human perception of cyber security: a Meta analysis. J. Inform. Secur. 14 (02), 93–110.

Ariely, D., 2008. Predictably irrational: The hidden forces that shape our decisions. New York.

Aristotle, J.O., Urmson, J.L., Ackrill, 1983. Nichomachean Ethics. In: Ross, W.D., Urmson, J.O., Ackrill, J.L. (Eds.). Oxford University Press, pp. 123–149.

Atlam, H.F., Wills, G.B., 2020. IoT security, privacy, safety and ethics. Digital Twin Technol. Smart Cities 123–149.

Bada, M., Sasse, A., Nurse, J.R.C., 2015. Cyber security awareness campaigns: why do they fail to change behavior?. In: International Conference on Cyber Security for Sustainable Society.

BCS, 2024. Code of conduct. Available at: https://www.bcs.org/membership-and-registrations/become-a-member/bcs-code-of-conduct/(Accessed: 12 June 2024).

Beauchamp, T.L., Childress, J.F., 1989. Principles of Biomedical Ethics. Oxford.

Bednar, K., Spiekermann-Hoff, S., 2020. The power to design: exploring utilitarianism, deontology and Virtue Ethics in three technology case studies. In: ETHICOMP 2020, p. 396.

Bentham, J., 1876. An Introduction to the Principles of Morals and Legislation. Clarendon Press, Oxford.

Brännmark, J., 2017. Respect for persons in bioethics: towards a human rights-based account. Human Rights Rev. 18 (2), 171–187.

Briggs, P., Jeske, D., Coventry, L., 2017. Behavior change interventions for cybersecurity. In: Little, L., Sillence, E., Joinson, A. (Eds.), Behavior Change Research and Theory. Elvesier, Amsterdam, pp. 115–136.

Camerer, C., 2003. Behavioral game theory: experiments in strategic interaction. New York.

Camerer, C. F., 2004. Prospect theory in the wild: Evidence from the field. In: Camerer, C. F., Loewenstein, G., Rabin, M. (Eds.), Advances in behavioral economics. Princeton and, Oxford, pp. 148-161.

Canca, C., 2020. Operationalizing AI ethics principles. Commun. ACM 63 (12), 18–21.

Caraban, A., Karapanos, E., Gonçalves, D., Campos, P., 2019. 23 ways to nudge: a review of technology-mediated nudging in human-computer interaction. In: CHI 2019.

Chowdhury, N.H., Adam, M.T., Teubner, T., 2020. Time pressure in human cybersecurity behavior: theoretical framework and countermeasures. Comp. Secur. 97, 101963.

Cisco. 2017. Annual cybersecurity report.

Coventry, L., Briggs, P., Jeske, D., & van Moorsel, A., 2014. SCENE: a structured means for creating and evaluating behavioral nudges in a cyber security environment. In A. Marcus (Ed.), Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience, pp. 229–239.
Craggs, B., Rashid, A., 2017. Smart cyber-physical systems: beyond usable security to security ergonomics by design. In: Proceedings of the 3rd International Workshop on Software Engineering for Smart Cyber-Physical Systems. IEEE Press, pp. 22–25.
Craggs, B., 2019. A just culture is fundamental: extending security ergonomics by design. In: 2019 IEEE/ACM 5th International Workshop on Software Engineering for Smart Cyber-Physical Systems (SEsCPS), pp. 46–49.
CREST, 2024. Code of ethics. Available at: https://www.crest-approved.org/about-us/code-of-ethics/(Accessed: 12 June 2024).
Cuthbertson, J., Penney, G., 2023. Ethical decision making in disaster and emergency management: a systematic review of the literature. Prehosp. Disas. Med 1–6.
de Bruin, M. and Mersinas, K., 2024. Individual and Contextual Variables of Cyber Security Behaviour–An empirical analysis of national culture, industry, organisation, and individual variables of (in) secure human behaviour. *arXiv preprint https://arxiv.org/pdf/2405.16215*.
Dekker, S., 2018. Just culture: Restoring Trust and Accountability in Your Organization. CRC Press.
Dhingra, N., Gorn, Z., Kener, A., Dana, J., 2012. The default pull: an experimental demonstration of subtle default effects on preferences. Judgm. Decis Mak 7 (1), 69–76.
Dittrich, D., Kenneally, E., 2012. The Menlo report: Ethical Principles Guiding Information and Communication Technology Research. US Department of Homeland Security.
Durojaiye, T., Mersinas, K., Watling, D., 2021. What influences people's view of cyber security culture in higher education institutions? An empirical study. In: The Sixth International Conference on Cyber-Technologies and Cyber-Systems, CYBER 2021. Barcelona, Spain.
Elia, J., 2009. Transparency rights, technology, and trust. Ethics Inf. Technol 11, 145–153.
Fenech, J., Richards, D., Formosa, P., 2024. Ethical principles shaping values-based cybersecurity decision-making. Comp Secur, 103795.
Fishbein, M., Ajzen, I., 1975. Belief, attitude, intention, and behavior: An introduction to theory and research. Reading, MA.
Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., 2018. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach 28, 689–707.
Formosa, P., Wilson, M., Richards, D., 2021. A principlist framework for cybersecurity ethics. Comp Secur 109, 102382.
Furnell, S., 2010. Jumping security hurdles. Comp Fraud Secur 2010 (6), 10–14.
General Data Protection Regulation. 2018. Retrieved from gdpr-info.eu.
Georgiadou, A., Mouzakitis, S., Askounis, D., 2022. Detecting insider threat via a cyber-security culture framework. J. Comp. Inform. Syst. 62 (4), 706–716.
Gillon, R., 1994. Medical ethics: four principles plus attention to scope. Br Med. J 309 (6948), 184–188.
Hagendorff, T., 2020. The ethics of AI ethics: an evaluation of guidelines. Minds Mach. 30 (1), 99–120.
Hastings, G., Stead, M., Webb, J., 2004. Fear appeals in social marketing: strategic and ethical reasons for concern. Psychol. Market. 21 (11), 961–986.
Anon. ISC2, 2024. Code of ethics. Available at: https://www.isc2.org/ethics (Accessed: 12 June 2024).
Jahn, W.T., 2011. The 4 basic ethical principles that apply to forensic activities are respect for autonomy, beneficence, nonmaleficence, and justice. J. Chiropr. Med. 10, 225–226.
Jobin, A., Ienca, M., Vayena, E., 2019. Artificial intelligence: the global landscape of ethics guidelines. Nat. Machine Intellig. 1, 389–399.
Kahneman, D., 2011. Thinking fast and slow. Allen Lane and Penguin Books, New York.
Kant, I., 1998. Groundwork of the Metaphysics of Morals [1785]. Cambridge University Press.
Leikas, J., Koivisto, R., Gotcheva, N., 2019. Ethical framework for designing autonomous intelligent systems. J. Open Innov. 5, 12.
Levy, N., 2017. Nudges in a post-truth world. J. Med. Ethics 43 (8), 495–500.
Loi, M., Christen, M., 2020. Ethical frameworks for cybersecurity. The Ethics of Cybersecurity, pp. 73–95.
Münscher, R., Vetter, M., Scheuerle, T., 2016. A review and taxonomy of choice architecture techniques. J. Behav. Decis. Mak 29 (5), 511–524.
Mersinas, K., Bada, M., 2023. Behavior change approaches for cyber security and the need for ethics. In: *Proceedings of the International Conference on Cybersecurity, Situational Awareness and Social Media: Cyber Science 2023;* 3-4 July. Copenhagen, Denmark. Springer.
Mersinas, K., Hartig, B., Martin, K.M., Seltzer, A., 2015. Experimental elicitation of risk behaviour amongst information security professionals. In: the *Workshop on the Economics of Information Security* (*WEIS)*. Delft, Netherlands.
Mersinas, K., Hartig, B., Martin, K.M., Seltzer, A., 2016. Are information security professionals expected value maximizers?: An experiment and survey-based test. J. Cybersecur. 2 (1), 57–70.
Mersinas, K., Sobb, T., Sample, C., Bakdash, J.Z., Ormrod, D., 2019. Training data and rationality. In: Proceedings of the European Conference on the Impact of Artificial Intelligence and Robotics, p. 225.
Meyer, E., 2014. The Culture map: Breaking through the Invisible Boundaries of Global Business. Public Affairs.
Mill, J.S. 1859. Utilitarianism. London.
Morgan, G., Gordijn, B., 2020. A care-based stakeholder approach to ethics of cybersecurity in business. In: Christen, M., Gordijin, B., Loi, M. (Eds.), The Ethics of Cybersecurity. Springer Open, pp. 119–138.
Neigel, A.R., Claypoole, V.L., Waldfogle, G.E., Acharya, S., Hancock, G.M., 2020. Holistic cyber hygiene education: accounting for the human factors. Comp. Secur. 92, 101731.
Nisbett, R. 2004. The geography of thought: how Asians and Westerners think differently - and why. London.
O'Neil, C., 2016. Weapons of math destruction: how big data increases inequality and threatens democracy. New York.
Ponemon Institute, 2019. Cost of a Data Breach Report. IBM Security, North Traverse City, MI, USA.
Posey, C., Folger, R., 2020. An exploratory examination of organizational insiders' descriptive and normative perceptions of cyber-relevant rights and responsibilities. Comp. Secur. 99, 102038.
POSTnote, 2006. Parliamentary Office of Science and Technology. Pervasive Computing.
Pruemmer, J., van Steen, T., van den Berg, B., 2023. A systematic review of current cybersecurity training methods. Comp. Secur., 103585
Rawls, J., 2001. Justice As fairness: A restatement. Harvard University Press.
Renaud, K., Dupuis, M., 2019. Cyber Security fear appeals: unexpectedly complicated. In: New Security Paradigms Workshop (NSPW '19), September 23-26 2019.
Sadeghi, B., Richards, D., Formosa, P., McEwan, M., Bajwa, M.H.A., Hitchens, M., Ryan, M., 2023. Modelling the ethical priorities influencing decision-making in cybersecurity contexts. Organiz. Cybersecur. J. 3 (2), 127–149.
Safi, R., Browne, G.J., Naini, A.J., 2021. Mis-spending on information security measures: theory and experimental evidence. Int. J. Inf. Manage 57, 102291.
Schneier, B. 2015. Data and Goliath. The hidden battles to collect your data and control your world. New York.
Shipp, L., Mersinas, K., Mushtaq, H., Panteli, N., 2023. Digital Trust: A Literature Review. UK Academy for Information Systems.
Simon, H.A., 1972. In: McGuire, C.B., Radner, R. (Eds.), Theories of bounded rationality. Decision and Organization, pp. 161–176.
Simon, H.A., 1991. Bounded rationality and organizational learning. Organization Science 2 (1), 125–134.
Sunstein, C.R. 2016. The ethics of influence. Government in the age of behavioral Science. New York.
Thaler, R.H., Sunstein, C.R., 2008. *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press, New Haven.
Van Schaik, P., Renaud, K., Wilson, C., Jansen, J., Onibokun, J., 2020. Risk as affect: the affect heuristic in cybersecurity. Comp. Secur. 90, 101651.
Van Staveren, I., 2007. Beyond utilitarianism and deontology: ethics in economics. Rev. Political Econ. 19 (1), 21–35.
Varelius, J., 2006. The value of autonomy in medical ethics. Med. Health Care Philos. 9 (3), 377–388.
Vishwanath, A., Neo, L.S., Goh, P., Lee, S., Khader, M., Ong, G., Chin, J., 2020. Cyber hygiene: the concept, its measure, and its initial tests. Decis. Support Syst 128, 113160.
Weber, K., Kleine, N., 2020. Cybersecurity in health care. In: Christen, M., Gordijin, B., Loi, M. (Eds.), The Ethics of Cybersecurity. Springer Open, pp. 139–153.
Xue, B., Xu, F., Luo, X., Warkentin, M., 2021. Ethical leadership and employee information security policy (ISP) violation: exploring dual-mediation paths. Organ. Cybersec. J. 1 (1), 5–23.
Yaghmaei, E., et al., 2017. Canvas white paper 1 – cybersecurity and ethics. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3091909. Last access 7 July 2019.
Zimmermann, V., Renaud, K., 2021. The nudge puzzle: matching nudge interventions to cybersecurity decisions. ACM Transac. Computer-Human Interaction (TOCHI) 28 (1), 1–45.
Zuboff, S. 2019. The age of surveillance capitalism. London.

Dr. Konstantinos Mersinas, PhD, CISSP, is an Associate Professor at the Information Security Group, Royal Holloway, University of London. Konstantinos worked in various industry cybersecurity roles before moving to academia. His research interests lie with human and behavioural aspects of cybersecurity, maritime security, and cybercrime. He has repeatedly advised the UK All-Party Parliamentary Group (APPG) on Cybersecurity, the UK Fraud Act and Digital Fraud Committee, and a number of UK Government Departments. He is co-founder of the research group HIVE (Hub for Interdisciplinary research into Vulnerability to Exploitation). Konstantinos collaborates with the NATO Cooperative Cyber Defence Centre of Excellence (https://ccdcoe.org) in Tallinn, Estonia, and the Pukhov Institute in the Ukraine. He is Director at the International Cyber Security Center of Excellence (https://incs-coe.org), an international centre founded between UK, USA, and Japan, that promotes cybersecurity research globally.