Philosophy & Ethics of Risk

Better secure than sorry?
Assessing the quantum cybersecurity risk

Eline de Jong[1]

Abstract

In today's digitalised society, encryption plays a crucial role in securing sensitive data. However, the looming threat of quantum computers capable of breaking current encryption methods poses a significant cybersecurity risk, known as the "cryptopocalypse" or Q-Day. This paper examines the ethical justifiability of transitioning to quantum-safe cryptography (QSC) as a preemptive measure against this threat.

First, the distinction between risk and uncertainty is established, highlighting the decision to treat the quantum cybersecurity threat as a risk as value-laden. Applying Gardiner's Rawlsian Core Precautionary Principle, the paper argues that a precautionary response to the potential harm posed by cryptographically relevant quantum computers is ethically defensible, particularly given the potentially catastrophic consequences of a large-scale cybersecurity breach.

However, several ethical issues arise in the application of this principle in this particular case. These include a lack of awareness about the potential harm among stakeholders, uncertainty about responsibilities for addressing the threat, and concerns regarding the fair distribution of risks and benefits associated with a transition to QSC.

Further research is needed to explore these ethical issues in depth and promote an ethical application of QSC transition measures. Additionally, the paper suggests parallels between applying the precautionary principle in the context of emerging technologies like quantum computing and other domains, such as natural or ecological disasters, which share similar uncertainties and ethical considerations.

Key words

Quantum computing, cybersecurity, philosophy and ethics of risk, uncertainty, precautionary principle

[1] **Eline de Jong**, Institute for Logic, Language and Computation, Institute of Physics, Qusoft Research Center for Quantum Software, University of Amsterdam, Amsterdam, The Netherlands

Corresponding author: Eline de Jong, e.l.dejong@uva.nl

1. Introduction

Q-Day

In today's digitalised society, encryption is key. Everything we store or transmit online, from sensitive healthcare data to banking details, and from classified government data to personal communication, is secured by 'encryption': A way of scrambling (or: encoding) data beyond recognition, only to be descrambled by a party who is authorised to access the information and possesses the key to do so. Encryption is like putting a lock on your digital information or communication, so that it is not visible or accessible for everybody. The "lock" that we use to secure our information is basically a very hard, practically unsolvable mathematical problem; the "key" is inside information to find its solution.

The insolvability of the mathematical problems that underlie currently used encryption methods is what secures our digital data. It would take regular computers an unreasonable amount of time to solve a problem like factorisation (i.e. the reverse of multiplication), which is for that reason one of the most widely used bases for encryption. The promise of strong enough quantum computers is that these will be able solve such problems within a *reasonable* amount of time. 'Strong enough' here means quantum computers that have the capacity to run very complex algorithms. One example is Shor's algorithm, which is the key to solving factorisation problems. The day that scientists and engineers manage to create a cryptographically relevant quantum computer (CRQC) - i.e. quantum computer able to run algorithms that can break current encryption schemes – is called *Q-Day* or the "cryptopocalyps". By that time, quantum computers render current encryption obsolete and seriously compromise cybersecurity as we know it.

A call for action

Living in a world where data security is paramount, the looming cybersecurity threat of quantum computers is a critical concern. Activated by this threat, cryptographers, security specialists, and government authorities are working hard on the development of *quantum-safe cryptography* (QSC), i.e. encryption methods that can resist attacks from both classical and quantum computers. In accordance with these efforts, there is a much-heard call for action to transition to these new encryption schemes rather sooner than later. Organisations are urged to start preparing for the complex and time-consuming migration to quantum-safe security solutions, so that they can smoothly make the transition by the time that official standards become available.

At a first glance, this call for action seems to make perfect sense: Scientists and engineers are currently working on a technology that will, once realised, seriously compromise something that is highly important to all of us (i.e. cybersecurity), and so, we must anticipate this threat and take precautionary measures. Migrating to QSC solutions, and making the preparations to do so, is largely depicted as a no regret-measure. "Better secure, than sorry". At a second look, this risk-based call for action may not be as straightforward as it seems, considering that i) no (cryptographically relevant) quantum computers exist yet and given it is impossible to state with certainty when they will – if at all; and ii) there are other serious threats, such as those related to climate change, that are in terms of potential harm not inferior to the expected cybersecurity threat posed by quantum computers, but which nevertheless meet a far less proactive, widely supported response.

Outlook

In this paper, I will focus on the first consideration that highlights the risk-*assuming* - rather than risk-*based* - character of the call for a transition towards quantum-safe cryptography. I will scrutinise this call for action on the ground that the (epistemic) uncertainty about the development of quantum computers makes such a call *value-laden* (that is: not value-neutral), and hence requires ethical analysis: Notwithstanding its initial plausibility, how strong is the case for calling for a migration to QSC from an ethical point of view? To answer this question, I will first characterise the assumed risk as a *potential harm* that requires decision-making under *uncertainty* (Section 1). Next, I will evaluate the call for action by Gardiner's criteria for the application of the precautionary principle (Section 2). After arguing that precaution is defensible in this case, I explore the ethical issues of bringing it into practice in Section 3.

2. The quantum cybersecurity threat

Potential harm

The call for a migration to QSC is a response to the expected threat that future quantum computers pose to cybersecurity. A proactive migration is pictured as the solution for the risk of being 'too late': not migrating in time means that encrypted data will be no longer protected and thus at risk of being hacked. However, the response to this expected risk by proactive migration to QSC might also come with risks. Most salient is the risk of spilling resources in case a CRQC will not see the light of day anytime soon, or in case the threat changes because new quantum algorithms are being developed, making currently developed QSC standards outdated well before their implementation. Both cases show that, in addition to the risk of acting *too late*, there is also a risk associated with acting *too early*.

'Risk' in its most basic sense relates to an *unwanted event* which may or may not occur. In this sense, the coming of CRQCs qualifies as a risk: there is a chance of an unwanted event occurring. However, commonly, the term 'risk' is used to indicate the *probability* of the unwanted event occurring (Hansson, 2004). When this probability is not known, we speak of 'uncertainty'. In our case, we know that if a CRQC becomes available, cybersecurity will be compromised; but we lack the knowledge about whether or when this will happen. Because of this lack of knowledge about the probability of having a CRQC at time *t*, and so, the lack of knowledge about the probability of the unwanted event of a large-scale cybersecurity breach to occur, the expected threat in our case seems to disqualify as a (probability-based) risk. Rather, the case is marked by *uncertainty*. Hence, acknowledging the unknown probabilities at hand, it is more appropriate to speak about 'potential harm' or 'expected risk' here. A migration to QSC (or not) then counts as a *decision under uncertainty*.

Uncertainty reduction

To avoid that (epistemic) uncertainties prevent us from effective decision-making, we often reduce them to probabilities (Hansson, 2004). This "helps us to achieve a cognitively manageable representation of the world" (Hansson, 2004: p.12). In the call for a transition to QSC, the *unknown* probability – i.e. uncertainty - of the arrival of CRQCs is reduced to the belief that these will be available within a certain number of years. With this uncertainty reduction, the potential harm associated with CRQCs becomes a risk. Putting it differently, treating the potential cybersecurity threat as a risk is an act of belief, necessarily involving a degree of speculation.

As Hansson rightfully notes, the reduction of uncertainty is not a purely epistemic move. In a way, it is an ethical one as well. In the process of uncertainty reduction, we appeal to certain values - in our case, primarily to security: Because the stakes are high, we *decide* to treat the potential harm posed by CRQCs as if its probability is known. To some extent, marking a potential harm as a risk compares to accepting a hypothesis in science: It is not a pure observation, but a decision based on values, moral and scientific respectively. It is important to point out that such a decision is value-laden, because by making this explicit, it becomes relevant and possible to evaluate uncertainty reductions – in this case: treating the quantum cybersecurity threat as a risk – from an ethical point of view. As reducing the uncertainty of the quantum cybersecurity threat to a risk will have far-reaching implications for everyone who – to put it shortly - uses the Internet, it is important to critically assess the reasoning for a transition towards quantum-safe cryptography.

Unignorable disastrousness

Price (2022) offers an interesting perspective to think about uncertainty reduction in the context of risk. He takes the costs of potential false negatives and those of false positives to guide thinking about extreme technological risks. A *false positive* here would mean identifying a risk that is actually not present (also known as a type-I error), while a *false negative* would mean failing to identify a risk that is actually present (also known as a type-II error). In more plain language: a false positive is 'false alarm', while a false negative compares to 'false relaxation'.

Price states that: "The more disastrous a potential failure, the more improbable it needs to be before we can safely ignore it." (2022: p.14) In our context, this can be paraphrased as: the higher the potential harm, the less probability is needed to justify precautionary measures. Or in terms of false positive and false negatives: if the potential harm of an unwanted event is high, the costs of a false positive (i.e. the costs that come with preparing for an event that does not occur) will be outweighed by the costs of a false negative (i.e. the costs that come with not being prepared when the event does occur). Now, Price's argumentation is intuitively appealing, and his simple rule seems plausible – however, it does not provide much guidance about how to weigh the costs of a false negative against the costs that come with a false positive. Where lies the threshold for 'unignorable disastrousness'?

Rawlsian Core Precautionary Principle

Gardiner's (2005) account of the Core Precautionary Principle might give a useful answer to this question. The precautionary principle could be summarised as the 'better safe than sorry'-principle: when it is reasonable to suspect a risk but there is no conclusive evidence that this risk is real, the precautionary principle prescribes that we may (or should) take measures against the potential hazard (Hansson, 2022). So, potential harm and uncertainty are two main characteristics of cases where a precautionary response may be the most appropriate response. The basic idea behind the precautionary principle is that uncertainty should not be a reason for failing to act to prevent potential catastrophic harm. The principle thus offers a solution for making decisions under uncertainty, and often aligns with our ethical intuitions. However, as Gardiner (2005) rightly points out, crucial questions remain open: What size of the potential harm and which level of uncertainty would we find reasonable to trigger the precautionary principle?

According to Gardiner, there are several criteria of cases where precaution is warranted. These criteria can be taken as limiting conditions for the application of what he calls a "Core Precautionary Principle" (2005: p.2). To define these conditions, Gardiner draws on Rawls' maximin rule (i.e. 'maximise the minimum'), which was designed to be risk-averse and prioritise the worst-case scenario for the least advantaged. Rawls formulated three general conditions under which it is rational (that is: acceptable for everyone) to follow the maximin rule. Building on these conditions, Gardiner argues that the precautionary principle may be rightly applied under the general circumstances that:

1. decision-makers either lack, or have reason to sharply discount, information about the probabilities of the possible outcomes of their actions (*absence of reliable probabilities condition*);
2. decision-makers care relatively little for potential gains that might be made above the minimum that can be guaranteed by the maximin approach (*care little for gains condition*);
3. decision-makers face unacceptable alternatives (*unacceptable outcomes condition*).

Gardiner takes these Rawlsian conditions to constitute the application of the precautionary principle, "such that if those criteria are met, the precautionary principle demands action." (2005: p.15)

Assessing the quantum cybersecurity threat

Now, does the case of the quantum cybersecurity threat satisfy the Rawlsian criteria for a precautionary approach? The first condition implies that the precautionary principle only properly applies in the epistemic situation of *uncertainty*. As we concluded that the quantum cybersecurity threat is marked by uncertainty, the "absence of reliable probabilities" condition seems to be met. When considering the third condition, it seems reasonable to believe that the costs of seriously compromised cybersecurity are high and even catastrophic, which would satisfy the "unacceptable outcomes" condition. Whether or not the second condition is met is debatable. Indeed, the quantum threat case is more about *avoiding losses* than it is about gains. But what makes this case complex is that it 'avoiding a loss' could motivate both migrating (avoiding the loss when a CRQC becomes available) and not migrating to QSC (avoiding the loss in case of a false positive).

One could argue that the "care little for gains" condition is met since the costs of a precautionary response – in this case a migration to QSC – can be considered manageable or more manageable than the costs of large-scale cybersecurity breaches. This is to say that the costs of a false positive (type-I error/false alarm) are more acceptable than the costs of a false negative (type II-error/false relaxation). However, the uncertainty around the cause of the potential threat – i.e. the availability of a CRQC – interferes with the plausibility of this claim. If there is only a very insecure basis to identify a certain risk *and* if the measures to prevent or mitigate that risk are very costly and require significant efforts, the acceptance of a false positive becomes less of a no-regret option. How to escape a 'problem of paralysis' here?

Currently, no solid estimations are available of the costs associated with a QSC-migration. On the other hand, it is easy to imagine the devastating effects of seriously compromised cybersecurity systems. If we agree that the 'cryptopocalyps'-scenario is unacceptable, and if there are no compelling reasons to dismiss this scenario, then not knowing the costs of the QSC-migration might not matter. Put another way, if the scenario of a false negative is absolutely unacceptable, we are automatically pushed into accepting the scenario of a false positive – unless the latter would also be unconditionally unacceptable. But that is the matter: Since there *is* a scenario in which the costs of a QSC-migration are acceptable (namely: in case a CRQC becomes available), while there is *no* reasonable scenario in which we would find the consequences of seriously compromised cybersecurity acceptable, we have a reason to prefer a chance of a false positive over the chance of a false negative in this case. The asymmetry in acceptability of the two scenarios, i.e. the *conditional acceptability* of the migration costs versus the *unconditional unacceptability* of seriously compromised cybersecurity, substantiates the idea that a false positive in this case is preferable over a false negative, thus satisfying the second condition. After all then, if we fare on Gardiner's guidelines, a precautionary response to the potential harm posed by CRQCs is ethically defensible.

3. Exploration of ethical issues

Who is involved

Although applying the Rawlsian Core Precautionary Principle in the case of the quantum cybersecurity threat seems to be justifiable, it does not follow that its application is a *fait accompli*. To further think about the application of the principle in the form of a migration to QSC, it is necessary to understand *who is involved in which way*.

Hansson (2018) distinguishes between three fundamental forms of direct involvement that people have in a risky situation and their relationships: the risk-exposed, the beneficiaries (of a risk being taken), and decision-makers. The fact that the case under consideration here is strictly speaking not one of risk does not seem to affect the relevance of these groups. For Hansson the identification of the people that hold certain roles is the first step in what he calls an 'ethical risk assessment', but we can also take the identification of these three groups as a starting point for exploring potential hurdles that stand between an ethically justifiable precautionary measure (i.e. a transition to QSC) and its (ethical) application. Just like an ethical risk assessment should include an analysis of the people involved, so must an ethical application of the precautionary principle take different stakeholder groups into account. In the remainder of this section, potential issues that may stand in the way of an ethically proper application of the precautionary principle in the case of the quantum cybersecurity threat are tentatively explored.

Risk-exposed

As roughly everyone who uses the Internet will be subjected to seriously compromised cybersecurity incase strong enough quantum computers become available, everyone is potentially risk-exposed. Thus, it is in everyone's interest to avoid the potential cybersecurity risk posed by future quantum computers. However, not everyone might be *aware* of this potential harm, and those who have heard of it might not feel a sense of urgency to proactively respond to it. So before speaking about the *ethical* application of the precautionary principle, it is necessary to think about the hurdle that must be taken in order to apply the principle at all: *awareness*.

It is problematic if the potentially risk-exposed are not aware that they actually are in this group, because awareness of the fact that they are seems crucial to protecting their interests. In the slipstream of a lack of awareness then rises a question of responsibility: Whose responsibility is it to raise awareness of a potential threat? A preliminary answer would be: Those who *know* that there is a potential threat. When it comes to raising awareness there thus seems to be an important role for experts to get the message out there. But it is a challenge on its own to actually get the message across. Knowing *that* there is a potential threat might not automatically translate into a felt urgency, let alone into action. This brings us to the group of decision-makers: Who can and who should act?

Decision-makers

In the case of the quantum cybersecurity threat, it is not obvious who counts as a decision-maker: Should governments take the lead in addressing this threat by enforcing and coordinating a migration to QSC? Or do organisations have a duty to switch over to an available alternative that is at least as secure and potentially more secure than current systems (Nyholm, 2022)? Uncertainty about who belongs to the group of decision-makers reflects uncertainty about responsibilities. The current inaction when it comes to preparing for a large-scale migration to QSC and the waiting game that appears to be going on, might partly be the result of these uncertainties.

Besides uncertainty about responsibility, it might also be unclear what is needed to be an effective decision-maker in this case. If we, for example, think that a certain group has the responsibility to be a decision-maker, that does not make them *de facto* decision-makers for they might lack the abilities to effectively act. Concretely, some organisations may not have the resources to perform the migration to QSC. Moreover, if everyone is at risk but if some are dependent on others for addressing that risk (that is: everyone is risk-exposed but not everyone belongs to the group of decision-makers), an issue of fairness may rise: How can we make sure that everyone in the risk-exposed group is equally well protected? Will quantum-safe solutions be accessible to everyone in an equitable way? Who is responsible for seeing to this? Here, questions of responsibility merge into concerns about *fairness*.

Beneficiaries

Since everyone is potentially risk-exposed, it could be said that no one would unequivocally benefit from taking the risk of cybersecurity schemes becoming obsolete due to the availability of CRQCs. However, any party with 'hacking-motives' might be labeled as a beneficiary, as the absence of quantum-safe encryption would give them ample opportunities to access previously restricted information once quantum computing capacities become available. In a way, this group could be seen as the cause of the potential risk at hand. At the same time, this group is likely to belong to the group of the risk-exposed as well. Also, some actors may think that there is a bigger chance of benefitting from taking the risk of a false negative, than from taking the risk of a false positive, which would make them assumed beneficiaries of risking the potential harm posed by CRQCs. However, the discussion in Section 1 shows that this reasoning is flawed.

The perceived benefit from 'taking the risk' in this case is potentially harmful in the sense that it could obstruct an appropriate precautionary response. This again highlights the importance of what could be called true awareness, referring to a kind of genuine understanding of the matter at hand.

4. Conclusion and discussion

Large enough quantum computers will seriously compromise cybersecurity as we know it, and thereby, disrupt society dramatically. In response to this scenario, there is a widely supported call for action to migrate to quantum-safe cryptography (QSC): encryption methods that can withstand attacks from both classical and quantum computers. Although initially plausible, this call becomes less self-evident when considering that cryptographically relevant quantum computers (CRQCs) do not exist yet, and that it is impossible to predict when or whether they can be expected. In other words: Why should we prepare (that is: invest time and money) for an uncertain scenario like this? In this paper, the call for a transition to QSC was scrutinised for its ethical justifiability.

In Section 1, I concluded that the quantum cybersecurity threat should be taken as an *uncertainty* rather than a *risk*, for there are no known probabilities of the occurrence of the unwanted event, i.e. a severe cybersecurity breach. This difference is important in so far as it shows that it is a (value-laden) decision to treat this potential threat as a risk, making it appropriate to ethically assess this decision. To do so, it was discussed whether the application of the precautionary principle – in the form of a QSC-migration – can be justified in this case. Guided by Gardiner's Rawlsian Core Precautionary Principle and the conditions for its proper application, a precautionary response to the potential harm posed by CRQCs was judged as ethically defensible. This conclusion is congruent with Price's (2022) suggestion that if the potential harm of an unwanted event is sufficiently high, the costs of a false positive are to be preferred over the costs of a false negative.

In Section 2, it was explored what hurdles may stand in the way of an ethical application of the precautionary principle here. Drawing on Hansson's (2018) distinction between the risk-exposed, the decision-makers and the beneficiaries as the three fundamental groups involved in a situation of risk, three potential (ethical) issues were identified. First is a *lack of awareness* about the potential harm; second is *uncertainty about responsibilities* for addressing the potential harm; and third is a concern for an *unfair distribution* of the potential harm. Further research is needed to elaborate on these and other potential issues relating to a QSC-transition, in order to promote an ethical application of an ethically defensible measure.

Furthermore, this case showed how a precautionary principle could be applied to a technology that does not exist yet. It would be interesting to explore similarities and differences with the application of the precautionary principle in other cases, such natural or ecological disasters, which share the same inherent uncertainty that also marks the potential threats of a novel technology.

5. References

Gardiner, S. M. (2006). A core precautionary principle. *Journal of Political Philosophy*, *14*(1), 33-60.

Hansson, S. O. (2022). Risk. *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.).

Hansson, S. O. (2018). How to perform an ethical risk analysis (eRA). *Risk Analysis*, *38*(9), 1820-1829.

Hansson, S. O. (2004). Philosophical perspectives on risk. *Techné: Research in Philosophy and Technology*, *8*(1), 10-35.

Nyholm, S. (2022). The Ethics of Transitioning Toward a Driverless Future. *Test-Driving the Future: Autonomous Vehicles and the Ethics of Technological Change*, 59.

Price, H. (2022). Risk and scientific reputation: Lessons from cold fusion. arXiv:2201.03776

Rawls, J. (1999 [1971]). *A Theory of Justice*, Cambridge, MA: Harvard University Press. Revised edition.