WE NEED TO TALK – OPENING A DISCUSSION ABOUT ETHICS IN INFOSEC

Ivan Kwiatkowski
Kaspersky, France

ivan.kwiatkowski@kaspersky.com

ABSTRACT

Infosec is not like other jobs. We handle personal data, sensitive information, vulnerabilities that can affect thousands of computers. Our skills are sought after by the most powerful companies and governments. Yet we like to see ourselves as technologists; morally agnostic technicians who focus on solving virtual-world problems.

Reuters' recent article about UAE's Project Raven evoked strong reactions in many members of the community, myself included. It showed how infosec skills can be used to make the world a worse place – that, we already knew. But it also revealed the thought processes and motivations of the people involved. Looking back at the discussions our community has been having on social media in recent years, we can see that these justifications were already echoing:

- 'Everything I do is legal.'
- 'Exploits don't torture people. People torture people.'
- 'Morality is relative.'

I have witnessed several professionals defending the notion that technology and ethics have nothing to do with each other. I find this alarming as this vision might, in fact, be the reason why some of us, deprived of an established moral compass, end up getting lost. It doesn't have to be this way: generations of thinkers such as Aristotle, Kant and Rawls have been studying the concepts of right and wrong for centuries. In this talk, I will present various schools of thought pertaining to the philosophy of justice, and explore how they could help us solve some of the dilemmas the infosec community is facing.

PREAMBLE

It is widely recognized that technology and its omnipresence in our societies profoundly affects the way we lead our lives. As such, the nature of its impact is constantly questioned. Does technology empower users? Does it provide them with freedom and agency? Or is it a force that will be used against humanity to impose new, subtler forms of oppression?

Cybersecurity is found at a strategic crossroads between the numerous technology fields and is usually viewed as a tool to guarantee ethical outcomes: ensuring that data will not leak, that adversaries will not be able to disrupt operations, and so on. Being presented as a solution to ethical problems places cybersecurity in a strange blind spot where its own dilemmas might be overlooked. In fact, a number of practitioners in the field tend to envision themselves as morally agnostic problem

solvers. Suppliers of politically neutral code or information that will somehow make the world a better place. I do not believe that to be the case. Yet, ethics in the specific context of cybersecurity does not appear to be an established topic and the extent of the discussion on the matter seems to culminate with the 'white hat' / 'black hat' distinction. That Manichean perspective fails to encompass the complexities involved in the difficult problems that can be encountered in cybersecurity and perhaps only serves to illustrate how much a discussion on that very issue is warranted.

HOW DID WE GET HERE?

What chain of event leads an industry to offer 'ethical hacking' certifications that contain zero questions about ethics? The answer, in my opinion, is threefold.

- **The cybersecurity field is young.** It has only existed for a few decades and during that time has changed at such a rapid pace that it might not have had sufficient time for introspection. However, this reason does not stand up on its own as other IT fields which have started to expand even more recently (such as artificial intelligence) are acutely aware of the moral implications of their work.
- **The cybersecurity field lacks diversity.** The diversity problem in cybersecurity has been widely documented [1]. But beyond the obvious gender gap, it appears that the nature of the field attracts a certain demographic: people who have a certain disregard for authority or established rules. It might also be said that hacking in itself involves power dynamics that appeal to bigger egos. This (very crude) psychological profile does not paint the picture of a crowd that would have a lot of interest in submitting itself to an ethical framework. In addition, as it is a very specialized field, it is less common for people from other domains to join cybersecurity later in their careers, where they would bring their unique experience and outlook.
- **A lack of literary education.** Information security curriculums tend to focus exclusively on technical subjects and do not provide students with sufficient background in philosophy [2]. As such, they may only rely on their personal intuitions to tackle moral issues instead of benefiting from the works produced by generations of thinkers.

Consequently, this paper has a very modest scope. It does not propose a grand unified theory of justice applied to cybersecurity. Instead, it aims at acknowledging the moral dilemmas that exist in the field and present existing ethical frameworks that could provide answers – though not definitive ones – to some of them.

THE MORAL DILEMMAS OF CYBERSECURITY

Cybersecurity deals with power. Virtually every nation in the world has identified the strategic importance of the field and how it transforms the operation of the state. SIGINT has changed the way intelligence is collected. Cyber-sabotage and disinformation campaigns have already started to affect how war is waged. Since the Snowden revelations, it has become difficult to cling to the belief that cybersecurity work happens in a vacuum, that its product doesn't affect the foundations of our society. On a smaller scale, even companies and individuals who handle sensitive software or information may, purposefully or otherwise, act in a way that will have important repercussions for the rest of the world. Unethical or inadequate handling of those responsibilities may lead to the following risks:

- **Harm to privacy.** The primary role of cybersecurity is to safeguard the vast troves of information amassed by the modern ecosystem in which we live – ensuring confidentiality, integrity and availability. Failure to protect this information (or voluntary attempts to access it in an unauthorized fashion) violates individuals' right to a private life.
- **Harm to property.** As more and more aspects of our lives are handled in datacentres, it is not just information that is at risk of being stolen. There are few financial assets left that cannot, one way or another, be transferred digitally. Hackers, sometimes even assumed to be backed by nation states, have conducted sophisticated heists in the pursuit of monetary gain.
- **Harm to individuals.** Finally, in the worst-case scenario, cybersecurity failures can lead to actual physical harm or even death. An attack against *Saudi Aramco* [3] specifically tried to disable ICS safety systems to trigger an explosion in one of its plants. It is also rumoured that journalist Jamal Khashoggi's communications were compromised with the Pegasus malware before his assassination.

While there generally is not much debate about the importance of protecting people's privacy, property and physical integrity, there are a few situations where actors advocate for exceptions and dilemmas arise. This paper focuses on three of them, each tied to one of the risks described above.

Dilemma #1: legitimate hacking

Intelligence agencies and armies justify their hacking operations by appealing to the reason of state. They describe cyber capabilities as a weapon that they are required to use to protect the interests of the nation. Some countries have also set up a pervasive surveillance apparatus that is supposedly invaluable in preventing terrorism. Companies may also be tempted to resort to offensive actions, in the context of 'hack back' operations, for instance. Finally, some individuals, claiming the title of 'hacktivists', feel like their fight for a perceived greater good makes hacking acceptable.

Dilemma #2: vulnerability handling

The infosec community has yet to reach an agreement on the correct way to handle a software or hardware vulnerability one has discovered. Many options are available:

- Responsible disclosure: the researcher reports the vulnerability privately to the software vendor who has a chance to patch it before making an announcement. Critics of responsible disclosure argue that software vendors may not treat bug reports seriously unless a form of coercion is applied, for instance through the public pressure to provide a fix that would result from a public disclosure.
- Doing nothing: some believe that the process of disclosure leads to increased exploitation because of the attention brought to the vulnerability and the low rate of patch installations. They postulate that less overall harm will be caused if no one ever reports the bug.
- Selling the vulnerability to a broker: there is a lucrative market for software vulnerabilities that can allow researchers to profit from such discoveries. Their proponents would argue that researchers have no moral duty to fix other people's bugs and that they are entitled to the fruits of their work (in this case, the payout from an exploit vendor).

Dilemma #3: dual-use software

As was hinted by the mention of exploit brokers, a number of companies specialize in selling hacking tools or services. Their business consists of developing commercial-grade malware and discovering high-impact vulnerabilities that they sell to their customers. The few companies that will speak publicly on the matter swear that they fully comply with the laws of their respective countries and that they will only do business with democratically elected officials who have a good track record of protecting human rights. In practice, there have been documented cases where such companies have made suspect decisions in that regard [4].

BLANKET OBJECTIONS TO DISCUSSIONS ABOUT ETHICS

The presentation of the three dilemmas above broaches on the arguments that are sometimes used by members of the infosec community (some of them prestigious) to reject any discussion centred on ethics. In this section, I will examine them in detail and explain why they cannot by themselves address the moral issues at stake.

Morality and legality

It might be tempting for someone doing questionable work to find vindication in the idea that the particular legal framework surrounding them does not explicitly forbid their activity. This argument implies that the state condones whatever hasn't been criminalized. Furthermore, it follows that what is legal must also be moral.

Political philosophy discusses the organization of the state, where its power originates and the limits it should have. While there are many accounts on these matters, libertarian thinkers such as Robert Nozick advocate for a minimal state whose role is mostly to guarantee what they see as one of the most fundamental human rights: liberty. This ideology is rooted in the concept of self-ownership: I am the owner of my own body and should have full agency over what I choose to do with it. Therefore, they reject paternalist laws (e.g. being forced to wear a safety belt), taxation which they equate with state-organized forced labour, and believe that the government has no business dealing in morality. Libertarians would not, for instance, support laws that prevent same-sex marriage – as the subject would reside solidly outside the scope of the state. They not only reject the equivalence between morality and legality, but also assert that they should never overlap.

Seventeenth century philosopher John Locke offers another observation on the limits of the legislative powers of states. He puts forward the concept of 'unalienable rights', rights that are so fundamental that no government – not even a democratically elected one – can override. Even if there were support from the majority, he says, it would not be right to enact laws that violate people's rights to life, liberty or property. It follows that whatever is legal would not necessarily be moral (as some authoritarian states demonstrate) and what is moral is not necessary legal (same-sex marriage debates do a good job of illustrating that).

Moral relativism or nihilism

The example of same-sex marriage could prompt the following observation: morality seems to be evolving over time and is so deeply rooted in culture that attempts to define it are pointless. Different people might have significantly different perceptions of what morality consists of. What would be the worth of a moral framework that only applies to a few select individuals? Moreover, generations of

philosophical thinkers have worked on this very question but have failed to provide a definitive answer. Who are we to believe, one may ask, that we can succeed where our betters have not? The conclusion would be that this issue of ethics in cybersecurity should be discarded altogether, as it is nothing more than a simple matter of opinion, if not an unresolvable problem.

While it is true that different societies will each develop their unique moral code, it is interesting to note that all of them have one. This would tend to show that while the question of morality is unsolvable, it is also inescapable. Moral relativism also implies on some level that even the most heinous acts (such as murder or slavery) can be justified in some cultural context, which makes this a delicate position – though not an impossible one – to defend.

INCENTIVES TO VIOLATE ETHICS

These blanket objections out of the way, it is worth taking some time to reflect on the way infosec practitioners are led to engage in unethical behaviour. It is probably fair to assume that most members of the field would not accept an outright unethical job. Simone Margaritelli, author of bettercap, relates how a company in the United Arab Emirates tried to recruit him to work on massive-scale interception and intrusion capabilities in the country [5]. He turned down the offer. Yet a recent *Reuters* investigation revealed the inner workings of a private intelligence contractor who hired several former NSA members to spy on UAE political dissidents [6].

It would be easy to condemn these people for their actions. But it is perhaps more helpful to think about the possible ways that the infosec ecosystem has created the conditions to make these bad decisions possible. Intelligence agencies, which specialize in driving people to do things they normally would not, have a well-known persuasion framework called MICE. The acronym stands for money, ideology, compromise, and ego.

Reuters mentions that the analysts presented in the article earned more than $200,000 a year. Their managers could be paid twice as much. It is quite straightforward to see how *money* can be used to incentivize unethical conduct. The main appeal of exploit brokers are the enormous payouts they offer for premium vulnerabilities. Cybersecurity workers may be convinced to act unethically through ideology, if they are persuaded that they would contribute to some greater good that supersedes the amoral act. Common arguments include the fight against terrorism or appeal to a nationalist sentiment. *Ego*, to which our community is particularly susceptible, consists of driving someone through flattery. Less direct approaches entail appealing to the prestige of joining a highly selective team or organization, the opportunity to work on extremely complex technical challenges (such as organizing mass surveillance), or the offer to operate above (or with no regards to) the law. It is worth pointing out that ego does not have to be manipulated by third parties to lead to questionable behaviour: there are various examples of people releasing sensitive tools or information for social media credit. Finally, in the case of *compromise*, a *Bloomberg* report describes the extreme duress under which North Korean cybersecurity workers must acquire funds by any means necessary [7].

Opaque structures such as the one presented by *Reuters* or (one would assume) intelligence agencies will also withhold critical information from their employees through compartmentalization to make sure that they are unable to apprehend the unpalatable objectives they are serving unbeknownst to them. A subtler form of compromise would be to reveal to them gradually the unethical nature of their actions and threaten to expose their participation, or argue that they have already crossed the Rubicon.

Since cybersecurity deals with power, and members of the infosec community are the wielders of this power, they should be wary not only of themselves, but also of the numerous people who will wish to acquire it for their own agenda. This paper does not make the point that government work (or any other work for that matter) is inherently unethical. It does, however, offer that infosec practitioners should be familiar with manipulation techniques that might be used against them, regardless of the source, and have a clear idea of what they personally consider ethical practice.

THE THEORIES OF JUSTICE

This vision of what constitutes ethical practice does not have to be reimagined from scratch. In the last part of this paper, I would like to introduce some of the major theories from the philosophy of justice and how they could be applied to the dilemmas presented earlier.

Utilitarianism

The doctrine of utilitarianism was founded by Jeremy Bentham, an English moral philosopher from the 18th century. Its core idea is intuitively appealing: all creatures feel pleasure and pain, and are governed by the will to maximize the first while minimizing the latter. Therefore, the moral thing to do is to act in a way that will maximize utility for the community as a whole – i.e. that provides the most happiness or prevents the most pain. In other words, utilitarianism promotes the greatest good for the greatest number. It is a *consequentialist* theory in the sense that it places the moral worth of an act in what results from it.

Utilitarianism has many critiques, such as:

- It promotes sacrificing the few for the good of the many and generally doesn't recognize fundamental rights. For instance, torturing someone to obtain information about an imminent terrorist plot might be acceptable if it produced reliable intelligence. In fact, the utilitarianist theory might not even forbid torturing a relative of the suspected terrorist.
- The happiest society may not be the fairest at all. Utilitarianism focuses on maximizing happiness but ignores how it is spread.
- Consequences of our actions are difficult to predict.
- The utility calculus is complex in itself, as it involves aggregating elements that might not compare well or cannot be translated to a single unit. For instance, what would be the dollar price of a human life? Or, closer to cybersecurity, what is the dollar price of privacy?

John Stuart Mill addresses the first of these issues by arguing that respecting people's fundamental rights will always lead to the maximum utility, at least in the long run.

Applying the utilitarianist doctrine to the vulnerability handling dilemma illustrates the speculative nature of consequentialist reasoning. It is unclear how getting a vulnerability patched would exactly affect global utility. It seems that irresponsible disclosure (publishing an advisory with no prior coordination with the vendor) would lead to many computers being exploited before a patch can be released, which translates to sysadmin grief and lost business and therefore constitutes an undesirable outcome. Selling the exploit to a broker increases the happiness of the researcher, but is likely to lead to suffering for everyone it is used against. This course of action can only be justified if the researcher has sufficient faith that their vulnerability will only be used in a way that increases social utility, which in practice is unlikely or at least difficult to verify. Responsible disclosure is harder to

evaluate, as it will simultaneously lead to more people being protected by the vulnerability as well as more people being attacked with it. Depending on the severity of the vulnerability and the vendor's track record of handling security issues, it is probably right to report it privately. Utilitarian researchers who prefer to err on the safe side may opt to sit on their findings to avoid affecting overall happiness, provided they're confident that no one else will find the bug.

Transcendental idealism

If you believe that people have unalienable fundamental rights that are intrinsic (as opposed to basic rights that would be guaranteed because they benefit society as a whole), you may not be satisfied with the utilitarian account. Renowned philosopher Immanuel Kant observes that humans are gifted with free will. He recognizes the existence of pain and pleasure, but disputes that we are governed by them: because we have free will, we can refuse to act on our impulses and desires. Freedom, for Kant, means obeying laws that we have given ourselves, and morality resides solely in the decision to use this freedom for good. Contrary to utilitarianism, this theory has no interest in the consequences of actions. It is *categorical*. Actions are moral in themselves, regardless of whether they succeed or not. The focus is placed on intent: doing the right thing isn't enough, it must also be done for the right reasons. And the only right reason in the eyes of Kant is duty: we should act not for some perceived benefit or outcome but solely on the basis that it conforms to moral law, for morality's sake.

But the notion of duty may seem subjective. If we must act according to laws we've given ourselves, what should they be? Kant's answer is the 'categorical imperative', dictated by pure practical reason, for which he offers a number of formulations. One of them is the formula of humanity as an end: 'Act in such a way that you treat humanity, whether in your own person or in the person of any other, always at the same time as an end and never merely as means.'

To understand what Kant means, let's go back to the 'legitimate hacking' dilemma. Many countries have published doctrines outlining when they may resort to offensive cyber operations, and will admit to placing certain people under surveillance (for instance, terrorist suspects). Whether or not this ends up saving lives makes no difference to the categorical thinker: they would argue that violating an individual's privacy for any perceived gain amounts to using that person as a means to an end. Failing to respect someone's dignity, even a terrorist's dignity, by treating them as a mere way of producing an effect, no matter how desirable, constitutes a lapse in moral judgement. Hacking back would likely be rejected on the same grounds: it is neither doing the right thing, nor for the right reasons.

Is Kant's criteria for morality too stringent? French philosopher Benjamin Constant submitted the following challenge to him: what if you were hiding a friend, and a murderer knocked at your door looking for them. Wouldn't it be morally acceptable to lie to the murderer? Kant doubled down. It's not that the murderer is entitled to the truth, it's that lying in itself is directly at odds with the supreme principle of morality and as such is categorically wrong.

The veil of ignorance

While we are all bound by the laws of our countries, we have never explicitly agreed to them – we were simply born in a society which happens to possess a particular legal framework. It can be argued that choosing to remain in that society and benefiting from its rules constitutes tacit consent. Modern American philosopher John Rawls focuses on the idea of a social contract, and studies under

which conditions it could have moral strength. He points out that if all the members of a society gathered to determine the principles of justice, they would have a hard time agreeing. Even if they did, disparities in power and knowledge would likely skew the resulting contract in the favour of a select elite.

To ensure that the resulting contract would be fair, Rawls devised a thought experiment that assumes a position of equality between all participants. What if, he asks, everyone was placed behind a 'veil of ignorance' that obscures who they are, even from themselves? Beyond the veil, participants would not know their social status, ethnicity, education, and so on. Then and only then they would agree to fair terms.

Despotism would be rejected outright, as it would be obvious to most participants that they would not end up in a position of power. For the same reason, they would turn away from utilitarianism as there would be a chance for them to end up in a minority sacrificed for the greater good. Rawls' theory is that two principles would emerge in such conditions:

- That all citizens should have equal rights, such as freedom of speech and opinion, liberty, etc.
- That inequalities in society can be tolerated as long as they benefit the least well off.

For instance, under these principles, it is acceptable to heavily tax the richest members of society as long as the money is used to improve the living conditions of those who are poorer – libertarians strongly disagree with this position on the grounds that it violates a fundamental right to property. Critics of Rawls' theory also dispute that those two principles would necessarily emerge from this thought experiment. Still, the veil of ignorance is a tool that can help evaluate the moral qualities of a decision in an easy way. Behind the veil, would we consent to the free distribution of dual-use technology? How would we feel about the idea that some companies sell surveillance software to foreign governments? I believe that even the most risk-seeking gamblers would recognize that after the veil was lifted, they would find themselves on the wrong end of exploits more often than not. On a different level, it seems clear that surveillance capabilities overwhelmingly benefit those who are already in a position of power; so the injustices they cause (violation of privacy, chilling effects on free speech) would not be tolerable.

CONCLUSION

This paper has presented a few moral frameworks in the hopes that they will encourage curious readers to learn more about them. It is not meant as a formal or comprehensive account of any of them, nor does it prioritize them in any way.

In the introduction, I explained how I felt that the lack of literary education in our field creates an environment that facilitates unethical behaviour. What more can we do beyond modest attempts (such as this one) to compensate for this blind spot? I believe that it would behove our community to allocate more attention to the subject of ethics. Physicians, who also wield significant power in our society, have devised the Hippocratic Oath which serves as a reminder of the duties and obligations that come along with this power. Maybe it is time our profession adopted a global code of conduct as well. Established and respected organizations such as the Electronic Frontier Foundation in the US or the Chaos Computer Club in Europe could play a key role in pushing such standards.

Conference organizers have the power to guide the discourse inside the infosec community. In particular, it may not be beneficial anymore to reserve keynote slots for celebrities from the field.

Celebrities whose success, in some cases, may even be explained by ruthless business tactics or suspicious dealings. Conference organizers have the opportunity to correct our field's lack of diversity and natural tendency to operate in isolation by inviting speakers that not only belong to underprivileged minorities, but also to other communities. I believe it would be particularly enlightening to hear philosophers or victims of cyber-abuse discuss how they perceive our community, its shortcomings, and ways we could together become the change we want to see in the world.

REFERENCES

[1] Dallaway, E. Closing the Gender Gap in Cybersecurity. https://www.crest-approved.org/wp-content/uploads/CREST-Closing-the-Gender-Gap-in-Cyber-Security.pdf.

[2] Blanken-Webb, J.; Palmer, I.; Deshaies, S-E.; Burbules, N. C.; Campbell, R. H.; Bashir, M. A Case Study-based Cybersecurity Ethics Curriculum. https://www.usenix.org/system/files/conference/ase18/ase18-paper_blanken-webb.pdf.

[3] Groll, E. Cyberattack Targets Safety System at Saudi Aramco. https://foreignpolicy.com/2017/12/21/cyber-attack-targets-safety-system-at-saudi-aramco/.

[4] Schwartz, M. Cyberwar for Sale. New York Times. https://www.nytimes.com/2017/01/04/magazine/cyberwar-for-sale.html.

[5] Margaritelli, S. How the United Arab Emirates Intelligence Tried to Hire Me to Spy on Its People. https://www.evilsocket.net/2016/07/27/How-The-United-Arab-Emirates-Intelligence-Tried-to-Hire-me-to-Spy-on-its-People/.

[6] Schectman, J.; Bing, C. American hackers helped UAE spy on Al Jazeera chairman, BBC host. Reuters. https://www.reuters.com/investigates/special-report/usa-raven-media/.

[7] Kim, S. Inside North Korea's Hacker Army. Bloomberg. https://www.bloomberg.com/news/features/2018-02-07/inside-kim-jong-un-s-hacker-army.