Ethical Frameworks for Cybersecurity

Loi, Michele ; Christen, Markus

DOI: https://doi.org/10.1007/978-3-030-29053-5_4

Chapter 4
Ethical Frameworks for Cybersecurity

Michele Loi and Markus Christen

Abstract This chapter presents several ethical frameworks that are useful for analysing ethical questions of cybersecurity. It begins with two frameworks that are important in practice: the principlist framework employed in the Menlo Report on cybersecurity research and the rights-based principle that is influential in the law, in particular EU law. It is argued that since the harms and benefits caused by cybersecurity operations and policies are of a probabilistic nature, both approaches cannot avoid dealing with risk and probability. Therefore, the chapter turns to the ethics of risk, showing that it is a necessary complement to such approaches. The ethics of risk are discussed in more detail by considering two consequentialist approaches (utilitarianism and maximin consequentialism), deontological approaches and contractualist approaches to risk at length, highlighting the difficulties raised by special cases. Finally, Nissenbaum's 'contextual integrity' approach is introduced, which has become an important framework for understanding privacy, both descriptively and normatively. A revised version of this framework is proposed for identifying and ethically assessing changes brought about by cybersecurity measures and policies, not only in relation to privacy but more generally to the key expectations concerning human interactions within the practice.

Keywords Consequentialism · Contextual integrity · Cybersecurity · Ethics of risk · Human rights · Principlism

M. Loi (✉)
Digital Society Initiative, University of Zurich, Zurich, Switzerland

Institute of Biomedical Ethics and History of Medicine, Zurich, Switzerland
e-mail: michele.loi@uzh.ch

M. Christen
UZH Digital Society Initiative, Zürich, Switzerland
e-mail: christen@ethik.uzh.ch

M. Christen et al. (eds.), *The Ethics of Cybersecurity*, The International Library
of Ethics, Law and Technology 21,
https://doi.org/10.1007/978-3-030-29053-5_4

4.1 Introduction

The term cyber*security* explicitly conveys its main ethical goal, namely to create a state of being free from danger or threat in cyberspace, if we follow the general definition of the English term 'security' (Oxford Dictionary). However, in ethics, the concept of security rarely plays a central role in theory building. For example, if we search the *Stanford Encyclopedia of Philosophy* for 'security', the term only appears in the entry under information ethics (which is the context that interests us here) and in political philosophy, referring to the security of nation states. This is remarkable, as from a purely biological perspective, organisms (and groups of social animals) invest considerable resources in protecting themselves against threats. Certainly, conditions resulting from insecurity such as harm or injustice are central topics in ethical theorising. Nevertheless, the positive orientation aimed to overcome those conditions refer to values like justice or benevolence, not security (probably with the exception of social security).

Why is this? One reason could be that the term 'security' used in a more general sense has certain negative connotations, particularly within ethics. These may refer to the problems that result when security is enforced by states through coercive capacities, to the observation that authoritarian regimes often rely on security when actually promoting injustice, or to the more general impression that a state of security involves a static and closed setting of societies. In that sense, within moral theory security is usually not an ethical value of its own, but rather an *instrumental value* to protect *ethical values* (but see also the considerations in Chap. 3) Thus, as an instrumental value, security can also be unethical, when either the protected goals or the means used to establish security are unethical. The same holds for cybersecurity.

Cybersecurity, understood broadly, is usually considered as a whole bundle of technologies and policies to protect the cyber-infrastructure. Following Hildebrandt (2013), we can distinguish three main classes of technology for cybersecurity: technologies that ensure confidentiality of information (including authentication of the intended recipients of communication); technologies that detect and counter online threats and vulnerabilities; and technologies that detect and counter cybercrime such as forgery, fraud, child pornography and copyright violations committed in cyberspace. In each of those application domains, different ethical problems emerge.

Given that cybersecurity is by itself not a genuine ethical value, we may pose a follow-up question of how to analyse the ethical questions raised by enforcing cybersecurity. In this chapter, we present several ethical frameworks useful for analysing ethical questions that arise in the context of cybersecurity. We start with two frameworks that are important in practice: the principlist framework employed in the Menlo Report on cybersecurity research (Sect. 4.2) and the rights-based principle that is influential in the law, in particular EU law (Sect. 4.3). We show that since the harms and benefits caused by cybersecurity operations and policies are often probable, rather than certain, both approaches cannot avoid dealing with risk and probability. Therefore, we turn to the ethics of risk, demonstrating that it is a necessary complement to such approaches (Sect. 4.4). Section 4.5 considers the ethics of risk in more detail by considering at length two consequentialist approaches (utilitarianism and maximin consequentialism), deontological approaches and contractualist approaches to risk, highlighting the difficulties raised by special cases. Finally, in Sect. 4.6, we introduce Nissenbaum's 'contextual integrity' approach and extend it to address all the human interactions (and not only informational exchanges) affected by new cybersecurity applications.

4.2 Principlism

The Menlo report was intended to guide research in cybersecurity, understood traditionally as a form of investigation aimed at generalisable knowledge for the benefit of society, and *in so far as it deals with human subjects*. However, it can also be applied more broadly to cybersecurity operations that involve a research component, e.g. acts of inspections and the collection of intelligence, such as those carried out by computer emergency response teams, if there is direct interaction with a human or if there are human data (Johnson, Bellovin, and Keromytis 2011). Cybersecurity—"the subdiscipline of computer science concerned with ensuring simultaneously the confidentiality, integrity, and availability of IT systems against the attacks of some set of adversaries" (Spring and Illari 2018, para. 1) can arguably produce general knowledge (Spring and Illari 2018) of a particular form. The general knowledge produced does not take the form of scientific theories, rather the discovery and modelling of peculiar *mechanisms* (e.g. mechanisms that disrupt the intended working of an information system). This knowledge of mechanisms provides, in the long run and in a patchwork way, cybersecurity experts with general knowledge on how to detect and respond to information security challenges, and how to improve cybersecurity defences (Spring and Illari 2018).

Principlism is a system of ethics based on a limited number of principles (usually 3 or 4) with a grounding in common-sense morality and professional ethical practice (see also Chap. 7). An instance of principlism is the Belmont Report for the protection of human research subjects, which includes three principles: Respect for Persons, Beneficence, and Justice. The Menlo Report (US Department of Homeland Security Science and Technology Directorate) adapted this approach to the context of Information and Communication Technology Research (Kenneally et al. 2010; Kenneally and Bailey 2013), using the same principles and highlighting ways of applying them to the cybersecurity domain.

Principlism is a form of deontology (deontology = the study of duty). The main principles of the theory can be regarded as the sources of prima facie duties in the sense of W.D. Ross (2002). According to Ross, an action's moral rightness cannot be explained in terms of its being productive of the good; rather, it should be analysed by considering prima facie duties. For example, if I fulfil my promise to you, what makes it *right* that I do so is not the consequences of fulfilling my promise but rather the fact that I promised. Of course, this is not to imply that I should respect my promise even when this would produce disastrous consequences. The way Ross explains this is by claiming that the duty to 'respect one's promises' is not *the only* duty and it is only a prima facie duty. A person also has a duty to *relieve distress,* which (in certain situations) may override the duty to keep one's promise. The prima facie duty to keep one's promise makes it *right* to keep one's promise if it is a stronger prima facie duty than conflicting prima facie duties, or if there are no other prima facie duties. The theory of prima facie duties is an alternative to the consequentialist theory that all conflicts of duties should be resolved by asking which action produces the most good. Instead, with prima facie duties there is no higher-order theory to determine how conflicts of duties are to be resolved.

It is not difficult to see that the logic of Ross's prima facie duties can be applied to principlism. The three (or four) principles in principlism can be regarded as prima facie duties: from the moral point of view, we *always* have good reasons to respect persons, to pursue the good of others, to avoid harming them, and to act justly in the absence of countervailing considerations. However, in practice, the duties implied by those principles may conflict and, when this happens, the principles must be balanced against each other. In the tradition of principlism, the balance of different duties occurs according to intersubjective agreements that, as in prima facie duties theories, are not theoretically predetermined in advance.

The principlist approach is a modest, minimalist framework that affords significant flexibility. It leaves to the researchers, or cybersecurity operatives, the difficult task of identifying the specific factors and circumstances that should carry weight in deliberations concerning a concrete case and the even more difficult task of weighing these considerations against each other when trade-offs occur.

Let us now briefly introduce the three principles of the Menlo Report. Respect for persons concerns all those cases in which data may be linked with identifiable persons, e.g. data concerning communication between individuals or IP addresses which may be linked to individuals. Respect also involves all research in which consent *can* be asked and in which it is realistically considered a necessary condition of research, for example some forms of experimental (psychological) research on human factors in cybersecurity, performed in the lab with research subjects recruited for that purpose (e.g. Hadlington 2017). One area of cybersecurity research that involves such methods is research on human factors of cybersecurity, which includes the experimental study of user acceptance, confusion, frustration, cognitive workload, error/risk reduction and the optimisation of error-tolerant systems (Boyce et al. 2011). Realistically, however, consent is often impracticable; in such contexts, the principle of beneficence may be the basis of a duty to do research when the cost-benefit ratio clearly favours it (Kenneally et al. 2010). The benefit principle applies in all generality to cybersecurity research; it should be understood as the principle of maximising probable benefit and minimising probable harm. Minimising harm also requires considering the full spectrum of risks to persons, including reputational, emotional, financial and physical harm (Kenneally et al. 2010). Justice involves a distributive aspect, concerning the fair distribution of the benefits and possible burdens of research. So for example, research should not be designed in such a way that one group benefits from the research while another group bears the burdens (e.g. re-identification).

4.3 Human Rights

The idea of a balance, familiar in the context of prima facie duties, is often used to discuss a trade-off between the extent to which human rights can be respected and security be achieved. The existence of a trade-off implies the weighing of different duties: e.g. which duty—protecting the security of personal information (e.g. by favouring the diffusion of encryption technology) or preventing criminal attacks (e.g. by limiting the diffusion of encryption technology or requiring device makers to build back doors)—should take priority in a given context?

Note that the duty of protecting the security of personal information is here both a duty of cybersecurity *and* a duty in relation to human rights (the human right to privacy). This should not be a surprise. Indeed, cybersecurity technology that aims to protect privacy and confidentiality, such as encryption, is in general aligned with human rights; the threat to human rights is typically not cybersecurity, but *inadequate* cybersecurity or the lack thereof. However, there might be cases in which cybersecurity technology for the protection of privacy and confidentiality is *both* a means to privacy *and* a threat. Cybersecurity technologies such as encryption are naturally accompanied by *authentication* (which distinguishes those who have the right to obtain the non-encrypted information from the rest); authentication involves certification and the management of credentials. This requires the collection of information about individuals, which may expose users to privacy infringement.

Other kinds of cybersecurity technologies—those involved in monitoring web trafficking and fighting cybercrime—are in more direct conflict with human rights. Monitoring is associated with surveillance and surveillance involves threats of censorship (which can be a violation of the human right to free speech) and eavesdropping (which can be violation of the human right to due process). Moreover, monitoring is associated with profiling. Profiling "may be used by the police or security agencies to find criminals or terrorists; by airports to decide who to check more carefully" (Yaghmaei et al. 2017: 29–30). Hence, profiling is associated with potential violations of the human right against *discrimination*, because in profiling "people are approached, judged or treated in a certain way because these have characteristics that fit a certain profile and that are associated with certain other traits (i.e. traits other than by which they are identified as belonging to the profile)" (Yaghmaei et al. 2017: 29). The main ethical issue in profiling is not privacy, although personal information may be used to build profiles. It is the fact that "profiling may inflict all kinds of undeserved harm on people, from nuisance to false accusations to even, in extreme cases, imprisonment of innocent people" (Yaghmaei et al. 2017: 29–30). This happens because in profiling "a generalisation is made based on limited information about a person" (Yaghmaei et al. 2017: 30). The statistical discrimination involved in *any* form of profiling is only in conflict with the *human right* to non-discrimination when profiling involves specific (typically, legally protected) categories:

> The fundamental right of non-discrimination concerns the prohibition of discrimination in the context of occupation or employment, the provision of goods and services or other important domains of everyday life such as housing, social security or healthcare. Such prohibitions, which vary across jurisdictions, are limited to a set of grounds and do not touch price discrimination based on economic calculation or actuarial approaches to insurance. (Hildebrandt 2013, 368)

Protecting the human right to non-discrimination is one of the goals of (most) data protection regulation and is enshrined in Chapter III of the EU Charter, which

> includes […] gender equality (Article 23) [and] also prohibits '[a]ny discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation' (Article 21). The underlying objectives of equality and non-discrimination principles have been further pursued in the EU secondary law such as the Equal Treatment Directive in the context of employment (Directive 2006/54/EC) and the Directive implementing the principle of equal treatment between persons irrespective of racial or ethnic origin (Directive 2000/43). (Jasmontaite et al. 2017, 81; see also Chapter 5)

The cybersecurity technologies protecting individuals from cybercrime may conflict with human rights. Cybercrime may be defined to include four different broad categories of crime: *cybertrespass, cybervandalism, cyberpiracy* and *computer fraud* (Brey 2007). The first concerns gaining unauthorised access to data and information systems, the second disrupting processes and corrupting data, the third reproducing and distributing software or content which violates intellectual property and the fourth the misrepresentation of identity or information for the sake of deception for personal gain (Brey 2007).

The tension between the third type of cybersecurity and human rights should be clear from the outset, for the fight against cybercrime often involves "technologies to gain secret access to computing systems, to capture, observe and/or intercept data and content" (Hildebrandt 2013: 371). However, gaining access to and capturing data involves exactly the kind of cyber-threats to the privacy of information and confidentiality of communication that the first kind of cybersecurity technologies is designed to protect people from.

Hildebrandt (2013) observes that the expression 'to balance' can be used in this context to indicate two very different concepts. In the sense of a trade-off, the concept of a balance implies that it is necessary to curtail, imperfectly realise or narrowly specify a right's content in order to achieve a high enough level of security. But the core of the human right in question should not be compromised to achieve a marginal gain in cybersecurity and other ways of enhancing cybersecurity without undermining rights have to be explored, even if they are significantly less efficient, easy to realise or comprehensive. The idea of a 'balance' may also refer to something different from a trade-off. Balance, as in the expression of 'checks and balances', indicates quite a different concept. This is the idea that any increase in security measures needs to be accompanied by a proportional increase in alternative safeguards of the human rights, which cybersecurity risks undermining. Importantly, balancing cybersecurity and human right, in this sense, means creating checks and balances to protect human rights that may be threatened by heightened cybersecurity measures.

What are the rights that need to be balanced with cybersecurity? According to Hildebrandt, those rights are privacy, data protection, non-discrimination, due process and free speech. We have already mentioned examples involving some of these above. With the emergence of the Internet of Things (IoT), the right to physical integrity becomes also paramount, due to the capacity of attacks to undermine the physical integrity of individuals whose life-sustaining functions depend on the proper functioning of ICT mechanisms, for example in the health domain (Weber 2010; Mittelstadt 2017; Weber et al. 2018). For example, it is the physical integrity of a person that is a stake, if a 'black hat' hacker—a hacker moved by malicious intent—aims to access the software in a pacemaker in order to disrupt it and kill or harm the person who has it (Newman 2017).

Interestingly, Hildebrandt argues that if privacy is understood as "the freedom from unreasonable constraints on the construction of one's identity" (Agre and Rotenberg 1998: 7) then the other four rights are actually implied by the right to privacy in the era of smart environments (but arguably this extension does not include the fifth right we added to Hildebrand's list, of physical integrity). Hildebrandt explains the connection as follows: data collection and the profiling of the data subject define our identity for others and make us vulnerable to be defined by other people in ways that we would not choose to endorse; profiling enables discrimination practices against specific individuals or types or categories or groups of individuals—it bypasses conscious, reflective attitudes and plans that are key to being able to use due process. Free speech is also affected by the inability to control processes that steer our thinking (and expression) in ways that are unreflective, sometimes even unconscious. This includes "freedom from monitoring, filtering, and blocking of Internet traffic" (Hildebrandt 2013: 369). Of course, not all forms of monitoring, filtering and blocking of traffic have a negative impact on the human interests that the human right to free speech is meant to protect. The problem is, however, that essentially the same technologies that allow an Internet service provider, for example, to inspect traffic to identify and block malware, or other illegal content (including pirated media) may also be used to monitor and filter the contents of speech in a politically non-neutral way, which counts as a violation to the core interest that the human right to free speech is meant to protect. Thus, all cybersecurity technologies involving the monitoring and filtering are potential threats to this right. Interestingly, European law allows Internet service providers to inspect packages against malware and other security threats if this results from their own initiatives, but prohibit courts to oblige them to do so, to protect copyright (Hildebrandt 2013: 369). This example demonstrates that courts themselves (in this case the European Court of Justice) engage in balancing (in both senses of the expression) when interpreting the scope of fundamental human rights. In this case, the courts may have reasoned that citizens' interest in avoiding cybertrespass and cybervandalism has sufficient weight to justify the use of monitoring and filtering technology in spite of the risks involved, whereas citizens' (and companies') interests in avoiding *cyberpiracy* do not. Alternatively, they may have reasoned that the monitoring and filtering of malware, given its nature, is less likely to imply censorship consequences than the monitoring and filtering of content related to intellectual property.

The following example, inspired by a real-world case study (Dittrich et al. 2011), illustrates the principlist and rights-based approach applied to the deployment of cybersecurity technology for *monitoring* computer systems in a response to a cybersecurity attack.

> *An information warfare monitor:* You are investigating a malicious botnet, the victims of which included the foreign embassies of dozens of countries, the Tibetan government-in-exile and multinational consulting firms. You begin your research by reviewing data collected by passive monitoring of suspected victim networks, which confirms the intrusions and identifies the malware. You collect more data from compromised computers with the owners' consent, monitor the command and control (C&C) infrastructure enough to understand the attackers' activities and to enable notification of infected parties at the appropriate time, work with government authorities in multiple jurisdictions to take down the attacker's C&C infrastructure, and store and handle data securely. (Adapted from Dittrich et al. 2011)

An information warfare monitor poses threats to right to privacy and of free speech of the suspected and actual victims (which may be particularly relevant for an exiled government). These threats are posed by the passive monitoring of suspected victim monitors (without consent) and subsequent data collection from the affected computers (with consent). In terms of the principlist approach, informed consent and notification fulfil the duty of *respect of persons.* In terms of the rights-based approach, they can be regarded as a way to balance (in the sense of checks and balances) the risk to the privacy of the victims caused from monitoring. Informed consent, it may be claimed, reduces the vulnerability to which a privacy breach and surveillance expose the subject of the right. Moreover, from a principlist point of view, security measures taken in the storing and handling of data from the computers of the victim (e.g. encryption, anonymisation, etc.) fulfil the duty of *beneficence* (which includes nonmaleficence as risk reduction). From the perspective of a human rights approach, they can be seen as a way to balance (in the sense of 'checks and balances') the heightened risk to privacy and informational self-determination of all other persons that the data in the infected computers may identify.

4.4 From Principlism and Human Rights to the Ethics of Risk

Hildebrandt advocates a legal approach (the 'triple test'; explained below) which involves both balancing as a trade-off and balancing as in 'checks and balances'. Some kind of trade-off is unavoidable when considering a rich and diversified set of human rights, because the duty implied by respect for one right may contradict the duty implied by respect for a different right. However, the idea of accepting a trade-off involving a human right may appear to contradict the very idea of a right, if a right is a side-constraint; that is, a rigid constraint defining the permissible scope of all other moral actions (Nozick 1974), or a 'trump card' (Dworkin 1977); that, is a consideration defeating all other utility considerations. According to those views, rights are different from other interests because they are the kind of things that societies cannot violate *even when* the violation clearly leads to a maximisation of aggregate interests (Rawls 1999: 3).

However, unless rights are very few and limited in the kind of duties they entail,[1] they are very likely to logically contradict each other in practical contexts. This is especially true of human rights as they are quite numerous and tend to have significant implications in terms of the resources and duties required by society to satisfy them.

The way Hildebrandt and John Rawls[2] address the problem of trade-offs involving rights is by acknowledging the necessity of limiting rights "without losing their substance" (Hildebrandt 2013: 375). What that means, in practice, is that one has to draw a distinction between the core elements of a right, which ought never be sacrificed (what Rawls calls "the central range of applications" [Rawls 1982: 11]) and those elements that are peripheral and should be satisfied, when possible, and sacrificed when they conflict with the core elements of another right. The hope is to be able to achieve, in a rationally defensible way, what Rawls calls a fully adequate scheme of rights and liberties. In doing so, pragmatic elements (what historical experience teaches us about the co-possibility of satisfying different rights within a coherent institutional arrangement) also play a role. However, deciding what applications of a human right are central to its meaning requires some kind of theory about the social function of the right in question.[3]

Hildebrand's triple test, which derives from an interpretation of the second paragraph of Art. 8 of the European Convention of Human Rights (binding for the 52 states of the Council of Europe), requires that a right's infringement "must be in accordance with the law, necessary in a democratic society and have a legitimate aim" (Hildebrandt 2013: 375). The necessity requirement "is understood as a requirement of proportionality between infringing measure and legitimate aim" (Hildebrandt 2013: 376). Proportionality is, philosophically, a difficult notion, but in the context of Hildbrandt's reasoning it may be interpreted, again, as a weighing

[1] This is arguably the case of a framework that only includes Nozickian libertarian ownership rights. These are strict negative rights prohibiting aggression and other forms of non-consensual interference aimed at dispossessing individuals of the fruits of their labor and of voluntary exchanges with other individuals.

[2] See for instance (Rawls 1982, 1996 Lecture VII: §8–11).

[3] This, of course, leaves open the question of how to address a conflict of rights when the clash involves the peripheral area of both rights, or the core area of both rights. There is no time here to dwell on the analysis of this problem. Perhaps it is acceptable to claim that it is compatible with respect for human rights to decide democratically which of two rights to sacrifice when both are involved only peripherally; the real tragic case is the one of a conflict between the cores of two rights, and perhaps a viable approach here is compensation (not necessarily only monetary). One potential solution that appears problematic here is a *maximising* one (i.e. to choose the combination of rights that maximises a given parameter). Any sufficiently *pluralist* conception of the fundamental interests and values behind such rights entails that there is no single metric to be maximised. That element of pluralism is perhaps what distinguishes, most fundamentally, a rights approach from a utilitarian one.

of the likelihood that, should a given privacy infringement not be allowed, an interest in the central range of application of some other right will be at risk, combined with a weighing of the likelihood that the cybersecurity measure adopted will not undermine the overall protection of the core human interests protected by the right in the core range of application of the right. An illustration of this could be the interpretation offered above of a high court decision to allow ISP to monitor and filter Internet traffic against malware and other cyberthreats, but to prohibit lower courts to oblige ISP to monitor and filter Internet traffic against violations of copyright laws.

Note, however, that even in a human rights approach, it is impossible to escape some probabilistic assessment of the risks of violating a right. Thus a right-based theory, no less than principlism, involves the assessment of risk and probabilities at some level of analysis. The evaluation of probabilities is explicit in the idea of risk-benefit analysis that is also explicitly invoked by the Menlo report in the application of the benevolence principle in practice.

It seems legitimate to conclude that the ethical assessment of cybersecurity always depends on risk assessment of a probabilistic form. Risk-assessment is normally understood as an aspect of the consequentialist approaches that justify the line of action that produces the biggest net benefit. When the outcomes are uncertain, actions and policies can only be assessed in terms of their *expected* net benefit. However, beyond utilitarianism (that is *only* concerned with outcomes) risk-benefit assessments are an integral aspect of any ethical framework that assesses the morality actions *also* in relation to their outcomes; for example, it is invoked by most interpretations of the duty of *beneficence* in principlist approaches in research ethics. Note that the Menlo Report states very clearly that the risk-benefit assessment under the heading of beneficence is not meant to be restricted in scope to research subjects. Instead, "[…] researchers should systematically assess risks and benefits across all stakeholders. In so doing, researchers should be mindful that risks to individual subjects are weighed against the benefits to society, not to the benefit of individual researchers or research subjects themselves" (Dittrich and Kenneally 2012 L 9).

Balancing a cybersecurity measure that poses a threat to privacy with heightened privacy guarantees requires an assessment of proportionality between the risk that a cybersecurity measure is meant to protect society against and the threat (free speech, due process, non-discrimination or data protection) that it constitutes against a human right. This presupposes a consideration of the *probability* of the violation of a right in the core area of application of such right.

4.5 Cybersecurity and the Ethics of Risk

In what follows, we shall consider a single cybersecurity case as a way of illustrating different approaches to the ethics of risk.

Responding to ransomware: You are the leader of a CERT team and you have identified ransomware (a software virus that encrypts the data in the computers infected and directs the victims to a payment service where, after paying 1000€, the victims can obtain the decryption key). You know that a partner software company has already begun to code an algorithm to decrypt the data; you estimate that the company has a 65% chance of success within one month (and a 0% chance of succeeding later). At the moment, 1000 computers are affected, all belonging to the network of an important hospital. Unfortunately, it is impossible to reconstruct what data was saved in each computer and the date of the latest backup. The probability that an alteration or deletion of data in a single computer will cause the death of a patient is 1/1000 for each device.

You can choose one of two response strategies:

- *Policy A*: you quarantine all the affected computers and shoot down the payment servers. These measures, with foreseen 100% efficacy, will prevent the spread of the infection and reduce the incentives for attackers to involve other computers in similar attacks in the near future. However, the malware is designed to detect your response and retaliate to it. It will irreversibly introduce random changes in the data in ways that are extremely hard to detect, or simply delete it. It is not possible to identify the data causally linked to the lives of patients in a reasonable amount of time.
- *Policy B:* you do not isolate the affected system and do not bring down the payment server; after one month, either you have obtained the decrypting tool with no losses; or you have not, in which case the infection will have spread to other 1,000,000 computers, with an expected aggregate economic loss for your society of €400,000,000, mostly consisting of donations of €500 to the hackers.

4.5.1 Expected Utility Maximisation

According to the moral theory of utilitarianism, the moral appraisal of any action is solely a function of the utility consequences of that action, i.e. of the sum total of well-being (or happiness) produced. (The net amount of aggregate well-being due to an action may also be negative if well-being losses are greater than gains.) Three features of utilitarianism are worth noticing: it is consequentialist, welfarist (the ethical appraisal of consequences only considers the well-being of sentient beings involved) and aggregative (individual losses of well-being to one individual may be compensated by greater gains to others). Utilitarianism is also a strictly *maximising* theory: the *right* action is the one that maximises well-being in the aggregate. Even an action that produces a net gain of well-being relative to a previous state of the world is *wrong*, if a different action leading to a *greater* increase of utility is feasible.

Since the consequences of virtually every action are to some degree uncertain, any action-guiding version of utilitarianism must *not* assess actions based on the outcome that actually materialises. The action-guiding version of utilitarianism prescribes the maximisation of *aggregate expected utility*, by which one means the probability-weighted average of utility in all possible states of the world that an action could cause.

The ethical dilemma for our case is to compare an expected disutility of €260,000,000€ (65% chance of a possible €400,000,000 damage if the decryption tool is not developed) with the probability of causing one or more deaths. The probability that no single computer is essential to the life of a patient is $(999/1000)^{1000}$, which entails a $1-(999/1000)^{1000}$ —roughly a 63%—chance that one person will die because of the first policy. Thus, policy A imposes a significant risk to a single individual. As a guide to cases like this, the guidance by utilitarian risk-benefit assessment strikes some as counterintuitive. It requires the decision-maker to compare a high expected likelihood of death, for a single person, with aggregate disutility for a large group, formed by individuals each of whom suffers a very small loss compared to death. It may seem plausible that, no matter how large in the aggregate, the sum of many small individuals losses cannot justify imposing a high risk of death for a single person. Utilitarianism, however, implies that the opposite must be the case: no matter how valuable a personal life (assuming a finite value), the aggregate of small damages inflicted to a group will count for more, if the group is large enough.

4.5.2 The Maximin Rule

A close relative of utilitarianism (or better, expected utility consequentialism) is what one may call *maximin* consequentialism. According to the maximin rule, in Hansson's formulation:

> the utility of a mixture of potential outcomes is equal to the lowest utility associated with any of these outcomes. (Hansson 2003: 296)

The 'mixture' of the potential outcomes of an action is the set of all outcomes whose probability of occurring is more than zero. The maximin rule orders the desirability of actions according to the desirability of their worst possible outcomes. The algorithm for the cybersecurity professional in the case at hand is:

1. assess the total utility of the worst outcome (O_A) associated with A, considered as if it were certain;
2. assess the total utility of worst outcome (O_B) associated with B, considered as if it were certain;
3. if $U(O_A) > U(O_B)$, choose A; if $U(O_A) < U(O_B)$, choose B, if $U(O_A) = U(O_B)$ draw a lottery with a 50% chance of A and B.

The worst outcome for action A is the certain death of one person; the worst outcome for action B is a certain damage of €400,000,000. The maximin approach requires that we compare the two outcomes and choose the lesser of the two. Note that this approach suffers from an objection analogous to utilitarianism, namely that, unless an individual life has an infinite moral value, it may justify the sacrifice of a human life to avoid a large sum of individually limited economic damages.

Maximin is also subject to another objection. Suppose that O_A is an outcome with a very small probability, e.g. a 1/1,000,000,000 chance of causing non-permanent health damage to all patients, amounting to a loss of 1,000,000,000€ in medical expenses and compensation. Utilitarianism entails that O_A should be chosen, because the expected *dis*utility of O_B, being certain, is much higher, than the disutility of O_A, which is discounted by its low probability. Maximim requires choosing O_B, because it does not discount the disvalue of O_A because of its low probability. Many would find utilitarianism more plausible than Maximin, given that in everyday life we consider it rational to engage in activities, such as crossing the street, which have a very small probability of leading to very bad outcomes (death after being hit by a car), even for the sake of a very small utility gains (e.g. purchasing ice cream).

Arguably, a significant proportion of those who believe that an individual life should be considered more important than a loss of €400,000,000 (distributed in small €500 losses for each individual), may nonetheless agree that strategy A is justifiable, given that the risk of causing death is so small. For example, we allow people to drive cars, in spite of the fact that allowing car driving increases the risk of death for innocent pedestrians, which may in fact be higher. Maximin consequentialism, however, obliges you to base your decision on what the worst possible outcome is for each scenario, in a method that is totally insensitive to its probability.

Therefore, the problem with this approach is that it would prohibit all cybersecurity measures that have some probability, no matter how low, of causing very significant harm as a side-effect (no matter how unlikely the causal chain that would lead to such outcome). Another problem is the difficulty of enumerating the low-probability events that may be associated with a given policy. As Hansson points out, we have to stop considering low-probability events that may follow from our actions at a certain point, and there may be no non-arbitrary cut-off point. This would introduce a degree of moral arbitrariness in the moral evaluation of such risks that counts against adopting the Maximin rule (Hansson 2003: 296).

4.5.3 Deontological and Rights-Based Theories

Deontological approaches are typically built around a list of morally prohibited acts, that is, acts that are prohibited no matter what, i.e. irrespective of the consequences. Suppose, for example, that it is not permissible to expose the private health condition of an individual to the public against his consent. A strict deontological moral system entails that it is always wrong to do so, even if, let us suppose, knowing this information would allow millions of shareholders of a company led by the sick man to reduce their exposure to financial risk. Let us refer to the acts that are prohibited—even when they would maximise utility—as ‘violations of deontological constraints’. Deontological approaches to *risk* claim that moral agents act wrongly if acting involves a non-null risk of violating a deontological constraint.

(Absolutist) rights-based theories are similar to deontological theories, but they are framed in a manner that shifts our attention to the person obligations are owed to, rather than to the agent who is obligated. If persons have rights, certain things cannot be done to them no matter how good the general consequences, while other things are owed to them, no matter what the costs are. By extension, rights-based theories *of risk* claim that moral agents ought not to perform actions that have a more than a null risk of violating the rights of other people. For example, every innocent person may be believed to have a negative right to life, entailing a duty of other people not to act in ways that would cause that person to die.

Let us move to a more rigorous formulation of such views. Following Hansson, let us define:

> *Probabilistic absolutism:*
>
> [for deontological theories]: If it is morally prohibited to perform a certain action, then this prohibition extends to all mixtures in which this action has non- zero probability.
>
> [for rights-based theories]: If someone has a moral right that a certain action not be performed, then this right extends to all mixtures in which this action has non-zero probability. (Hansson 2003: 298)

In Hansson’s terminology, *mixtures* are value carriers (actions, outcomes). For example, in the CERT case, the CERT manager is addressing the following two mixtures:

> A: shutting down the payment server, limiting the range of computers affected by ransomware and indirectly causing a person’s death;
>
> B: not shutting down the payment server, allowing ransomware attacks to continue and allowing economic damage to occur.

According to probabilistic absolutism, if ‘indirectly causing an (innocent) person’s death’ is impermissible, then every act that has a small probability of causing a person death is impermissible too. Thus, probabilistic absolutism prohibits A even when the probability of harming a patient is very low (e.g. equal to or less than 0.001% in the variation of the ransomware scenario discussed in Sect. 4.5.2).

The problem with this theory is that it is, in general, too demanding for the moral subject who, by virtue of some apparently innocent act, associated with some terrible outcome by virtue of a very unlikely chain of events, risks violating his duties. It also prevents the execution of many acts of beneficence (often attempts to do the good have a very small probability of doing some evil). Often, agents will face a dilemma in which they will violate duties whichever option they choose.

Some of the implausible consequences of probabilistic absolutism are avoided by risk-deontological and risk-rights-based theories acknowledging a *probability limit*.

> *Probability limit for risk-deontological theories:* Each prohibition of an action is associated with a probability limit. The prohibition extends to a mixture that contains the action if and

> only if the action has, in that mixture, a probability that is above the probability limit. (Hansson 2003: 298)[4]

In the threshold approach, risk-deontological (or risk-rights-based) constraints generate moral duties *only if* the risk of violating a deontological constraint (or another person's rights) is higher than a given *threshold value*. Therefore, it is legitimate to ignore risk-deontological (or risk-rights-based) prohibitions when we do actions that only have a very low chance of causing violations of these constraints.

This approach may seem to deliver a reasonable method to assess the scenario described above. With a probability threshold set to 5%, policy A would be impermissible in the first case discussed (where the risk of death of a patient was >60%) but not in the second one (where the probability of health damage was extremely low).

The main problem with the theory is that it appears difficult to justify such thresholds (e.g. how low should the probability of killing an innocent be to allow it to occur?). Not only it is difficult to justify a single threshold, but it seems even harder to justify different thresholds for different types of harm (e.g. how high should the threshold for allowing economic damage be set, in comparison to the threshold for causing death?) a priori.

Justice theories may explain some intuitions concerning the imposition of risk. Some of these theories imply that it is *ceteris paribus* ethically wrong to impose risk on individuals who are already vulnerable to risk instead of targeting less vulnerable people (Wolff and De-Shalit 2007; Ferretti 2009, 2016). For example, if a threat exists that could lead to the irremediable loss of equally sensitive data, it is *ceteris paribus* wrong to let the risk be imposed on poor instead of wealthier people. This is because, for the former, losing €500 due to the ransom may involve a significant sacrifice of economic security, which may increase their exposure to other kinds of risk (e.g. tackling disease or unemployment). Ferretti's (2016) theory focuses on total risk, suggesting that the threshold level should be different when duties affect persons in circumstances that already add to/reduce their total risk level. Similar implications can be drawn from capability-based theories of disadvantage and risk (Wolff and De-Shalit 2007; Murphy and Gardoni 2012).[5]

These *non-deontological* theories explain intuitions, which may be quite widespread, that what counts as an "acceptable level of risk" depends on both the kind of risk in question and the situation of the person affected by this risk. In contrast to the latter, risk-deontological (or risk-rights based) theories of risk assume an equal risk-threshold for all. The risk-deontological approach as such does not provide

[4] The probability limit for rights-based theories can be defined along similar lines.

[5] These theories measure the impact of risk in terms of their impact on capabilities, defined as genuine opportunities to achieve *valuable* functionings (Sen 2009; Nussbaum 2006). The approach by Wolff and De Shalit (2007) focuses in particular on the fact that certain categories of risks tend to affect more than one capability. It attributes more harmful effects to 'cross-category risks' and 'inverse cross-category risks'.

any principled guidance to assign different levels of risks in different cases.[6] In order to justify a *different* risk threshold, one needs to appeal to some independent conception of *fairness* in risk distribution. One last approach we will consider is the one provided by *contractualism.*

4.5.4 Contractualism and Risk

Aggregative views in general (not just aggregative views on risk) are exposed to peculiar counterexamples; the cybersecurity response to ransomware in Sect. 4.5.1 may be taken as one such example. The cybersecurity response A, which imposes a 65% risk that a person will die, seems morally objectionable because the sum of individual small losses, no matter how large, cannot justify imposing a significant risk of death to a single person.

The philosopher Thomas Scanlon has proposed *contractualism* as an alternative to utilitarianism. Contractualism compares the strength of the individual claims without aggregating them (Scanlon 1998: 235). Scanlon's way of comparing individual complaints has later been labelled the MiniMax Complaint principle, which states that "when we would not be violating any moral constraints, we are morally required to act in the way that minimises the strongest individual complaint" (Horton 2017, 55). In our example, the relevant complaints concern (a) the life of one individual person whose medical treatment depends on the integrity of the encrypted data and (b) the individual loss of €500 of one individual, not yet affected, who will end up paying a ransom for his encrypted data if further attacks are not prevented by shooting down the payment server. Since the complaint against death is greater than the complaint against a ransom, one ought not to quarantine the computers and to shoot down the payment servers.

There is a lively philosophical debate on how to interpret the MiniMax Complaint principle in cases involving risk. Consider the choice between two vaccines, assuming that choosing either one is necessary to avoid the spread in the population of an epidemic that will unavoidably kill everyone on Earth. Vaccine A has a one in a million chance of killing the user as a side effect; vaccine B leads to the certain paralysis of one limb for all users. The *ex post* version of the MiniMax Complaint (Scanlon 1998; Reibetanz 1998; Otsuka 2015), requires choosing B, since it adopts the perspective of a person who is certainly going to die as a result of A. Here it is assumed that in a population of several billion people it is almost certain that someone will die, but the identity of this person cannot be known in advance. In the *ex post* approach, the claim of the *statistical individual who will unavoidably die* is stronger (for *ex post* contractualism) than the claims of every person who, if the

[6] However, they can be used to represent all the appropriate beliefs. For example, a deontological theory can be a simple list of many different duties and rights, associated with specific probabilities specified at the level of concrete situations.

other vaccine is chosen, will only end up paralysed. Many find this counterintuitive.

An alternative theory is *ex ante* contractualism (Lenman 2008; James 2012; Frick 2015). A simple *ex ante* version compares complaints in terms of *expected* harm, that is to say, the outcome is weighted by the probability of its occurrence. Thus, the risk of 1 in a billion chance of losing life may be considered weaker than having a paralysed limb with full certainty. Thus the *ex ante* view justifies using vaccine A. This is considered more plausible by those who think, for example, that compulsory vaccination for non-lethal diseases is not necessarily morally wrong, even it is known in advance that some people will die because of lethal complications.

Ex ante contractualism may appear to have plausible implications in the case of a CERT's response to ransomware. When the risk of a patient's death (for each patient) is very low, it entails that it is permissible to quarantine the system and put the server used for the payments of the ransom offline. When the risk is significant, it prohibits sacrificing the patients.

But even *ex ante* contractualism has detractors. The objections against it can be explained more easily by focusing on a different case:

> *A choice of anti-malware:* You are dealing with malware that turns the affected computers into nodes in a botnet performing a distributed denial-of-service attack against servers in an important hospital, which risks placing the lives of its patients at risk. You have three anti-malware tools in your arsenal, all of which are effective against the malware. However, the malware is designed to retaliate by wiping out the entire hard disk, as soon as it is disconnected from the malicious server. A preliminary study of the malware shows that it could be fought with three different software approaches. Each of them fails in specific ways to limit the damage. Due to time and resource constraints, you can develop only one of these before the malware spreads, causing morally intolerable human damage. Which one do you develop?

- Anti-malware 1: it protects all computers but deletes all Excel and Word files during installation.
- Anti-malware 2: it only works on non-Apple operating systems, which entails that Apple systems will have to be quarantined (and will lose all data). Ten percent of the computers in the botnet are Apple ones.
- Anti-malware 3: it works perfectly on all computers, except on those with some specific UUIDs, Universal Unique Identifiers, assigned by the malware itself. It is impossible to determine the UUID generated by the malware without triggering a malware response that would erase all data. Hence, for every practical purpose, the UUID of each infected computer can be considered unknown and unknowable. It is known, however, that the malware will wipe out all the data if the last numerical digit of the UUID it assigned to device is 0. Since every Arabic numeral has the same chance of being the last numerical digit in these UUIDs, every computer has an *ex ante* 10% probability of being wiped out completely and a 90% probability of being rescued completely.

Let us begin by comparing Anti-malware 1 vs. 2. *Ex ante* contractualism here entails weighing the *ex ante* complaint of Mac users (having the hard-disc completely wiped out) vs. the *ex ante* complaint of other users (having only text and spreadsheet files deleted), considered individually. Since Mac users have the strongest *ex ante* complaint (they are 100% sure of having all their files deleted), contractualism requires that you choose anti-malware 1. In the imaged scenario, Apple software runs on 10% of the affected computers; note, however, that contractualism would have implied the same response if there had been a single Mac user in the botnet.

Let us now consider anti-malware 1 vs. anti-malware 3. Suppose that you have established empirically that each computer owner strongly prefers a lottery with a 90% chance of rescuing the data and a 10% probability of losing all data in the computer, compared to the certain loss of all their text and spreadsheet files. *Ex ante* utilitarianism entails, in this case, that you ought to choose anti-malware 3.

Is the choice of malware 3 morally unobjectionable? Similar cases in moral philosophy have been criticised for two reasons. First, it treats identified individuals, such as owners of Mac computers, differently from *statistical* individuals, e.g. owners of computers with a UUID whose last numeral digit is 0, whose identity can be determined only *after* they suffered from the harm. However, the difference between statistical individuals and identified individuals seems entirely morally arbitrary—in no way are statistical individuals less worthy of respect. Second, it uses statistical individuals as means: their interests are sacrificed to promote the aggregate good (Rüger 2018).[7]

In summary, it seems reasonable to expect that some situations faced in cybersecurity analysis and operation deal with outcomes that are not certain, but to which probabilities (often, mere subjective probabilities) can be assigned. Unfortunately, utilitarianism suffers from known objections (sacrificing the individual for the greater good) and there are hard cases in which the most intuitively plausible version of contractualism is no different from utilitarianism in this respect.

4.6 Contextual Integrity

Contextual integrity is a framework for understanding privacy, both descriptively (i.e. why do people find some technologies upsetting?) and normatively (should society favour the introduction of certain technologies?) (Nissenbaum 2004, 2009). The main insight of this theory is that privacy violations consist of violations of social norms concerning the transmission of information between persons. The relevant social norms are specific for the social contexts/practices and the social roles that individuals have within those practices. For example, the transmission of information between patient and physician in a hospital, spouses within a family, priest

[7] Philosophers have tried to avoid these types of problems by providing more sophisticated formulations of both *ex ante* and *ex post* versions of contractualism. All appear to be vulnerable to counterexamples and, for this reasons, it has been argued that the Minimax Complaint view should be abandoned altogether when dealing with risk (Horton 2017).

and confessor within the church, employer and employee within a company, policemen and citizen within the state, need not be (and usually are not) governed by the same informational norms. Individuals have privacy when established expectations concerning the way information should be transmitted are respected—this is compatible with people expecting different people in different contexts to handle their information in very different ways. However, not all changes of social norms and expectations concerning information should be considered violations of privacy since, as we shall see, some changes in informational norms may be justified, all things considered.

Contextual integrity is a mildly conservative theory. The violation of a contextual integrity norm provides a prima facie case for considering a new practice (e.g. the introduction of a new cybersecurity technology) as a sensitive privacy issue. However, the overall evaluation of the innovation may turn out to be justified in the end. Thus, the theory has a conservative bias, but it does not support conservative prescriptions in every case. Violating established expectations can be significantly harmful,[8] but it may not be wrong overall. The conservative bias of the theory can be overcome by pointing out, following the work of Michael Walzer (1983), that a transformation even in an established social norm can provide a more sensible method to achieve the goals that actors in a practice are set to achieve, without altering the most general relevant principles applying to the domain, and without violating the fundamental rights and interests of all those affected (Nissenbaum 2009: Chap. 8).

In recent work (2009), Nissenbaum explains how to use the theory as a basis for the empirical analysis of technologies that are perceived as raising a privacy problem; the feeling of a technology being problematic is explained as a consequence of its violation of expectations concerning information, given the existing context-relevant social norms. The moral assessment is driven by the assessment of the goal of the practice and the framework of more general principles and values applying across domains. Nissenbaum's privacy as contextual integrity is directly relevant to assessing cybersecurity technologies whose goal is to ensure the confidentiality of information. It is also pertinent to assessing technologies for detecting online threats and counter cybercrime, since such technologies are likely to affect the way information is accessed and used as a side effect.

[8] Nissenbaum (2004, Chap. 8) justifies the conservative inclination of the theory by considering arguments for conservativism provided by the radical utilitarian philosopher Jeremy Bentham (1747–1832) and the conservative philosopher Edmund Burke (1729–1797). Bentham argues that laws contradicting established ones tend to undermine the sense of security that derives from established expectations about the law. Thus, radical legal innovations could bring about—at least during the transition to a new legal regime—a utility loss, making it more difficult for agents to plan rationally in the pursuit of their own goals. Burke, on the other hand, considers established customs as the product of accumulated wisdom, which normally exceeds the ability of the individual minds to build models of social interactions and solutions for social problems that work in practice. Arguably, both arguments apply also to abrupt changes in conventional norms concerning information.

Moreover, some aspects of Nissenbaum's framework can be expanded and applied beyond its original scope, i.e. privacy. In particular, let us assume that the moral importance of contextual integrity derives from the value (in terms of security, peace of mind and the ability to rationally plan one's life) of fulfilling expectations. If so, there is no reason to consider only expectations connected with *informational* norm, as Nissenbaum's approach does. Her theory can be generalised into a more overarching theory that requires cybersecurity agents to consider established social norms and expectations concerning the actions (e.g. 'investigating a crime', 'assessing the trustworthiness of an employee', 'responding to an emergency in a patient') and not only those associated with the way information is accessed, transacted and used.

We thus conclude this essay by sketching a methodology for the ethical assessment of cybersecurity technology, which is essentially a version of Nissenbaum's contextual integrity privacy framework (2009: Chap. 9), extended to include social norms and expectations affecting all human interactions that are constitutive of an established social practice. The approach applies to all cases in which the adoption of a cybersecurity policy, or technology, affects the way information is exchanged. It also applies to all cases in which it affects the relations between people with established roles (roles linked to stable expectations) within the institution (e.g. hospital, company) or practice (e.g. diagnosis, marketing) that is affected by them. Following Nissenbaum, the framework consists of the following steps:

1. Establish the prevailing context of the cybersecurity measures in question (e.g. finance, law-enforcement, administration, business, medicine or some combination of more than one context);
2. Ascertain the information attributes (e.g. citizen's name, age, amount and entity of commercial transactions, purchase type) affected by the cybersecurity measures proposed; ascertain what aspects of human interactions (which are not defined by informational exchanges) are affected
3. Determine what changes in the principles/social norms governing the transmission of information are foreseeably due to the cybersecurity measures; determine other foreseeable changes in human interactions and modalities of operation in practice;
4. Red flags: if the new cybersecurity measures generate changes in the actors (e.g. client, financial institution employee, police investigator, nurse, physician), attributes (e.g. the kind of information/interaction affected) or relevant social norms, flag the measure as a prima facie violation of the contextual integrity of the domain in question. This counts as a prima facie violation and counts against the measure unless it can be justified in steps 5 and 6 below.
5. For a technology that has raised a red flag, determine what are the socially valuable goals and the core EU values and rights affected by the change in informational norms and expectations concerning the social interactions that have been detected;

6. For a technology that has raised a red flag, determine if the changes caused in this way improve the prospects of the actors to achieve the valuable goals of the practice; determine also whether they conflict with core EU values and rights.

4.7 Conclusions

This chapter presented several ethical frameworks for evaluating cybersecurity threats, countermeasures and policies. The chapter began with an examination of two influential approaches, the principlist approach (especially influential for the ethics of cybersecurity *research*) and the human rights approach (especially important for the law, in particular EU law). Both approaches are non-utilitarian, in that they do not define as morally right, or morally required, those cybersecurity acts (or policies) that maximise the good, defined as a single value (e.g. utility, or happiness). We then demonstrated that both these non-utilitarian approaches raise questions about the ethics of risks and present different ethical approaches to evaluating risk. Finally, we presented Helen Nissenbaum's contextual integrity theory both as a framework to understand why some technological changes are perceived as problematic and as a normative approach to assess whether they count as privacy violations all things considered. We proposed a revised version of Nissenbaum's contextual integrity framework for identifying and ethically assessing changes brought about by cybersecurity measures and policies, not only in relation to privacy but more generally to the key expectations concerning human interactions within the practice.

Acknowledgements The chapter was created with funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 700540 and the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 16.0052-1.

References

Agre PE (1998) Introduction. In: Agre PE , Rotenberg M (eds) Technology and privacy: the new landscape. The MIT Press, Cambridge, MA/London, pp 1–28, at 7

Boyce MW, Duma KM, Hettinger LJ (2011) Human performance in cybersecurity: a research agenda. Proc Hum Factors Ergon Soc Annu Meet 55(1):1115–1119. https://doi.org/10.1177/1071181311551233

Brey P (2007) Ethical aspects of information security and privacy. In: Security, privacy, and trust in modern data management, data-centric systems and applications. Springer, Berlin/Heidelberg, pp 21–36. http://link.springer.com/content/pdf/10.1007/978-3-540-69861-6.pdf#page=36

Dittrich D, Kenneally E (2012) The Menlo report: ethical principles guiding information and communication technology research. US Department of Homeland Security

Dittrich D, Bailey M, Dietrich S (2011) Building an active computer security ethics community. IEEE Secur Priv 9(4):32–40

Dworkin R (1977) Taking rights seriously. Harvard University Press, Cambridge, MA

Ferretti MP (2009) Risk and distributive justice: the case of regulating new technologies. Sci Eng Ethics 16(3):501–515. https://doi.org/10.1007/s11948-009-9172-z

Ferretti MP (2016) Risk imposition and freedom. Pol Philos Econ 15(3):261–279. https://doi.org/10.1177/1470594X15605437

Frick J (2015) Contractualism and social risk. Philos Pub Affairs 43(3):175–223. https://doi.org/10.1111/papa.12058

Hadlington L (2017) Human factors in cybersecurity; examining the link between internet addiction, impulsivity, attitudes towards cybersecurity, and risky cybersecurity behaviours. Heliyon 3(7):e00346. https://doi.org/10.1016/j.heliyon.2017.e00346

Hansson SO (2003) Ethical criteria of risk acceptance. Erkenntnis 59(3):291–309

Hildebrandt M (2013) Balance or trade-off? Online security technologies and fundamental rights. Philos Tech 26(4):357–379. https://doi.org/10.1007/s13347-013-0104-0

Horton J (2017) Aggregation, complaints, and risk. Philos Pub Affairs 45(1):54–81. https://doi.org/10.1111/papa.12084

James A (2012) Contractualism's (not so) slippery slope. Leg Theory 18(3):263–292. https://doi.org/10.1017/S135232521200002X

Jasmontaite L, Fuster GG, Gutwirth S et al (2017) Canvas White Paper 2 – cybersecurity and law. SSRN scholarly paper ID 3091939. Rochester: Social Science Research Network. https://papers.ssrn.com/abstract=3091939. Last access 7 July 2019

Johnson ML, Bellovin SM, Keromytis AD (2011) Computer security research with human subjects: risks, benefits and informed consent. In: International conference on financial cryptography and data security. Springer, Berlin, pp 131–137

Kenneally E, Bailey M (2013) Cyber-security research ethics dialogue & strategy workshop

Kenneally E, Michael Bailey M, Maughan D (2010) A framework for understanding and applying ethical principles in network and security research. In: International conference on financial cryptography and data security. Springer, Berlin, pp 240–246

Lenman J (2008) Contractualism and risk imposition. Pol Philos Econ 7(1):99–122. https://doi.org/10.1177/1470594X07085153

Mittelstadt B (2017) Designing the health-related internet of things: ethical principles and guidelines. Information 8(3). http://www.mdpi.com/2078-2489/8/3/77htm. Last access 7 July 2019

Murphy C, Gardoni P (2012) The capability approach in risk analysis. In: Handbook of risk theory. Springer, Dordrecht, pp 979–997

Newman LH (2017) Medical devices are the next security nightmare. Wired. https://www.wired.com/2017/03/medical-devices-next-security-nightmare/. Last access 7 July 2019

Nissenbaum H (2004) Privacy as contextual integrity. Wash Law Rev 79(1):119

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

Nozick R (1974) Anarchy, state, and utopia. Basic Books, New York

Nussbaum MC (2006) Frontiers of justice: disability, nationality, species membership. The Belknap Press, Cambridge, MA

Otsuka M (2015) Risking life and limb: how to discount harms by their improbability. In: Cohen GI, Daniels N, Eyal N (eds) Identified versus statistical lives: an interdisciplinary perspective. Oxford University Press, Oxford

Rawls J (1982) The basic liberties and their priority. The Tanner Lectures on Human Values 3:3–87

Rawls J (1996) Political liberalism, expanded edn. Columbia University Press, New York

Rawls J (1999) A theory of justice, 2nd edn. Harvard University Press, Cambridge, MA

Reibetanz S (1998) Contractualism and aggregation. Ethics 108(2):296–311. https://doi.org/10.1086/233806

Ross WD (2002) The right and the good. Stratton-Lake P (ed) Oxford University Press, Oxford

Rüger K (2018) On ex ante contractualism. J Ethics Soc Philos 13(3). https://doi.org/10.26556/jesp.v13i3.323

Scanlon T (1998) What we owe to each other. Belknap Press of Harvard University Press, Cambridge, MA

Sen AK (2009) The idea of justice. Harvard University Press, Cambridge, MA

Spring JM, Illari P (2018) Building general knowledge of mechanisms in information security. Philos Tech. https://doi.org/10.1007/s13347-018-0329-z

Walzer M (1983) Spheres of justice: a defense of pluralism and equality. Basic Books, New York

Weber RH (2010) Internet of things–new security and privacy challenges. Comput Law Secur Rev 26(1):23–30

Weber K, Loi M, Christen M (2018) Digital medicine, cybersecurity and ethics: an uneasy relationship. Am J Bioeth 18(9):52–53

Wolff J, De-Shalit A (2007) Disadvantage. Oxford political theory. Oxford University Press, Oxford

Yaghmaei E, van de Poel I, Christen M et al (2017) Canvas White Paper 1 – cybersecurity and ethics. SSRN scholarly paper ID 3091909. Social Science Research Network, Rochester. https://papers.ssrn.com/abstract=3091909. Last access 7 July 2019