Ethical Approaches to Cybersecurity

Kevin Macnish and Jeroen van der Ham

The Oxford Handbook of Digital Ethics
Edited by Carissa Véliz

Subject: Philosophy, Moral Philosophy Online Publication Date: Mar 2022
DOI: 10.1093/oxfordhb/9780198857815.013.28

Abstract and Keywords

This chapter examines current research on cybersecurity ethics. It frames this around three different approaches to the subject. The first ('bottom up') considers ethical issues arising in different case studies and developing groupings of these issues, such as those relating to privacy, those to security, etc. The second approach ('pragmatist') considers ethical approaches currently used in cybersecurity practice, focusing on the confidentiality, integrity, availability (CIA) triad. The third approach ('top down') takes its starting point as broader ethical theories, which are then applied to cybersecurity. The authors present a novel top-down approach, defining security as the inverse of risk and then drawing on recent literature on the ethics of risk. The chapter concludes with a review of the strengths and weaknesses of each approach.

Keywords: cybersecurity, security, confidentiality, risk, privacy, justice

Introduction

The ethics of cybersecurity falls under the broader topic of the ethics of security. However, the ethics of security is an under-explored field in philosophy, and indeed more generally in academia. The greatest area of research into security has been in the discipline of International Relations and its sub-discipline of Security Studies. Yet here, the focus has traditionally been on the military and national security, turning in the past few decades to broader questions of political uses of security (the Copenhagen School) and issues of international security as seen through questions of emancipation (the Welsh School) or the movement of people across international borders (the Paris School) (Browning and McDonald 2011). None of these, however, translate easily into an understanding of the ethics of cybersecurity, which has domestic and international elements.

An obvious question arises as to whether there is even a need for an ethics of cybersecurity. Is not security, after all, a good thing? Clearly it is, but it does not follow that security can be achieved without cost, and those costs must be balanced against any gains in security. Such trade-offs may be not only financial, but also ethical. That we do not accept

security as the main, still less the only, value in society is clear from the fact that we reject notions of police states which may promise (however falsely) security from attack in exchange for a surrender of liberties. At the same time, nor do we generally embrace total anarchy in a celebration of those liberties. The relationship between security and liberty is more finely balanced than that, with liberty requiring security in order for liberties to be enjoyed. Hence security is valued, at least in part, for its securing of liberties.

Furthermore, security aims at enforcing divisions, keeping some people safe from other people or things. Such divisions may be justified in at least some cases, but not always. What should we say, for instance, of the security of a tyrant? Furthermore, security can never guarantee it will be 100 per cent successful. In practice, there are likely to be cases of false positives (those identified as attackers, who are not) and false negatives (those not identified as attackers, who should be). Hence, security can impose costs on the innocent and fail to protect as intended.

So, there are ethical issues that arise from and surround questions of the practice of security in general. It is not surprising then that there are similar issues and questions surrounding cybersecurity in particular. The ethics of cybersecurity seeks to identify, isolate, and address those issues and questions. Work examining the ethics of cybersecurity can therefore focus on the issues faced or at a more meta-level in considering the approaches taken to understanding those issues. We focus here on the latter.

In this chapter, we therefore examine three approaches to this task. The first is the ‘bottom-up’ approach of considering case studies and literature reviews to examine numerous incidents of cybersecurity successes and failures to identify ethical issues embedded in each. For readers interested in concrete cases, these can be found in abundance in this literature. The second approach is to look at the current practices in the professional domain of cybersecurity. These are, we argue, dominated by the triad of confidentiality, integrity, and availability of data, or CIA. This triad has been broadly taken to determine the success of cybersecurity since the 1980s, holding that all necessary criteria have been met provided the confidentiality of the data, the integrity of the data, and the availability of the data can be assured. There are, as we argue below, some significant weaknesses with this approach, though. The third approach that we consider is ‘top down’, through considering theoretical frameworks that can be applied to the subject. In each approach, we argue, there are notable strengths and weaknesses such that for a full picture of the ethics of cybersecurity, a combination of the three should be embraced.

The chapter progresses by looking at each of the three approaches in more depth, taking the order suggested above. The aim is to introduce the reader to current thinking in the field of cybersecurity ethics, while also presenting a novel approach that contributes to the advancement of the field. This novel approach involves a top-down framework through recognizing security as the inverse of risk, introducing into the discussion the framework of risk analysis and the ethics of risk.

Bottom-up approaches

Bottom-up approaches to ethics seek to identify case studies as a means to identify the ethical issues faced in a particular field. This has been a popular and fruitful angle with which to approach the relatively new field of cybersecurity ethics. In this section, we consider different bottom-up approaches and note the benefits provided and challenges raised by these approaches.

In the early twenty-first century, a significant focus of cybersecurity ethics was placed on e-voting, voting in democratic elections through electronic media rather than the tried-and-tested means of placing an ‘x’ next to the name of the preferred candidate or party on a piece of paper. The benefits of e-voting are, at least on the surface: immediate tallies, lack of confusion and spoiled ballot papers, and diminished scope for manipulation and corrupt practices. However, several commentators questioned whether these benefits were as straightforward as suggested. What happens, for example, if a voting machine, or worse, a network of such machines, is hacked, and when should such a hack be recognized publicly? Before the vote, the election could be delayed and the system undermined. After the vote, the results of the election would be undermined and open to challenge (Robinson and Halderman 2011).

In their study, Robinson and Halderman presciently questioned how researchers should respond to political pressure (not) to investigate the potential for interference in electronic voting and the scope for collateral damage through the networking of e-voting systems to potentially unrelated networked systems (Robinson and Halderman 2011). Pieters went further to question the notion of trust that the public might have in e-voting machines, recognizing that there are (at least) two competing notions of explanation concerning human–technology interaction: explanation-for-confidence and explanation-for-trust (Pieters 2011b). In the former, the focus is placed on providing confidence in a technology, which can be achieved without people knowing the risks involved (i.e. when arrived at through reassurance by an expert). In the latter, for trust to be given to a system, there must be a full understanding of the risks and alternatives involved. According to Pieters, the danger in e-voting machines is that authorities seek to provide the public with too much information (to establish trust) when the public only demands sufficient reassurance for confidence. The excess of information could then serve to undermine confidence in the system. Contrarily, if the public demands sufficient information for trust but is only given the information needed for confidence-building measures, then the disconnect could undermine trust in the system.

More recently, in-depth cybersecurity cases have been considered as illustrations of ethical concerns in cybersecurity. One of the best known cases is the ENCORE programme. ENCORE was a trojan horse (seemingly innocuous code which hosts malicious code) that could be implanted on a computer and then used to communicate from that computer. ENCORE was developed by a team of researchers at Princeton and Georgia Tech and hosted on a number of popular websites. Through people visiting those websites, the trojan was implanted on computers around the world, but particularly of note was its im-

plantation on domestic computers in China, Egypt, and Iran. From these computers, the research team was able to test the respective national firewalls to see which webpages were available to users in these countries.

Researchers wanted to understand the efficacy and prevalence of national firewalls designed to prevent citizens from accessing webpages that may be deemed 'undesirable' by the government (typically through challenging the legitimacy of that government), and which can only be effectively tested from inside the respective state (Burnett and Feamster 2015). However, this meant that the computers of non-consenting citizens were being used to attempt to access 'blocked' websites, potentially bringing those citizens to the attention of the authorities in their countries**.** Furthermore, given the nature of the states in question, those authorities are not known for their sympathy or for taking time to appreciate that the attempts were made by malicious software operated from the United States and not by the user themselves. In this way, cybersecurity research could place a user at considerable risk without consent or even awareness of the risk undertaken, in contravention of the most basic research ethics principles that have guided academic research since the Second World War (Byers 2015; Macnish 2019a).

A third bottom-up approach has taken the form of literature reviews, pulling together existing writings on cybersecurity ethics, such as those considered above, and hence taking a more holistic approach to the issues than any one case study. This approach has been followed, for example, by van der Poel (van de Poel 2020). In the associated study on which this chapter is based, Yaghmaei et al. carried out a literature review of 236 texts concerning cybersecurity in the fields of business, medicine, and national security (Yaghmaei et al. 2017). From the results of this review, van der Poel argues that the ethical issues arising can be clustered into four areas: security, privacy, fairness, and accountability. This approach has also been taken by Jha (2015), although he identifies eight clusters: equity, rights, honesty, exercise of corporate power, privacy, accuracy, property, and accessibility. The literature review approach is also the starting point for Macnish and Fernández (2019), who identified ten clusters.[1]

There are several advantages to the bottom-up approach typified in the above papers, not least that the issues speak clearly to practitioners through case studies grounded in real-world experience. The challenges with e-voting machines, with ENCORE, and with the issues identified by Macnish and Fernández are challenges faced in cybersecurity practice rather than relying on abstract philosophical theory or thought experiments, the relevance of which may be hard to see for someone not used to dealing with such a level of abstraction.

At the same time, such approaches have their problems. First, they risk missing important examples which may provide their own, unique issues. Hence, while bottom-up approaches may be thorough in their examination of particular cases, they cannot be systematic because they don't examine all cases. It is simply not feasible to examine every case for reasons of time and availability of cases (many, for example, never come to the attention of the public or academic researchers).

Second, in the same way as there may be latitudinal gaps in understanding the complete picture of cybersecurity challenges, there will also be longitudinal gaps as new issues arise with the advent of new attacks and security practices. As such, when new cases come to the attention of the public and academics, new analyses are needed to examine whether new ethical issues have emerged.

A third problem is that of clustering. This is a necessary practice to manage the large number of ethical issues associated with cybersecurity. As noted already, van der Poel clusters the issues into four categories, Jha into eight, and Macnish and Fernández into ten. In each case, though, the clustering appears to be largely ad hoc and lacks any clear theoretical underpinnings, meaning that there is scope for issues arising in one cluster to appear in another cluster (such as Jha's rights and property). For analytic purposes, one risks mis-identifying an issue as belonging to only one category when it may in fact fall into two or more, or possibly some other category not identified. This risk could result in ethical issues falling between the cracks, or being responded to as one type of ethical concern and not the other. The problem gets aggravated when new cases emerge that invite new clustering exercises. It is notable that the lack of agreement on clustering and its theoretical paucity is not always recognized. At present, the approaches taken tend to cluster issues around authors' perceived commonalities rather than adopting a pre-existing framework. Theoretically grounded taxonomies for cybersecurity have yet to be proposed.

Finally, while a top-down approach can be effective at recognizing the plurality of ethical issues arising in practice, it is not effective in providing suggestions as to how to deal with value conflicts. Without theoretical underpinning, the reader is left with a list of ethical challenges but little guidance as to how to prioritize them. Which trade-offs, for example, are tolerable and which are unacceptable? The resulting picture is often one of ethical plurality that is silent about general practical guidance.

The bottom-up approach is therefore a valuable contribution to the literature in providing a thorough, albeit not systematic, analysis of a variety of real-world problems. Through its grounded nature, the discussion surrounding these problems can be of benefit to practitioners as well as theoreticians. However, there is a number of drawbacks to the approach as it is currently practiced, not least its specificity, the lack of theoretical underpinning in the framework informing the clustering of issues, and in the means of prioritizing ethical conflicts that emerge in cybersecurity practice.

Pragmatist approaches

The second approach that we consider here is that of cybersecurity practitioners. This approach draws on the pragmatist theory of philosophers such as Dewey, Rorty, and Putnam, and is concerned with current practices and norms of behaviour, as opposed to focusing on case studies or looking to grand theories such as deontology or utilitarianism.

As LaFollette argues, for pragmatists ‘meaningful inquiry originates in practice’ (2000: 400).

As noted in the introduction, the values embraced by this community concern the confidentiality, integrity, and availability of data (the CIA triad). If data is kept confidential, if its integrity is maintained, and if it remains available to the pertinent user, then cybersecurity is successful. These could be seen as the fundamental values practiced in the cybersecurity community for nearly half a century.

The provenance of the CIA triad lies in the Anderson Report, one of the earliest publications on computer security, which discussed security exposure in networked environments. The triad was echoed in a later paper by Saltzer and Schroeder. The CIA abbreviation itself was coined by Steve Lipner around 1986 (private conversation). Since then, the term has been used widely in reports, standards, and other publications on cybersecurity (van der Ham 2020).

In the decades following the introduction of the CIA triad, the concept has been used as a de facto definition of security in texts on cybersecurity, ranging from textbooks to International Organisation for Standardisation (ISO) standards. The CIA triad is also a fundamental element of the Common Vulnerability Scoring System (CVSS). A CVSS score is used to describe the severity of a vulnerability in a software product. The CIA elements are used in the calculation of the impact of vulnerabilities in the scoring system. CVSS scores have been included in vulnerability publications since 2005 (FIRST CVSS SIG 2020).

The benefits of the CIA approach lie in the fact that it is time-tested by practitioners. As with the bottom-up approach discussed above, it is therefore grounded in real-world practice and has the advantage of a history of discussion and debate through which it has been refined over the past fifty years.

At the same time, the debate concerning the CIA triad has not settled. While the three concerns of confidentiality, integrity, and availability have largely remained core to cybersecurity practice, they have also been challenged. It has been suggested that the triad be supplemented with the values of non-repudiation, possession, and utility, known as the *Parkerian hexad* (Parker 1983, 2012). Others have suggested that each of confidentiality, integrity, and availability may not be as core to the practice as seems at first sight (Spring et al. 2019; van der Ham 2020).

A second challenge arises from the fact that the triad is highly technical in its application. Given the provenance of the triad, and a relative lack of philosophical engagement with it, the traditional definitions of C, I, and A are technical rather than ethical. This has led to the aspects of each being largely binary such that they are either true or false, or in the case of CVSS impact, on a vulnerability having a ‘high’, ‘low’, or ‘none’ measurement risk. This scale can give a false sense of accomplishment, as the current status gives no

guarantees about the future (or even the past); that is, data may currently be confidential, but it does not follow that it was confidential yesterday or will be tomorrow.

A third challenge is that this definition leads to a focus on the C, I, and A aspects of individual assets. Furthermore, each aspect is taken as an absolute value that must be upheld, instead of performing a holistic risk assessment that may impact the security of an individual asset (such as an individual computer, server, or entire system). This approach does not take into account the broader costs and benefits to whole organizations. Hence, confidentiality is a property of an individual asset, and usually not a property of a context, such as a computer network or office environment. The CIA triad leads to individual measures on objects instead of a general approach to cybersecurity.

In the early days of cybersecurity, many of the risks we face today were not there. When the CIA triad was proposed, computers were not connected to a (local) network, and the internet as we know it today did not exist. Both Anderson and Saltzer and Schroeder focused on theoretical work regarding computer security, since there were few practical attacks on computers at the time of their research. However, the context in which we use computers has evolved significantly. From single, offline, room-filling computers, we now have computers in our pockets that are constantly connected. Most artefacts in the physical world are now affected in some way or another by computers. Similarly, the nature of adversaries has changed significantly.

Traditional cybersecurity in the 1990s and early 2000s focused on providing security at the edges of networks. With the advent of mobile devices and the gain in popularity of bring your own device (BYOD), and more recently still the advent of the Internet of Things, this approach is no longer tenable. The boundaries of security have moved with these developments, and yet the overall approach to security has not, which has had the effect of strengthening the tendency to go for individualistic approaches to security (van der Ham 2020).

Furthermore, the focus on individual assets and meeting each of the requirements of confidentiality, integrity, and availability has led to a sticking-plaster mentality. As van der Ham has noted:

> the individual, binary measures and narrow focus in turn often lead to stop-gap solutions. Once a vulnerability threatens to break confidentiality, a measure is put in place to ensure confidentiality again. The risk associated with that vulnerability is then often not taken into account, confidentiality is now guaranteed again, so the problem appears to be solved. This is then often repeated many times for every new vulnerability.
>
> (van der Ham 2020)

The focus on technical solutions in turn risks missing opportunities for non-technical solutions to security challenges. Consider the increasingly complex rules to create safe passwords: one needs to include alpha-numeric and special characters, the password must

consist of at least 10 digits that do not form a memorable word, and it should be changed frequently. As each of these stipulations has developed in response to particular security challenges, the result is poor user security, given people's inability to maintain numerous safe passwords (Grassi et al. 2017). Solutions exist to provide more secure authentication methods (such as multifactor authentication or password managers), yet these do not see widespread use.

Finally, as with bottom-up approaches, the pragmatist approach is unable to provide guidance on key issues. This is a particular problem for emerging technologies for which norms are yet to be established or in contexts, such as cybersecurity, in which norms may have developed but lack sufficient theoretical underpinnings. For example, an ongoing concern in the cybersecurity community is that of vulnerabilities exploitation procedures. Recent years have seen active discussions on how security researchers should disclose security vulnerabilities; it is now deemed standard practice to privately warn organizations of existing security vulnerabilities. (Google has defined 90 days to be an acceptable period after which vulnerabilities can be publicly disclosed (Google 2019).) However, it is less clear how organizations and governments should deal with vulnerabilities that they discover (Pupillo et al. 2018). There are only two countries (the United States and the United Kingdom) that have released policies on how they deal with vulnerabilities that governmental agencies discover themselves (Ambastha 2019; Bradford Franklin 2019; Jaikar 2017; White House 2017). For commercial companies, there is very little guidance as to how they should deal with vulnerabilities and exploits, which some companies actively try to purchase or sell. Neither the CIA triad nor the Parkerian hexad has anything to contribute to this discussion; once more, practitioners are left without ethical support.

Cybersecurity expert Bruce Schnier has contributed extensively to this space over the past decade. Schneier's early work focused on technical aspects of cybersecurity (Schneier 2011, 2017, 2019) and while he remains a public figure in this area, much of his later writing has incorporated the societal and ethical impacts of cybersecurity. In particular, Schneier has been highly critical of security measures that may look as if they are working but are, in effect, of little value. Such 'security theatre' is problematic in failing to provide security while at the same time leading to significant reductions in individual and group liberty (Schneier 2009). As with Wolter Pieters and Mariarosaria Taddeo (below), Schneier has focused on the impact of cybersecurity, and the digital environment in general, on trust as necessary for the functioning of society (Schneier 2011). To this end, policies which lead to a false sense of security, or which collect excessive data beyond what is necessary are highlighted and dismissed in his work (Schneier 2014, 2015). Most recently, he has focused on the developing Internet of Things and the policies which, he argues, would be most effective in guaranteeing users' security (Schneier 2018) and on public-interest technologists, practitioners who try to contribute to the public good through developing tools to help society or contributing to policy development (Schneier 2020).

The pragmatist approach therefore provides some insight into the values espoused by the cybersecurity community and the practices currently embraced as a result of those values. However, the lack of theoretical underpinning for those values and the challenges posed by emerging technologies such as the Internet of Things have meant that this approach is weak on providing ethical guidance outside a very specific (technical) sphere, and even within that sphere it is weakening as a means of providing security, let alone considering the potential ethical costs of that security.

Top-down approaches

So far, this chapter has considered bottom-up approaches, which look at individual case studies concerning cybersecurity and ethics, and pragmatist approaches, which look at how the community practices cybersecurity. We turn now in this section to look at top-down approaches. While these could involve the application of traditional deontological or utilitarian theories to cybersecurity, we are not aware of any sustained attempts to do this in the field.[2] Instead, we will look at three approaches: Pieters's application of Bruno Latour's actor network theory (ANT) and systems theory, Mariarosaria Taddeo's work on balance and trust in liberal societies, and the novel approach we propose here of applying the concept of security as risk. Each of these is top down in considering a unique theoretical perspective which is then applied to cybersecurity as a means of understanding ethical issues, rather than attempting to isolate the individual issues themselves or the practices of those working in the field.

Pieters has argued that Latour's ANT provides new and important insights into cybersecurity. ANT holds that human beings and technical artefacts are both actors in broader networks. Through the eyes of this approach, as Pieters notes, shooting is performed neither by a person without a gun, nor by a gun without a person. The person enables the gun to shoot as much as the gun enables the person to shoot. As such, both gun and person are equal actors in this particular network. Which actor initiates the action is, according to Pieters's account of this line of thinking, irrelevant. What is important to him is that it is the combination of the actors that performs the action.

The advantage of this approach, argues Pieters, is that it allows for the recognition of human factors in assessing security modelling. This human element is often discounted from models as human motives and actions are taken to be far less predictable than non-human systems. This discounting leads to a significant gap in these models (Dimkov et al. 2008, 2010; Probst and Hansen 2008). Rather than discounting the human factor, particularly when interaction between various human elements is concerned, Pieters proposes that the 'flat' model of ANT in which humans and artefacts are treated equally helps to remove apparent complications. Rather than looking to motivations, the focus shifts to the information that is moved.

Pieters has also drawn on Niklas Luhmann's systems theory to emphasize the human element in the construction of information security (Pieters 2011a). As with ANT, systems theory involves the human element in the overarching system. Security is necessarily a

human construct, argues Pieters, given that ‘actual security is dependent on perceived security’ (Pieters 2011a: 333), by which he means that the objective security of a network is partly dependent on how effective that network’s security is perceived to be by potential attackers. Pieters has argued that the systems theoretic approach can also help to highlight ethical issues that may otherwise be occluded (Pieters 2017). For example, in the debate surrounding e-voting, considerable emphasis was placed on the privacy of individuals. This was, at least in part, argues Pieters, because there are clear legal regulations concerning privacy in the voting booth. However, there is a risk of excessive focus in this area because of the practical connection to the law, which risks commentators missing other (arguably more) important ethical elements, such as power, trust, and security. Privacy of individual voters was, holds Pieters, a relatively minor concern next to the potential for the outcome of the voting process to be manipulated through a successful attack on the network hosting the voting machines.

Pieters’s account demonstrates the advantage that drawing on generalized theories in philosophy, and particularly those in the Continental tradition, can offer to an understanding of cybersecurity ethics. New insights are proposed that can be used to approach existing problems in new ways. At the same time, neither ANT nor systems theory are uncontroversial in the philosophy of technology. While Pieters’s arguments are predominantly pragmatic, demonstrating the benefits that can arise for cybersecurity practice from adopting this view, they risk deflecting the argument to philosophical discussions as to the merits or otherwise of Latour and Luhmann on a more general level.

As with Pieters, Mariarosaria Taddeo’s early work focused on questions of trust, and particularly trust in digital environments. In her case, Taddeo focused on whether trust could exist in online environments, arguing that it was clearly possible (Taddeo 2009, 2010a, 2010b; Turilli et al. 2010). From here, she moved towards cyberwarfare, where she became particularly noted for her work on using the just war tradition as a framework for analysis of cyberwarfare operations (Floridi and Taddeo 2016; Taddeo 2012, 2016), including an introduction to the ethics of cyber-conflicts (Taddeo 2021). While the military extension of cyber-operations has been the major focus of Taddeo’s work, she has also written on the ethics of civilian cybersecurity, such as we are considering in this chapter. In 2013, Taddeo edited a special edition of the journal *Philosophy and Technology*, focusing on balancing online security and civil rights (Taddeo 2013) which included articles on proportionality in cybersecurity (Hildebrandt 2013) and Thomism as a lens for interpreting ethics in online environments (Dainow 2013).

More recently, Taddeo’s work has incorporated a focus on artificial intelligence (AI), bringing together her earlier work on trust and cyberwarfare (Taddeo 2019; Taddeo et al. 2019) and including the impact of AI on cybersecurity (Taddeo 2019). This last is a developing field, with the impact of AI on attacks and defence still being determined (Patel et al. 2019). Throughout her work, Taddeo’s focus has remained largely on the need to find balance in liberal societies between stability in the online environment, achieved through the establishment of reliable environments, and trust of users in the maintenance of their rights in that environment. To this end, she has argued that trustworthy environments

should consist of system robustness, system resilience, and system response (Taddeo 2019; Taddeo et al. 2019). As with Pieters, then, Taddeo's work does not amount to a blunt application of a 'grand theory' approach to cybersecurity or cyberwarfare. However, through appeal to liberal notions of striking a balance between authority and individual liberty (Taddeo 2013) and the need to create trustworthy environments online through which users can feel safe to engage in free expression, she falls into what we have called the 'top-down' approach to cybersecurity ethics.

We propose an alternative top-down approach, and, we believe, one less controversial than either those of Latour or Luhmann and more structurally cohesive than that of Taddeo. We contend that security is the inverse of risk. If we take risk to be a function of probability and of harm (Hansson 2013), then as the probability of a harm occurring increases, so risk increases (and security decreases). Likewise, as the severity of the harm increases, so risk also increases and security decreases. Conversely, a decrease in the severity of harm threatened or the probability of that harm arising equates to an increase in security. In itself this is not, we believe, a radical solution and is indeed one largely adopted in industry, where security is often referred to as 'risk management'. However, as we demonstrate below, through drawing on recent work in the philosophy of risk, it promises significant insights into the ethics of cybersecurity.

The approach that we advocate draws on earlier conceptual analyses of security provided by Arnold Wolfers and David Baldwin (Wolfers 1952; Baldwin 1997), while also introducing the framework of ethical risk analysis to security in general and cybersecurity in practice (e.g. Hansson 1996, 2013). In what follows, we develop in brief the concept of security as the inverse of risk in general before applying that to cybersecurity and demonstrating the strengths that come from introducing the ethics of risk as a framework through which we can approach the ethics of cybersecurity.

First, Wolfers originally argued that 'security in any objective sense, measures the absence of threats to acquired values, in a subjective sense, the absence of fear that such values will be attacked' (Wolfers, 1952: 485). Wolfers provided an early recognition that there were both subjective and objective aspects of security in that a person could believe they are secure without being secure, or, contrariwise, believe they are not secure while actually being secure. To this distinction, Herington has more recently introduced a third, non-cognitive dimension of affect (Herington 2018: 181–185); it holds that a person could be secure (objective), believe that they are secure (subjective), but not *feel* secure (affective). In this way, it is possible to think or feel that a network is secure even when it is not, or to feel that a network is insecure when it is in fact highly secure.

This is not to say that objective and subjective states of security are unrelated. The above examples are extreme, and it is more likely that I will believe and feel marginally more or less secure than is in fact the case. Take the case of Ross and Rachel. Rachel loves Ross as an objective state. However, the chronically insecure Ross is scared that Rachel's love for him is fleeting at best, and so he is subjectively (in terms of both belief and affect) insecure. This causes Ross constantly to ask Rachel whether she still loves him, a practice

that annoys her. As a result of his persistent questioning of her love for him, she starts to draw away from Ross emotionally, which leads to an increase in his insecurity. A vicious circle develops such that eventually his subjective insecurity means that she ceases to love him. His subjective state of security in the relationship has therefore brought about a change in his objective state of security in that relationship.[3]

Second, Wolfers argued that security consists of a threat to values. Baldwin has challenged this point by noting that threats typically involve intentionality, and yet security is broader than this, involving the security of assets that can be challenged by natural events which lack intentionality (such as physical assets that can be endangered by flooding). Baldwin's response is to define security as 'a low probability of damage to acquired values' (Baldwin 1997: 13).[4] While we embrace this move, we argue that one can go further by understanding security as a function of probability and severity of damage to values. Given that probability and severity are aspects of risk, we hold that (high levels of) security involve(s) a low probability of severe harm and vice versa. Hence, when there is a low probability of harm occurring to a valued network, that network enjoys a relatively high level of security. Once the probability of harm occurring increases, or the severity of the harm which might occur intensifies, then the security of that network diminishes.

Baldwin's response to Wolfers, seen in the current context, raises the challenge as to the relevance of the intentionality of the attacker in a cybersecurity incident. Our proposal, in line with that suggested by Pieters, adopts a flatter approach that does not distinguish between human and non-human causes of cybersecurity incidents. While the presence of intention is traditionally seen as the distinguishing factor between security incidents (which require intention of an attacker) and safety incidents (which presume that there is no intention to attack), we think that this is an increasingly dated and irrelevant assumption. We hold that the intention of the attacker, while it may be pertinent in judging the person, is unimportant in approaching the *concept* of security and its ethical outworking as discussed here.

From the perspective of providing cybersecurity, challenges to the network may be posed by people or by natural events. Which of these is behind the challenge will likely have an impact on the response that is taken to that threat (a security measure), but not to the *level* of security or to the responsibility of those who are tasked with ensuring security. Analogously, we might say that my home could be equally under threat of destruction from a tornado or a terrorist. In either case, my security is comparatively low, irrespective of which threat is being considered. How I respond to the challenge posed will differ, but the actual state of security of my home is not a function of whether the challenge is posed by an entity or event with intention.

Third, we recognize a descriptive (as opposed to normative) benefit in the Wolfer's definition of security as the absence of threat to *acquired values*. Given that Wolfers was writing in the context of international relations, we presume that he is thinking about values to different states. In our broader context, though, the reference to values (or severity, which is a reflection of that which is valued) allows for any value to fill that space. Wal-

dron argues that we could expand the context 'to refer to the assured possession and enjoyment of any value, including liberty' (Waldron 2006: 504). However, in a descriptive account one should be able to cope with values with which one does not agree. Hence, it is meaningful to discuss the security of the Nazi party in Germany in 1945, even when we ourselves do not value that particular case of security. In the case of cybersecurity, then, it is similarly meaningful to discuss the security of a drug dealer's mobile device, even while law enforcement is seeking a way to dismantle that security and track the device. The security of the device may be valuable to the dealer, while the same security is not desirable to those in law enforcement attempting to hack the device.

Fourth, the notion of probability in the conceptual definition holds to a more recently identified aspect of security, in that it is always future-focused. While Waldron has suggested that this future focus concerns the security of that which is valued extending into the future (Waldron 2006: 474), Herington is more circumspect about this, recognizing merely the future focus on the concept (Herington 2012: 18). In this, we agree with Herington that security is about the future. Security concerns, for example, the safety of a person from attack or from being injured. Once they *are* attacked or injured, security ceases to be the pertinent concept. Rather than thinking of the person's security, we think of the harm they are experiencing. Indeed, the very fact that someone is experiencing harm suggests that any security they had prior to the harm either failed or was insufficient to the risk.

Finally, as noted by Herington, security always has a referent (Herington 2012: 13; see also Newey et al. 2012: 9). Security is the security of a person or thing (referent). However, to this we add that security always has a context: security is the security of an entity (referent) from a challenge (context). Hence, as noted above, we can discuss the security of a network from an attacker or from an earthquake, or the security of a mobile device from being hacked. This recognition side-steps the problems noted earlier in the CIA triad, which tends to focus attention merely on technical solutions in response to risks to individual assets. Broadening our understanding of the relevant referents and contexts encourages a more holistic approach to cybersecurity.

Understanding security in terms of the inverse of risk is not only an effective means of approaching security in general and cybersecurity in particular; it also carries with it the benefits that can come from applying the ethics of risk literature to the ethics of cybersecurity. The ethics of risk highlights ethical concerns that other approaches miss.

To develop an ethics of cybersecurity from the ethics of risk literature, we draw primarily on the work of Sven Ove Hansson, who has dominated the latter field in recent years. In particular, we hold that work in four areas is demonstrative of the benefits that a combining of the two fields can bring. These areas are: the distinction between objective and subjective harms, challenges in calculating probabilities, the recognition of fallacies, and the problems arising from risk thresholds and distribution. We will take these in turn.

There is a concern about the calculation of harms as only objective. Granted, there is often an objective element to harm (the loss of a hand is harmful to anyone); but there is also a subjective element. It is presumably worse for a concert violinist to lose their hand than for a philosopher to do so. With only one hand, the philosopher could continue to practice their profession while the violinist would not be able to do so. In the case of cybersecurity, as noted above, this may extend further, such that the drug dealer values the security of their mobile device very highly, while pursuing law enforcement agents will not value that security (to the extent that they will attempt to compromise it). Any objective value of that security is questionable in this case (Hansson 2010).

The second concern is that of calculating probabilities. While we can largely agree on the probability of a traffic accident occurring to us being more likely than being caught up in a terrorist incident, this does not prevent the latter seeming more likely than the former as a result of persistent reporting in the press and news media. Hence, there is an important subjective element in probability calculation. Furthermore, in calculating probabilities there are potentially infinite complications that could occur to make radical changes to those statistics. In order to manage these complications, externalities and feedback loops are typically simplified, or even removed. While this leads to a 'cleaner' resulting statistic, it is also a highly theoretical and unrealistic statistic precisely because of the simplification. The appearance may therefore be of certitude when in fact such certainty is misplaced. Hansson has described this as the *tuxedo fallacy*, imagining that we are calculating probabilities in an idealized casino rather than in the real world (Hansson 2009).

Third, there are problems arising from risk toleration thresholds and the distribution of risk. Risk toleration thresholds—the limits at which we are prepared to tolerate risks—have been demonstrated to differ among groups in (western) society. White males have frequently been shown to be more tolerant of risk than White women or non-White men (Flynn et al. 1994; Hermansson 2010). This tendency is perhaps not surprising in societies which have largely been designed by White men. For example, when Twitter was designed, by predominantly White men in Silicon Valley, the issue of potential harassment apparently did not enter the discussion as one possible outcome (McCaskill 2015).

When new risks are introduced into social discourse, it may appear that particular courses of action are preferable owing to the fact that they are 'less risky' than existing courses. Hansson describes this as the 'sheer size' fallacy, noting that just because B is less risky than A, and A is currently employed in society, it does not follow that B is the best course of action to follow (Hansson 2004). It may be that both A and B are unacceptably risky for the majority of society. Furthermore, through adopting B we may cease to look for less obvious alternatives (C) that would fall below the risk toleration threshold of a majority. Hence, in the Patriot Act, the US administration and the associated intelligence agencies apparently felt that it was less risky for those intelligence agencies to monitor the metadata of mobile phone calls of every person in the country than to allow for further terrorist attacks to occur. Following the Snowden revelations, though, it transpired

that not everyone agreed; possibly a majority of citizens in the United States disagreed (Rainie and Madden 2015; Stoycheff 2016; Bakir 2015; Krueger 2005).

Fourth, and related to the problem of the sheer size fallacy, is the concern with the distribution of risk. Jonathan Wolff has noted that the agent who makes a decision involving risk is not always the same agent who will benefit from that risk, or indeed pay the costs of that risk (Wolff 2010). For example, in the infamous example of Ford's decision not to recall the highly dangerous Pinto car in the United States in the 1970s, that decision was made for financial reasons by the board of Ford, who stood to benefit from not paying the costs of a recall (Macnish 2019b; Mcginn 2018). By contrast, the members of that board were almost certainly not driving Pintos themselves; the people paying the costs of that decision were not the beneficiaries. Such a scenario is ethically problematic as it enables risk-prone decision-makers to indulge in seeking their own gain at the expense of other people.

An analogous scenario to that of Pinto happens in cybersecurity when businesses underinvest in cybersecurity. In the event of a breach arising from such underinvestment where customer information is stolen, the business may suffer reputational damage and, increasingly, legal fines. However, it is the customers who suffer the loss of significant personal data, such as email addresses, physical addresses, and credit card details.

In summary, there is a number of benefits that can be drawn from a reconsideration of the concept of security in light of the concept of risk and the ethics of risk literature. Note that our account is still subject to some of the earlier criticisms levelled at bottom-up approaches, though. For example, much of the ethics of risk literature has itself been developed through bottom-up analyses, and therefore provides insights that are drawn from the sort of clustering against which we cautioned earlier in this chapter. Furthermore, there is no one theoretical stance on risk, still less an ethical framework for approaching risk that will provide the systematic approach that we argued was lacking from bottom-up approaches.

Moreover, this approach is highly theoretical in its stance, in that it defines cybersecurity and then seeks to identify ethical concerns arising through the definition. By being highly theoretical, it risks both becoming untethered from reality if the conceptual analysis is flawed and missing central ethical concerns that are identified through bottom-up and pragmatist approaches.

At the same time, through this approach we have introduced novel concerns that have not been discussed in the ethics of cybersecurity literature arising from bottom-up or pragmatist approaches. Concerns about measurement of probability (the tuxedo fallacy) for example, are real issues in a field increasingly referring to itself as 'risk management', in which risks are calculated to a precise degree by insurance companies. The fact that such risks cannot be calculated with a degree of accuracy should, we hold, provide pause for thought and caution in promises that can reasonably be made by cybersecurity providers. Likewise, differing risk thresholds throughout society and the problems of risk distribution coupled with White male bias imply that current cybersecurity risk thresholds may

fail to take account of certain sectors in society (particularly women, the poor, and ethnic minorities), while imposing risks on those same sectors should once more lead to serious reflection. To our knowledge, neither of these concerns has been raised in the ethics of cybersecurity literature discussed in this chapter.

Conclusion

In this chapter we have considered three approaches to understanding the ethics of cybersecurity. These three, bottom up, pragmatist, and top down, have each been shown to have strengths and weaknesses. Our own preferred approach of understanding security as the inverse of risk carries some strengths not apparent in other approaches, and we hold that it is valuable in highlighting ethical concerns that are overlooked by those other approaches.

It is also noteworthy that the approaches considered here are, to date, the leading approaches to cybersecurity ethics. There is little to no work explicitly exploring more traditional interpretations deriving from Kantian, rights-based, or utilitarian approaches, and yet these may bear considerable fruit, either through their own insights or through contributions to existing approaches.

We therefore believe that to continue to develop this new field in a thorough and systematic manner all three approaches are valuable. The development of case studies can continue to introduce new concerns and insights, while pragmatist approaches can track the development of new technologies and associated issues arising therefrom. Finally, the insights gained from the ethics of risk, and from broader theoretical approaches such as Latour's actor network theory can also provide insights that may not be obvious to those practicing cybersecurity. Furthermore, there is a wealth of insight to be gained from exploring utilitarian and deontological approaches to cybersecurity which remains untapped. Pursuing any one of these approaches to the exclusion of the others will risk leading to an overly narrow perspective that misses crucial concerns. Drawing them together into a cohesive whole promises to lead to a more thorough and complete understanding of the problems, coupled with a clearer way forward for practitioners.

References

Ambastha, Mimansa (2019), 'Taking a Hard Look at the Vulnerabilities Equities Process and Its National Security Implications'. *Berkeley Tech Law Journal Blog*. **https://btlj.org/2019/04/taking-a-hard-look-at-the-vulnerable-equities-process-in-national-security/**.

Bakir, Viand (2015), '"Veillant Panoptic Assemblage": Mutual Watching and Resistance to Mass Surveillance after Snowden', *Media and Communication* 3(3), 12–25.

Baldwin, David A. (1997), 'The Concept of Security', *Review of International Studies* 23(1), 5–26.

Bradford Franklin, Sharon (2019), 'The Need for Countries to Establish Robust and Transparent Vulnerabilities Equities Processes', *Fletcher Security Review* 6, 45.

Browning, Christopher S., and McDonald, Matt (2011), 'The Future of Critical Security Studies: Ethics and the Politics of Security', *European Journal of International Relations* 19(2), 235–255.

Burnett, Sam, and Feamster, Nick (2015), 'Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests', in *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*, New York, 653–667. doi: **https://doi.org/10.1145/2785956.2787485**.

Byers, John W. (2015), 'Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests—Public Review', Technical Report, **https://conferences.sigcomm.org/sigcomm/2015/pdf/reviews/226pr.pdf**, accessed 1 October 2021.

Dainow, Brandt (2013), 'What Can a Medieval Friar Teach Us About the Internet? Deriving Criteria of Justice for Cyberlaw from Thomist Natural Law Theory', *Philosophy & Technology* 26(4), 459–76. doi: **https://doi.org/10.1007/s13347-013-0110-2**.

Dimkov, Trajce, Wolter Pieters, and Pieter Hartel (2010), 'Portunes: Representing Attack Scenarios Spanning Through the Physical, Digital and Social Domain', in Alessandro Armando and Gavin Lowe, eds., *Joint Workshop on Automated Reasoning for Security Protocol Analysis and Issues in the Theory of Security* (Berlin, Heidelberg: Springer), 112–129.

Dimkov, Trajce, Qiang Tang, and Pieter H. Hartel (2008), 'On the Inability of Existing Security Models to Cope with Data Mobility in Dynamic Organizations', MODSEC@ MoDELS. **https://ris.utwente.nl/ws/portalfiles/portal/5100076/final_sent_at_31_august.pdf** (accessed 9 October 2021).

FIRST CVSS SIG. (2020), 'CVSS v3.1 Specification Document', FIRST—Forum of Incident Response and Security Teams, **https://www.first.org/cvss/v3.1/specification-document**, accessed 1 October 2021.

Floridi, Luciano, and Taddeo, Mariarosaria, eds (2016), *The Ethics of Information Warfare*, softcover reprint of original 2014 edn (Cham, Switzerland: Springer).

Flynn, James, Slovic, Paul, and Mertz, Chris K. (1994), 'Gender, Race, and Perception of Environmental Health Risks', *Risk Analysis* 14(6), 1101–1108.

Google (2019), 'Project Zero: Vulnerability Disclosure FAQ', *Project Zero* (blog), 31 July, **https://googleprojectzero.blogspot.com/p/vulnerability-disclosure-faq.html**, accessed 1 October 2021.

Grassi, Paul A., Perlner, Ray A., Newton, Elaine M., Regenscheid, Andrew R., Burr, William E., Richer, Justin P., Lefkovitz, Naomi B., Danker, Jamie M., Choong, Yee-Yin, Greene, Kristen K., Theofanos, Mary F. (2017), 'Digital Identity Guidelines: Authentication and Lifecycle Management [includes updates as of 03-02- 2020], Special Publication

(NIST SP), National Institute of Standards and Technology, Gaithersburg, MD, [online], **https://doi.org/10.6028/NIST.SP.800-63b** (accessed October 9, 2021).

Ham, Jeroen van der (2020), 'Towards a Better Understanding of Cybersecurity', *ACM Journal of Digital Threats: Research and Practice (DTRAP)* 2(3), 1–3.

Hansson, Sven Ove (1996), 'What Is Philosophy of Risk?', *Theoria* 62(1–2), 169–186. doi: **https://doi.org/10.1111/j.1755-2567.1996.tb00536.x**.

Hansson, Sven Ove (2004), 'Fallacies of Risk', *Journal of Risk Research* 7(3), 353–360. doi: **https://doi.org/10.1080/1366987042000176262**.

Hansson, Sven Ove (2009), 'From the Casino to the Jungle', *Synthese* 168(3), 423–432.

Hansson, Sven Ove (2010), 'Risk: Objective or Subjective, Facts or Values', *Journal of Risk Research* 13(2), 231–238. doi: **https://doi.org/10.1080/13669870903126226**.

Hansson, Sven Ove (2013), *The Ethics of Risk: Ethical Analysis in an Uncertain World* (New York: Palgrave Macmillan).

Herington, Jonathan (2012), 'The Concept of Security', in Michael Selgelid and Christian Enemark, eds, *Ethical and Security Aspects of Infectious Disease Control: Interdisciplinary Perspectives* (London: Routledge), 7–26.

Herington, Jonathan (2018), 'The Contribution of Security to Well-Being', *Journal of Ethics & Social Philosophy* 14, 179.

Hermansson, Hélène (2010), 'Towards a Fair Procedure for Risk Management', *Journal of Risk Research* 13(4), 501–515. doi: **https://doi.org/10.1080/13669870903305903**.

Hildebrandt, Mireille (2013), 'Balance or Trade-Off? Online Security Technologies and Fundamental Rights', *Philosophy & Technology* 26(4), 357–379. doi: **https://doi.org/10.1007/s13347-013-0104-0**.

Jaikar, Chris (2017), *Vulnerabilities Equities Process* (Washington DC: Congressional Research Service).

Jha, Davendranath (2015), 'Importance of Morality, Ethical Practices and Cyber Laws as Prelude to Cybersecurity', *CSI Communications* 39(2), 29–32.

Krueger, Brian S. (2005), 'Government Surveillance and Political Participation on the Internet', *Social Science Computer Review* 23(4), 439–452.

LaFollette, Hugh (2000), 'Pragmatic Ethics', in Hugh LaFollette, ed., *Blackwell Guide to Ethical Theory* (Malden, MA: Blackwell Publishing), 400–419, **https://digital.usfsp.edu/fac_publications/2242**, accessed 1 October 2021.

Macnish, Kevin (2019a), 'Informed Consent', in Carissa Veliz, ed., *Data, Privacy and the Individual* (Madrid: IE University Press), 1–16.

Macnish, Kevin (2019b), 'Introduction to Privacy', in Carissa Veliz, ed., *Data, Privacy and the Individual* (Madrid: IE University Press), 1–17.

Macnish, Kevin, and Fernández, Ana (2019), 'Smart Information Systems in Cybersecurity', *ORBIT Journal* 1(2). doi: **https://doi.org/10.29297/orbit.v2i2.105**.

Manjikian, Mary (2017), *Cybersecurity Ethics*, 1st edn (London and New York: Routledge).

Mccaskill, Nolan D. (2015), 'Twitter's Diversity Problem', *The Agenda*, 9 March, **https://www.politico.com/agenda/story/2015/09/twitters-diversity-problem-000218**, accessed 1 October 2021.

Mcginn, Robert (2018), *The Ethical Engineer: Contemporary Concepts and Cases* (Princeton, NJ: Princeton University Press).

Newey, Glen, (2012), 'Liberty, Security Notwithstanding', in Charles Husband and Yunis Alam, eds., *Social Cohesion, Securitization and Counter-Terrorism* (Helsinki: Helsinki Collegium for Advanced Studies) **https://helda.helsinki.fi/handle/10138/32359**, accessed 9 October 2021.

Parker, Donn, B. (1983), *Fighting Computer Crime* (New York: Scribner).

Parker, Donn, B. (2012), 'Toward a New Framework for Information Security?', in Seymour Bosworth, Michael E. Kabay, and Eric Whyne, eds., *Computer Security Handbook*, 6th ed. (New Jersey: Wiley), 3.1–3.23.

Patel, Andrew, Hatzakis, Tally, Macnish, Kevin, Ryan, Mark, and Kirichenko, Alexey (2019), 'D1.3 Cyberthreats and Countermeasures', Online resource D1.3. SHERPA, De Montfort University. doi: **https://doi.org/10.21253/DMU.7951292.v3**.

Pieters, Wolter (2011a), 'The (Social) Construction of Information Security', *The Information Society* 27(5), 326–335.

Pieters, Wolter (2011b), 'Explanation and Trust: What to Tell the User in Security and AI?', *Ethics and Information Technology* 13(1), 53–64. doi: **https://doi.org/10.1007/s10676-010-9253-3**.

Pieters, Wolter (2017), 'Beyond Individual-Centric Privacy: Information Technology in Social Systems', *The Information Society* 33(5), 271–281.

Poel, Ibo van de (2020), 'Core Values and Value Conflicts in Cybersecurity: Beyond Privacy Versus Security', in Markus Christen, Bert Gordijn, and Michele Loi, eds, *The Ethics of Cybersecurity*, The International Library of Ethics, Law and Technology (Cham: Springer International), 45–71. doi: **https://doi.org/10.1007/978-3-030-29053-5_3**.

Probst, Christian W., and René Rydhof Hansen (2008), 'An Extensible Analysable System Model', *Information Security Technical Report* 13(4), 235–246.

Pupillo, Lorenzo, Ferreira, Afonso, and Varisco, Gianluca (2018), 'Software Vulnerability Disclosure in Europe: Technology, Policies and Legal Challenges', *CEPS*, 28 June, **https://www.ceps.eu/ceps-publications/software-vulnerability-disclosure-europe-technology-policies-and-legal-challenges/**, accessed 1 October 2021.

Rainie, Lee, and Madden, Mary (2015), 'Americans' Privacy Strategies Post-Snowden', *Pew Research Center: Internet, Science & Tech* (blog), 16 March, **http://www.pewinternet.org/2015/03/16/americans-privacy-strategies-post-snowden/**, accessed 1 October 2021.

Robinson, David G., and Halderman, J. Alex (2011), 'Ethical Issues in E-Voting Security Analysis', in George Danezis, Sven Dietrich, and Kazue Sako, eds, *Financial Cryptography and Data Security*, Lecture Notes in Computer Science 7126 (Berlin: Springer), 119–130.

Schneier, Bruce (2009), *Schneier on Security*, 1st edn (Indianapolis: Wiley).

Schneier, Bruce (2011), *Secrets and Lies: Digital Security in a Networked World*, 1st edn (Indianapolis: Wiley).

Schneier, Bruce (2014), 'NSA Robots Are "Collecting" Your Data, Too, and They're Getting Away with It', *The Guardian*, 27 February, sec. Comment is free, **http://www.theguardian.com/commentisfree/2014/feb/27/nsa-robots-algorithm-surveillance-bruce-schneier**, accessed 1 October 2021.

Schneier, Bruce (2015), *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*, 1st edn (New York: W. W. Norton & Company).

Schneier, Bruce (2017), *Applied Cryptography: Protocols, Algorithms and Source Code in C*, 20th edn (Indianapolis: Wiley).

Schneier, Bruce (2018) *Click Here to Kill Everybody: Security and Survival in a Hyper-Connected World*, repr. edn (New York: W. W. Norton & Company).

Schneier, Bruce (2019), *We Have Root: Even More Advice from Schneier on Security*, 1st edn (Indianapolis: Wiley).

Schneier, Bruce (2020), 'Public-Interest Technology Resources', *Public-Interest Technology Resources* (blog), 24 February, **https://public-interest-tech.com/**, accessed 1 October 2021.

Spring, Jonathan M., Hatleback, Eric, Householder, Allen, Manion, Art, and Shick, Deana (2019), 'Prioritizing Vulnerability Response: A Stakeholder-Specific Vulnerability Categorization', *Carnegie Mellon University Software Engineering Institute*, November, 36.

Stoycheff, Elizabeth (2016), 'Under Surveillance: Examining Facebook's Spiral of Silence Effects in the Wake of NSA Internet Monitoring', *Journalism & Mass Communication Quarterly* 93(2), 1–16.

Taddeo, Mariarosaria (2009), 'Defining Trust and E-Trust: From Old Theories to New Problems', International Journal of Technology and Human Interaction (IJTHI) 2(April), **http://www.igi-global.com/article/defining-trust-trust/2939**, accessed 1 October 2021.

Taddeo, Mariarosaria (2010a), 'Trust in Technology: A Distinctive and a Problematic Relation', *Knowledge, Technology & Policy* 23(3–4), 283–286.

Taddeo, Mariarosaria (2010b) 'Modelling Trust in Artificial Agents, A First Step Toward the Analysis of e-Trust', *Minds and Machines* 20(2), 243–257. doi: **https://doi.org/10.1007/s11023-010-9201-3**.

Taddeo, Mariarosaria (2012), 'Information Warfare: A Philosophical Perspective', *Philosophy & Technology* 25(1), 105–120. doi: **https://doi.org/10.1007/s13347-011-0040-9**.

Taddeo, Mariarosaria (2013), 'Cyber Security and Individual Rights, Striking the Right Balance', *Philosophy & Technology* 26(4), 353–356. doi: **https://doi.org/10.1007/s13347-013-0140-9**.

Taddeo, Mariarosaria (2016), 'Just Information Warfare', *Topoi* 35(1) 213–224. doi: **https://doi.org/10.1007/s11245-014-9245-8**.

Taddeo, Mariarosaria (2019), 'Three Ethical Challenges of Applications of Artificial Intelligence in Cybersecurity', *Minds and Machines* 29(2), 187–91. doi: **https://doi.org/10.1007/s11023-019-09504-8**.

Taddeo, Mariarosaria (2021), *The Ethics of Cyber Conflicts: An Introduction*, 1st edn (London: Routledge).

Taddeo, Mariarosaria, McCutcheon, Tom, and Floridi, Luciano (2019), 'Trusting Artificial Intelligence in Cybersecurity Is a Double-Edged Sword', *Nature Machine Intelligence* 1(12), 557–560. doi: **https://doi.org/10.1038/s42256-019-0109-1**.

Turilli, Matteo, Vaccaro, Antonino, and Taddeo, Mariarosaria (2010), 'The Case of Online Trust', *Knowledge, Technology & Policy* 23(3), 333–345. doi: **https://doi.org/10.1007/s12130-010-9117-5**.

Waldron, Jeremy (2006), 'Safety and Security', *Nebraska Law Review* 85, 454.

White House (2017), 'Vulnerabilities Equities Policy and Process for the United States Government', White House Report, Washington, DC. **https://trumpwhitehouse.archives.gov/sites/whitehouse.gov/files/images/External%20-%20Unclassified%20VEP%20Charter%20FINAL.PDF**, accessed 9 October 2021.

Wolfers, Arnold (1952), '"National Security" as an Ambiguous Symbol', *Political Science Quarterly* 67(4), 481–502. doi: **https://doi.org/10.2307/2145138**.

Wolff, Jonathan (2010), 'Five Types of Risky Situation', *Law, Innovation and Technology* 2(2), 151–63. doi: **https://doi.org/10.5235/175799610794046177**.

Yaghmaei, Emad, van de Poel, Ibo, Christen, Markus, Weber, Karsten, Gordijn, Bert, Kleine, Nadine et al. (2017), 'Canvas White Paper 1—Cybersecurity and Ethics', *SSRN ELibrary*, **https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3091909**, accessed 1 October 2021.

Notes:

(1) While this study involves a less thorough literature research than van de Poel, the literature provided a background to interviews held with cybersecurity researchers and practitioners to build an understanding of the areas in which cybersecurity ethics literature complements cybersecurity practice and areas in which the two diverge.

(2) Arguably, Manjikian's *Cybersecurity Ethics: An Introduction* does examine various ethical concerns in cybersecurity from various traditional viewpoints. However, this is an introductory text and refrains from defending one particular approach, such as a Kantian or utilitarian framework, but rather demonstrates how Kantians, utilitarians, and virtue ethicists might each approach the concern (Manjikian 2017).

(3) We are grateful to Bert Gordijn for this illustration from the characters of the sitcom *Friends*.

(4) Although using Wolfer's notion of 'acquired values', neither Baldwin nor Wolfers elaborate on what is meant by this phrase. Drawing from their context in International Relations, we presume this to be the values held by individual nation states.

Kevin Macnish

Assistant Professor in Ethics and IT, University of Twente

Jeroen van der Ham

National Cyber Security Centre, The Netherlands