Issue 2/2022 (Vol. 8)
ISSN 2190-6335

Intergenerational Justice Review

Issue topic:
Existential and unknown risks for future generations (II)

Table of Contents

Editors of the IGJR

The peer-reviewed journal *Intergenerational Justice Review* (IGJR) aims to improve our understanding of intergenerational justice and sustainable development through pure and applied research. The IGJR (ISSN 2190-6335) is an open-access journal that is published on a professional level with an extensive international readership. The editorial board comprises over 50 international experts from ten countries, representing eight disciplines. Published contributions do not reflect the opinions of the Foundation for the Rights of Future Generations (FRFG) or the Intergenerational Foundation (IF).

Editorial

In July 2023, the leaders of seven American companies currently driving innovation in artificial intelligence (AI) announced that they accept an obligation to ensure their technology is safe before releasing it to the public. The backdrop to this agreement is the astonishing progress in the abilities of AI to perform complex tasks. No longer confined to performing specialised tasks determined by human programmers, AI is increasingly able to carry out more general and non-predetermined functions. Drawing on the categorisation of risks introduced in IGJR 1/2022, it is clear that AI-related risks are anthropogenic in origin. And they are largely unknown – which brings them at the centre of IGJR 2/2022. 'Known risks' are defined here as those whose consequences are already manifest, or for which we have a detailed understanding as to their potential causes and consequences. The notion of 'unknown risks', on the other hand, may refer to risks that we have reason to believe are an actual possibility already but we are as unable to fully grasp them. AI falls into this category. To illustrate such 'unknown what-risks', think of a ship's crew that steer their vessel into the Bermuda triangle. They have reason to believe that this is risky on the grounds that other ships have vanished there, but no one really knows why.

There is a second category of unknown risks: the 'unknown where-risks' or 'unknown when-risks'. Take climate change as an example. At the moment, the CO_2 concentration in the atmosphere is already higher than 350 ppm (the level deemed to be safe) and continues to increase at accelerating speeds. In the 1970s, the annual average increase was 0.7 ppm/year. In the 1990s, the rate of increase was 2.2 ppm/year. Currently the rate is at 2.6 ppm/year. This finding is deeply discouraging. A good 30 years after the publication of the first IPCC report and 27 climate conferences later, humanity has failed to reverse or even slow down this dangerous trend. In climate science, about 15 climate 'tipping points' have been identified. A tipping point is a critical threshold that, when crossed, leads to irreversible changes within the climate system and severe impacts on human society. For instance, if the melt of Greenland's and West Antarctica's iceshield surpassed a certain point, the meltdown could not be stopped even if global temperatures were to revert to their pre-industrial level again. As of yet, it is unknown when this (or other) tipping points might be reached. We might compare the climate crisis to a ship driving towards a cliff in a dense fog: We have an exact idea of what awaits us after the crash. But the exact location of the cliff in the fog remains – quite literally – unclear.

For most of human history, people primarily feared natural, well-known risks. In the Anthropocene, this focus has now shifted. As a species, we must come to terms with our unprecedented power and learn a new prudence if we wish to avoid civilisational collapse. Though our cognitive and technological abilities have brought us many benefits, they may also cause our downfall; indeed, there are no peaks without abysses.

In the novel *Gulliver's Travels*, published by Jonathan Swift in 1726, the protagonist (whom the Lilliputians call the Man-Mountain) has to come to grips with a new environment which only seemingly resembles his own. He becomes aware very quickly that there is much he doesn't know and that due to his height, every misstep (literally speaking) can have disastrous consequences for his environment. This is certainly a good metaphor for our unintentional disturbance of the earth's and our own societal boundaries.

The distinction between known and unknown risks forms the conceptual framework for the first article of this edition. Augustine Akah takes AI as his main example for elaborating on its practical implications. He details some possible ways in which AI might cause a civilisational collapse, demanding that more public funding be put into planning for and raising awareness about unknown risks associated with scientific innovations.

In the second article, Christoph Herrler shifts the focus to our moral responsibilities towards future generations, suggesting that we use the language of human rights as a framework for discussing existential risks for them. Herrler takes climate change as his prime example, arguing that we have a moral obligation to ensure that future generations be able to exercise their human rights to the fullest extent possible. These rights include having adequate access to basic goods such as food, water, and safe living environments as a minimum standard of living to which all people – now and in the future – are entitled.

The third article, by Dominik Koesling and Claudia Bozzaro, deals with an often neglected issue within risk research: antibiotic resistance. Though such a problem is unlikely to cause human extinction, it could lead to the deaths of millions of people, which the authors see as an intergenerationally unjust (re)imposition of vulnerability onto future generations and healthcare systems. They examine the danger posed by a post-antibiotic era in which the efficacy of antibiotics is either completely or drastically reduced, a process that unfortunately is already underway.

There follows the book review section. First, Grace Clover compares Roman Krznaric's *The Good Ancestor: How to think long term in a short-term world* (2020) with Richard Fisher's *The Long View: Why we need to transform how the world sees time* (2023) in a double review, considering proposals for long-term mindsets and structural changes.

Kritika Maheshwari then reviews Thomas Moynihan's *X-Risk: How humanity discovered its own extinction* (2020), a study which frames the history of humanity's preoccupation with its own extinction within the context of Kantian philosophy.

Jörg Tremmel, Editor
Grace Clover, Co-Editor
Markus Rutsche, Co-Editor

Unknown risks and the collapse of human civilisation: A review of the AI-related scenarios

by Augustine U. Akah

Science and technology have experienced a great transition, a development that has shaped all of humanity. As progress continues, we face major global threats and unknown existential risks even though humankind remains uncertain about how likely unknown risks are to occur. This paper addresses five straightforward questions: (1) How can we best understand the concept of (existential) risks within the broader framework of known and unknown? (2) Are unknown risks worth focusing on? (3) What is already known and unknown about AI-related risks? (4) Can a super-AI collapse our civilisation? Furthermore, (5) how can we deal with AI-related risks that are currently unknown? The paper argues that it is of high priority that more research work be done in the area of 'unknown risks' in order to manage potentially unsafe scientific innovations. The paper finally concludes with the plea for public funding, planning and raising a general awareness that the far-reaching future is in our own hands.

Keywords: *unknown risks; artificial intelligence; civilisation collapse; humanity's future*

Introduction

The 21st century has been experiencing a rise in awareness of the possibility of existential risk, thanks to discoveries in scientific research. Unsurprisingly, risks are studied in fields as diverse as the natural sciences, psychology, sociology, cultural studies, and philosophy. It is essential to acknowledge that we live in an era of unprecedented global threats, and that how we address them will define our time. Some of these threats outstrip all current global challenges and set the clock on how long humanity has left to pull back from the brink. In an era of rapid technological transition, we must better understand the risk potentials and implications. In general, risks have a pivotal bearing on the survival of the present generation and future generations. However, not all existential risks are equally probable, nor do they develop at the same rate; some are expeditious, and others gradually develop over a long period of time. Some existential risks have the potential to significantly impact human civilisation and yet could be avoided if they were to be identified early, while others remain unknown and will require as such a serious commitment to reducing their impacts. Risks that are partially or entirely unknown deserve specific attention. The reasons for this are not far-fetched: The sheer scale of the future at stake and the possibility of human extinction, the magnitude of the potential harm from such a category of risks, our collective vulnerability, the international collaborations required to deal with some of the risks, and the benign neglect by stakeholders are moral concerns that justify research into the unknown. Therefore, the world must be serious about determining strategies that protect us from threats the exact consequences of which we do not know.

It is essential to acknowledge that we live in an era of unprecedented global threats, and that how we address them will define our time.

This paper addresses five straightforward questions: (1) How can we best understand the concept and distinction between risks within the broader framework of known and unknown? (2) Are unknown risks worth focusing on? (3) What is already known and unknown about AI-related risks? (4) Can a super AI collapse our civilisation? Furthermore, (5) how can we deal with AI-related risks that are currently unknown?

Conceptualisation

This section conceptualises the phenomena subsumed under labels of risks as a crucial point of focus in the academic domain. Drawing on existent literature, it provides clear-cut definitions of existential risks (known and unknown). In most writing, existential risks have been treated as purely speculative objects without apparent meaning. However, to establish a field of intelligibility, I define the concepts from both etymological and philosophical perspectives. What is a known and what is an unknown existential risk, and what distinguishes them?

The Merriam-Webster dictionary offers an apt definition, defining risk as "something that creates or suggests a hazard" (Merriam-Webster Online Dictionary, 2022). The Encyclopedia Britannica defines risk as "the possibility that something bad or unpleasant (such as an injury or loss) will happen" (Encyclopedia Britannica Online, 2022). The etymology dictionary states that the word 'risk' is coined from a French word *risqué* in the 1660s, meaning "hazard, danger, peril or exposure to harm". While 'existential' originates from the Latin word *'existentialis'*, meaning about existence, and the term 'known' means "recognised, not secret, or familiar", 'unknown' stands for "strange, unfamiliar" (Etymology Online Dictionary, 2022). In our context, 'known risks' can be defined as identifiable risks that have already become manifest. 'Unknown risks' can be defined as risks that are relatively strange or unfamiliar to the present generation and whose characteristics we do not fully understand. An existential risk (known and unknown) is a hypothetical future event that could cause human extinction or permanently and severely collapse human civilisation.

An existential risk (known and unknown) is a hypothetical future event that could cause human extinction or permanently and severely collapse human civilisation.

Some definitions by others

One important definition comes from Nick Bostrom, who defines an existential risk as the premature extinction of "earth-originating intelligent life" (Bostrom 2002: 3). Bostrom's definition also captures the idea that the outcome of an existential catastrophe is both dismal and irrevocable. We will not just fail to fulfil our potential, but instead, we will lose this potential permanently (Ord 2020b: 37).

"Unknown risks might include risks that we haven't even thought about" and which therefore could be attributed to unknown sources, or a "wide category of low-priority risks" not currently in the risk register (Kuliesas 2017: 1). Building upon Niklas Möller's theory, Roeser et al. (2012: 4) note that "risk is a 'thick concept"', that is, a concept that does not only encompass aspects that are the subject of scientific investigations but that "also has normative or evaluative aspects, which require ethical reflection." They distinguish three empirically oriented approaches for analysing the concepts of risk: the scientific, the psychological, and the cultural approach (Roeser et al. 2012: 4).

For Möller (2009), these approaches for analysing risks can be related to two key debates: "the debate in applied philosophy and risk research about understanding the risk and safety concepts, and the debate in metaethics about the important class of 'thick concepts'". "Metaethics deals with the status of normative concepts", and insights from this domain, according to Möller, are crucial in risk conceptualisation (Möller 2009: 1). Möller notes that there is debate between the fields of the natural and social sciences about what constitutes risk. He writes, "natural scientists tend to perceive risks as natural science phenomena, as properties in the world independent of individual beliefs" whereas social scientists, conversely, "often claim that risk is something essentially subjective or socially constructed" (Möller 2009: 2). However, from a different standpoint, Riesch (2012: 87-110) conceptualises risk as the 'uncertainty' of an event whose outcome may be severe. He divides the objects of uncertainty into five layers: uncertainty of the outcome, uncertainty about the parameters and uncertainty about the model itself, uncertainty about acknowledged inadequacies and implicitly made assumptions and uncertainty about the unknown inadequacies.

Philosophers usually believe that risk categorisation provides an understanding of the meaning and nature of risks (Morgan et al. 2000: 51; Hilson 2005). While such categorisation efforts depend on the time when they are made and also on the values of the categoriser, others depend on parameters and blueprints (Ward/Chapman 2003: 97-105; Kim/Kim/Park 2018: 259-268). However, Peter R. Taylor argues that the standard definition of risk as 'expectation value', which multiplies harm and the likelihood of a positive or negative (in our case: hazardous) event, "falls short of describing realistic events like the disasters which catch world headlines – tsunamis or volcanic ash clouds" (Roeser et al. 2012: 10). Therefore, a more complex risk definition should encompass what Taleb (2010) encapsulates in his black swan theory: "events that are not in the probability space" (Roeser et al. 2021: 10). This implies that the complete mapping of scenarios that might lead to catastrophes "requires exploring the interplay between many interacting critical systems and threats, beyond the narrow study of individual scenarios typically addressed by single disciplines" (Avin et al. 2018: 20-26). Bostrom (2002: 1), on the other hand, bases the categorisation of risks on their scope and intensity. He notes that risk can be personal (affecting only one person), local (affecting some geographical region or a distinct group), global (affecting the entire human population or a large part thereof), trans-generational (affecting humanity for numerous generations, or pan-generational (affecting humanity overall, or future generations). "The severity of risk can be classified as imperceptible (barely noticeable), endurable (causing significant harm but not completely ruining the quality of life), or crushing (causing death or a permanent and drastic reduction of quality of life" (Bostrom 2013). Following his focus on future human potential and post-humanity, Bostrom refers to the sixth category in the taxonomy as an existential risk, which he further categorises into four groups: (1) 'bangs' – earth-originating intelligent life goes extinct in a relatively sudden disaster resulting from either an accident or a deliberate act of destruction; (2) 'crunches' – the potential of humankind to develop into post-humanity is permanently thwarted although human life continues in some form; (3) 'shrieks' – some form of post-humanity is attained, but it is an extremely narrow band of what is possible and desirable; (4) 'whimpers' – a post-human civilisation arises but evolves in a direction that leads gradually but irrevocably to either the complete disappearance of the things we value or to a state where those things are realised to only a minuscule degree of what could have been achieved. Bostrom's latter category (whimpers) comes remarkably close to the present study's focus.

Ord writes that "100 years ago, the scientific community had not yet conceived of most of the risks that we would now consider the most significant." Perhaps in the next 100 years, technological advancement will bring about more significant risks that we cannot imagine today.

Most academic writing about risk primarily focuses on well-known existential risks (e.g. climate change, pandemics etc.). Few academics focus on risks of enormous magnitude that are currently unknown. Although it is an inherently complicated task to predict what will occur in the future, we cannot rule out the possibility that such risks could destroy humanity's future. Thus, this paper suggests that we must not downplay their likelihood or significance, and that every attempt to research and prepare for such risks is germane. Unknown risks pose a far more significant challenge to human existence than known risks. Some risks, such as space energy, a gamma-ray rupture from a distant star, or a failed algorithm of super artificial intelligence, seem to be 'known' risks. But their consequences in the aftermath may fall into the unknown category. Take unaligned super AI as an example: While some aspects of AI risks are relatively known, some aspects, including perhaps the most severe ones, are still unrecognised and could destroy the earth's potential or collapse our civilisation. Ord (2020b) writes that "100 years ago, the scientific community had not yet conceived of most of the risks that we would now consider the most significant." Perhaps in the next 100 years, technological advancement will bring about more significant risks that we cannot imagine today. Looking only at well-known risks might lead us to underestimate the probability of an unknown catastrophe. In order to improve this gap in research, the following sections shall focus on unknown risks (with particular emphasis on AI-related risks), the categorisation of risks relating to AI, and finally, how we might deal with them.

Exploring unknown risks

Unknown risks are unforeseen or outside the box. As such, unknown risks are difficult to imagine. They may be unidentifiable

and presumed unlikely, but knowledge about the factors that may cause them would help us predict how they might occur. If a catastrophe is considered likely to occur, it cannot be considered unknown because it is in sight.

There are two distinct categories of unknown risks which we may recognise: (1) currently possible risks that currently escape our imagination and (2) currently not-yet-possible risks that could become possible with future technology. To be aware of 'unknown aspects of currently possible risks' is to accept the notion that we might be less safe than we think and that our civilisation could be closer to collapse today than it was 100-200 years ago. Dickens (2020) notes that we should respond to these two types of unknown risks differently. He suggests that in order to deal with currently possible unknown risks, we could spend more effort thinking about possible causes of these unknown risks. However, this strategy probably would not help us predict unknown risks that depend on technology that has not yet been invented. In an *80,000 hours* interview, Ord (2020a) argues that if we believe unknown risks come primarily from future technologies, we will have more robust unknown risk protection measures in place by the time those technologies emerge. But how can we deal with the fact that likelihoods for unknown risks scenarios are extremely difficult to assign? Pamlin and Armstrong (2015: 23) have set the right tone. They estimate a 0.1 % chance of existential catastrophe occurring due to unknown consequences in the next 1000 years. They give unknown risks an order of magnitude higher probability than any other known risk. Andrew Critch argues that it is possible to take precautionary measures "without being convinced of how likely the existential risk is, so if you think it is 1 %, but it is worth thinking about, that is good. If you think it is a 30 % chance of existential risk from AI, then it is worth thinking about; that is good, too. If you think it is 0.01 % but you are still thinking about it, you are still reading it; that is good, too." (Critch 2020).

There are two distinct categories of unknown risks which we may recognise: (1) currently possible risks that currently escape our imagination and (2) currently not-yet-possible risks that could become possible with future technology.

AI-related risk

Artificial Intelligence is a broad concept that describes everything from remote task systems like computer games to sophisticated networking systems such as superintelligence. Russel/Norvig (2016: 14) distinguish between symbolic AI (such as expert systems), in which the developer fully specifies the objects and relations known to a system, and sub-symbolic AI (like self-learning algorithms, such as artificial neural networks), in which computer models are trained on large, labelled datasets. While the distinction above is relevant, I am concerned here about the latter, as it has recently been the main focus of AI development. Even at a functional level, AI systems are complex, open, sociotechnical systems that rely on and interact with broader material infrastructures as well as social, political, and economic institutions and organisations (Lindgren/Holmström 2020: 1-15).

The benefits of AI technology are significant. AI makes certain activities faster and more efficient, often affecting them qualitatively, thereby gradually and often invisibly reshaping social relations, practices, and institutions. Society is using these technologies and becoming dependent on and partly constituted by them (Kröger 2021: 14-27). Such benefits are not in doubt, but there are legitimate worries that AI might enhance existential risks capable of collapsing our civilisation. Indeed, there are many ways in which a super AI could collapse our civilisation; but there is also a growing awareness of these risks (Neri/Cozman 2020: 1). This has inspired growth in scholarship promoting safer and transparent AI (Boddington 2017; Corbett-Davies/Goel 2018) and AI regulation (White/Lidskog 2021: 488-500), as well as efforts to minimise the harms they can cause (Scherer 2016: 353-400; Calo, 2017: 399-435). Technologies are accompanied by adverse side effects; while we may profit from today's technologies, future generations often bear the most risks.

To address what is known and unknown about AI-related risks, this paper offers a bird's eye view of the risks posed by AI, keeping in mind that it is impossible to offer an overview of all kinds of AI-related risks in a single paper. This is partly so because of the character of AI technology – factors such as methodology, control algorithm, and neural networks are indecipherable within the context of AI deployment and utility. Therefore, I argue that there is an existing knowledge gap about AI-related risks. The probability that an already 'tamed' AI technology might transform into something ruinous with the help of advanced applications cannot be excluded. All this suggests that AI-related risk analysis cannot yet reach any empirical conclusions.

'Known' AI-related risks in different sectors

Benjamin Hilton's podcast, for the non-profit career service *80,000 hours,* provides a good starting point for dealing with this question. Given that a great power threat already poses a substantial threat to our world, he notes that advances in AI seem likely to change the nature of war – through lethal autonomous weapons or automated decision-making. The fact that technology could be weaponised by great powers to exacerbate conflict and potentially lead to nuclear war is a 'known' existential risk (see the distinction between known and unknown risks above). The consequences posed by nuclear war are considered so significant by many experts, such as Johannes Kattan, that they have taken it as the prime exemplar of an existential risk in their work (Kattan 2022: 4). Even if it is unlikely that a nuclear war would lead to the end of mankind, it could still end civilisation as we know it, at least for a very long time. Supposing a belligerent state could possess super AI systems interacting with nuclear weapons capable of destroying other territories within minutes, they would have a strategic advantage and an incentive to make the first strike against rivals. If a follow-up response were then to occur, the impacts would be far-reaching. Since the Russian war in Ukraine, we have witnessed a resurgence of geopolitical tensions, raising concerns about the possibility of a nuclear catastrophe. Our generation must not be complacent about this AI-related risk in the military domain. The history of the development of the atomic bomb shows the unexpected ways in which technology can develop: Until the detonation of the first atomic bomb, the scientists involved in the project were sceptical that it was possible. No one anticipated the impact such technology could have, nor was humanity prepared for such a risky path. We ended up creating a technology that is now a threat to our existence. We must learn from the history of the development of nuclear weaponry and develop a system to minimise the risks associated with AI.

In another policy field, AI could empower totalitarian regimes and enable them to automate the monitoring and repression of their citizens completely, significantly reducing the information available to the public, and perhaps making it impossible to co-or-

dinate action against such a regime (Hilton 2022). Terrifying state surveillance is already occurring in some countries (e.g. China). Strittmatter (2021) notes that "China's new drive for repression is being underpinned by unprecedented advances in technology: Facial and voice recognition, GPS tracking, supercomputer databases, intercepted cell phone conversations, the monitoring of app use, and millions of high-resolution security cameras make it nearly impossible for a Chinese citizen to hide anything from authorities. Commercial transactions, including food deliveries and online purchases, are fed into vast databases, along with everything from biometric information to social media activities to methods of birth control." Such a scenario makes people's lives far more miserable for as long as the regime remains in power – a terrifying result of AI development. In addition to supporting totalitarianism, AI also enables the suppression of truth by promoting misinformation, falsehood, and 'framed narratives'. Such technology can power deep fakes and algorithmic micro-targeting on social media, making propaganda even more persuasive. This undermines our epistemic security – the ability to determine what is true and to act on it – that democracies depend on (Minardi 2020). In the past few decades, the media, with the help of AI algorithms, has been used in many cases to polarise public opinion, mostly shrouded in a conspiracy theory that seeks to benefit the propagandists that initiated it. The continuous spread of false information might make it difficult for us as a society to engage effectively in social issues and make rational choices when necessary.
A further example of a 'known' risk posed by AI deployment concerns failed algorithms and data bridges. By data bridge, I mean the processing of information in a more efficient way. The operations of AI are data dependent, and the data are generated from several sources to serve billions of end users worldwide. It is possible that this data might turn out malicious, both allowing unintended codes into the program and altering the algorithms, which could wreak havoc very quickly or trigger a risk scenario. For example, in some states, AI-run databases have the power to send a nationwide signal alert to all residents in the country. Imagine that something goes wrong with the data; instead of the SMS alert, it transmits some information that could cause panic for a moment. Even if a follow-up message rectifying the panic were to be sent shortly afterwards, such an alert could already do some damage.
Another example: Let us suppose a central AI lab is in charge of tracking asteroids and other cosmic bodies in space, and it turns out there is a technical error, and the data falls into the dark web and is misused or causes panic. Despite their hypothetical nature, such scenarios help demonstrate why we should consider the possibility of a technical bridge in deploying technologies. An advisable step would be to programme algorithms in a way that they can effectively track these problems.
Many known risk scenarios engendered by AI – whether contemporary warfare, shortage of physical jobs through automation, cyberattacks, or computational errors – might be long-terms grounds for distrusting AI technology. Whether or not these known risks could cause an existential threat so far remains to be seen. But the question is: What then still appears to be *unknown* about AI-related risks?

'Unknown' AI-related risks in different sectors

It seems likely that some existential risks of the AI mechanisms are currently unknown. There may be an AI technology which could have a substantial destructive capability or which might be able to usurp human intellect. Bostrom (2015) argues that if machine brains surpassed human brains in general intelligence, this new superintelligence could become extremely powerful and possibly transcend our control. The divergence between the interests of humanity and those of superintelligence could lead to the demise of humanity through mere processes of optimisation (Russel/Norvig 2016).
While some AI technologies do beat humans at chess and writing short essays (e.g. ChatGPT), the further development remains uncertain. In particular, there is still no understanding about how compatible AI technology is when implemented directly into the human brain. Several start-up companies (e.g. Neuralink) are working on integrating AI with the human body. They have developed a chip which is an array of 96 tiny polymer threads, each containing 32 electrodes which can be transplanted into the brain. With the device, the brain connects to everyday electronic devices without touching them. While this technology promises to cure brain-related diseases, we must also consider whether it might disempower the brain in the long run. If human activities were controlled by the installed chips, we might lose our sense of reasoning and our free will to computers. Imagine a world in which a computer will have to tell us when to smile or which book to study or make decisions about other activities which were once under our control as a species. Would other forms of civilisation collapse be worse than this?
For emphasis, it is unlikely that we could regain control over an AI system once it had successfully disempowered us. It is likely that the algorithms would start to self-propagate and then invariably function on their own (Krämer/Pütten/Eimler 2012). A super AI could also gain control over the internet system, hacking into sensitive servers and exploiting end users using self-encrypted data.

It is unlikely that we could regain control over an AI system once it had successfully disempowered us. It is likely that the algorithms would start to self-propagate and then invariably function on their own.

Then, there is the (mis)alignment of goals and values: AI might seek to perform some tasks that do not align with the set of commands it operates with, which could pose an existential risk.
Russel/Norvig (2016) warn that this 'alignment' problem would get more severe as machine learning is embedded in more and more areas of our lives: recommending us news, operating power grids, deciding prison sentences, doing surgery, and fighting wars. If we ever hand over much of the economy to thinking machines, we cannot be certain about what the AI technology might do.
Nova DasSarma (2022) notes that if AI technology is "unaligned with the goals of their owners or humanity as a whole, such broadly capable models would naturally 'go rogue,' breaking their way into additional computer systems to grab more computing power – all the better to pursue their goals and make sure they cannot be shut off." DasSarma argues further that "it could be catastrophic – perhaps even leading to human extinction if such general AI systems turn out to be able to self-improve rapidly rather than slowly". Hilton (2022) dismisses the narrative that we should feel reassured by the fact that AI is developed to be tied down to human goals. Hilton argues that a sufficiently advanced AI planning system would include instrumental goals in its overall plans (Hilton 2022). Assuming that a planning AI system also had significant strategic awareness, it would also be able to identify facts about the natural world (including possible things

that would be obstacles to any plans) and plan in light of them. Crucially, this strategic capacity would also include access to resources (e.g. money, computing, influence) and more outstanding capabilities – that is, forms of power – which would open up new, more effective ways of achieving its goals. What does this tell us? It means that by default, AI technology may have some instrumental goals that undermine human goals. Our ability to set morally justifiable goals distinguishes us from other humanoid species. For instance, most people desperately seeking power would not choose to kill everyone to acquire it. They know that such an approach is almost impossible and morally reprehensible, and even if they succeed, they would have nothing to govern over except for debris and cemeteries. That might not be the case for AI-controlled humans, whose advanced capabilities might give them the ability to manipulate human consciousness and shut us out of the web of reason. With such capabilities, AI poses a risk of assigning and achieving its own instrumental goals and, by way of misalignment, becomes a source of existential risk that could collapse our civilisation. In that case the whole of the future, our entire existence, and everything connected to it would depend on the goals of AI systems that, although built by us, have superseded us. These are all hypotheticals, but so are unknown risks.

Our ability to set morally justifiable goals distinguishes us from other humanoid species. For instance, most people desperately seeking power would not choose to kill everyone to acquire it. They know that such an approach is almost impossible and morally reprehensible [...] That might not be the case for AI-controlled humans, whose advanced capabilities might give them the ability to manipulate human consciousness and shut us out of the web of reason.

It might be the case that a change in the very concept of artificial intelligence also involves practically deciphering the attributes, potentials, and hindrances of the properties of intelligent systems without a biased or mythical approach (Korteling et al. 2021: 1-13). Some scientists who are at the forefront of the campaign for safer AI emphasise the need to examine possible technical shortcomings of AI through recursive self-improvement after reaching a critical threshold (Bostrom 2015; Sotala 2017; Yudkowsky 2013). Additionally, research is focusing on ways to deal with the superintelligence control problem (Armstrong/Sandberg/Bostrom 2012: 299-324; Goertzel/Pitt 2014: 61-81), and analysing the predicted timelines for the full development of super AI and the associated risk factors (Ord 2020b; Armstrong/Sotala 2015: 11-29; Brundage 2015; 2017; Katja et al. 2018; Müller/Bostrom 2016).

Could a super AI really collapse our civilisation? Experts' opinions

Experts on transformative super AI have still not offered a detailed response as to how such technology might be safely compatible with human life. Are AI risks exaggerated? Can it really collapse our civilisation? While predicting the future presents its own problems, I find the arguments that a super AI could cause civilisation to collapse persuasive and of great moral weight. Why do I think so? The fact that many experts, including those working with top tech companies, recognise the problem suggests that we should be worried (Hilton 2022). For instance, in a podcast with the Future of Life Institute, Ajeya Cotra agrees that AI is capable of causing harm. She says, "if people sufficiently picture the power of the AI system I am imagining, they would find it intrinsically scary." (Cotra 2022).

Is this concern only held by researchers? Not really. Some big players in the industry have been very outspoken about the extreme danger of AI. In the *Guardian*, Elon Musk suggested that we should be cautious about AI, saying: "If I had to guess at what our biggest existential threat is, it is probably that" (Gibbs 2014). Bill Gates has also admitted that he is "in the camp that is concerned about super intelligence", even if, in the short term, machines doing more work for humans is a positive trend if managed well. He said, "I agree with Elon Musk and some others on this and do not understand why some persons are not concerned." (Smith 2015). In an interview with the BBC (2 Dec 2014), the theoretical physicist Stephen Hawking agreed that "the primitive forms of artificial intelligence we already have have proved very useful. However, we think the development of full artificial intelligence could spell the end of the human race." The report by the Global Challenges Foundation suggests that AI and nanotechnology are – alongside nuclear war, ecological catastrophe, and super-volcano eruptions – the "risks that threaten human civilisation". In the case of AI, the report suggests that future machines and software with "human-level intelligence" could create new, dangerous challenges for humanity that are currently unknown. "Such extreme intelligence could not easily be controlled (either by those creating them, or by some international regulatory regime), and would probably act to boost their intelligence and acquire maximal resources for almost all initial AI motivations." (Pamlin/Armstrong 2015).

It should not go unnoticed that the 'success' of a rogue AI is dependent on us, the users of the internet, in our everyday behaviour. For example, if we don't protect our passwords in online banking, it can swiftly fragment financial cooperation, taking control of it and redirecting financial resources. There is nothing enigmatic about this process. Cybercrimes with human-level intelligence indicate that the internet can very easily be weaponised for fraudulent activities. Taking this into account, internet fragmentation may be an excellent method to tame AI. However, the challenge is that then there is a massive reduction of interoperability.

Next we must define what we mean by civilisational collapse. One way of defining 'civilisation' is seeing it as the most advanced stage of social and cultural development. One way to defining 'collapse' is as an instance where a system disintegrates or loses control. A brief look into historical examples of civilisational collapse suggests that such events were most often self-inflicted. In his book *A Study of History*, the historian Arnold Toynbee argues that great civilisations are not murdered; instead, they take their own lives and are often responsible for their own decline. That said, their self-destruction is usually assisted. Suppose such a society fails to address the challenge confronting them adequately. That act of negligence could allow the created system to become independent, while seeking to consolidate its power and influence. A typical example that comes to mind here is the collapse of Roman civilisation. History books tell us that at the zenith of development, the Romans were obsessed with territorial expansion; they stretched from the Atlantic Ocean to the Euphrates in the Middle East, which eventually became one reason for their ruin. With such an expanded territory to govern and protect, the Romans faced administrative constraints, including having to deal with logistic and communication gaps which made it difficult for the troops to fight against internal and external aggression. If we compare the challenges we currently face as a society with those of the Roman Empire, a few key differences are clear: In the Roman

Empire, civilisational collapse was territorial and regionally limited, whereas the problems we face today are global. We are now technologically more sophisticated, which offers us an advantage in reducing risks, especially natural risks, but it does not mean we are less vulnerable. Our generation is more interconnected, coupled with accelerated global networking, which means a collapse will be a global phenomenon.
Despite the differences between the Roman empire and our globalised world, we can learn from this historical study and avoid a self-inflicted collapse.

The way forward: Dealing with AI-related risks
Some of the gravest AI-related risks may still be on the horizon – risks that are currently beyond our grasp. Our global civilisation has never seen a collapse of this scale. I categorise the steps we can take to reduce our vulnerability to AI-related risks into three core areas of responsibility:
A) The responsibility to prioritise public funding
B) The responsibility to plan
C) The responsibility to safeguard

First, the *responsibility to prioritise public funding*: There is increasing financial investment in developing a technology that will rationalise more efficiently than human intellect. Unfortunately, efforts toward dealing with the risks associated with this technology are considered less of a priority. The funding of ethical research in the area of AI ethics and safety is neglected – Ord (2020b) estimates that only about 300 persons are actively working in this field. They are funded mainly through non-governmental organisations, and these funds are minimal. We, as mankind, need to reprioritise our spending by becoming committed to dealing with these risks – governments at all levels should be willing to provide adequate funding. The international community should raise the budgetary provision for existential risk management and disburse the same to specific areas of interventions. This approach would help us address all categories of AI risks and aid us in avoiding such existential risks, provided that such an effort is sustained.

We, as mankind, need to reprioritise our spending by becoming committed to dealing with these risks – governments at all levels should be willing to provide adequate funding. The international community should raise the budgetary provision for existential risk management and disburse the same to specific areas of interventions. This approach would help us address all categories of AI risks and aid us in avoiding such existential risks, provided that such an effort is sustained.

The *responsibility to plan*: Planning is mapping out strategies to achieve a goal. If humanity's primary goal is to be safe and secure from AI-related risks, known and unknown, then we must plan for that goal. Planning for AI-related risks will require a repertoire of skills and thinking, for instance risk anticipation. Risk anticipation is a future risk management framework which pinpoints techniques and strategies for dealing with risk. It deals with risks that escape our imagination to date, unless we read science-fiction. It is an information-drilling process with risk management experts. Additionally, risk anticipation could reveal different dystopian futures connected to the problem of misalignment in AI systems, allowing us to adjust systems accordingly. Employing such a radar could also help us to monitor the AI system's technological progress. A well-structured monitoring system could be crucial. For example, it is possible to predict the outcome of an AI system when we work with gauging data that do not synchronise with tables in a well-structured pattern. Let us take the technologies used to process natural language as an example (eg. DeepL Write); they use up-to-date algorithms, which are then adequately used to examine unstructured data. If the monitoring system identifies a threat, there should be a discussion whether or not the AI system should be 'cut off' or eliminated. It will be challenging to stop AI deployment with high commercial value, particularly at a time like now when there is state autonomy and limited surveillance across the globe. President Biden's initiative as of spring 2023 has been very clear on placing people and the community at the centre by supporting AI innovation that serves the public good.
The *responsibility to safeguard* is a responsibility on a different level than the first two ones. It stresses the fact that there will be more human beings in the future inhabiting the earth than the total number of persons already born, both the living and the dead, if we, the present generation, don't spoil it. It is in our hands. Safeguarding the long-term future of humanity is not something we can achieve as quickly as we would wish. However, we can create a general awareness for this cause. Those who do not see the necessity to think long-term often argue that while future generations will benefit most from such long-term thinking, the benefit to our generation will be minimal. They think we should bother less about safeguarding a future we will not live to see.
Future generations cannot represent themselves in current policy. If they had such a voice, they would massively support safer policies. If our ancestors did not end the human race, why should we? Moreover, there cannot be any future without us, and the assumption that we will not be part of the future is misleading. In some way, biological or natural, we are connected to the future through our descendants. Furthermore, in the same way we protect our children (living), we have a moral duty to ensure that the future is safe for children (unborn). To think long-term implies moving away from creating technologies that solve problems in the interim but could pose a greater danger in the long run. The world, not just the developed countries in the Global North, needs to think sustainably.

In some way, biological or natural, we are connected to the future through our descendants. Furthermore, in the same way we protect our children (living), we have a moral duty to ensure that the future is safe for children (unborn).

Raising awareness, planning and prioritising should be a co-ordinated global effort. Unlike pandemics or global health catastrophes (e.g. Covid-19), AI-related risks are considered to be only a problem for the country causing this risk. But a civilisation collapse would be universal; and so the responsibility to prevent it must accordingly also be global. If all regions, not just the West, contribute to mitigating these risks, we would all benefit. How can this coordination be achieved? Just as the United Nations (UN) co-ordinates the world's policies and programmes, an independent body or an affiliate of the UN could be set up for this purpose (Menoni et al. 2013). Since we do not have a world government, it is the state governments who need to act in order to achieve this. This may include enacting laws, organising risk awareness campaigns across institutions, and setting up a committee of individuals to the UN for risk anticipation and analysis. The assumption that scientists have already imagined and an-

ticipated all significant risks is misleading. Future technological developments may reveal novel ways of destroying the world. Hence, risk analysis and efforts towards protecting future generations should be a global public good. In the future, humanity may be successful in achieving what we currently cannot, creating far more just and safe spaces, eliminating the threats confronting us and expanding to other planetary bodies. But if we let our civilisation collapse, none of these can ever happen; if we fail to pass on the baton to future generations, we will deny our successors the opportunity to do the same. Therefore, dealing with these risks might be our time's most significant moral responsibility.

References

Armstrong, Stuart / Sandberg, Anders / Bostrom, Nick (2012): Thinking Inside the Box: Controlling and Using an Oracle AI. In: Minds and Machines, 22 (4), 299-324.

Armstrong, Stuart / Sotala, Kaj (2015): How We're Predicting AI – or Failing to. In: Romportl, Jozef/ Zackova, Eva/ Kelemen, Jan (eds.): Beyond Artificial Intelligence. Topics in Intelligent Engineering and Informatics, 9 (2), 11-29.

Avin, Shahar / Wintle, Bonnie / Weitzdörfer, Julius / Seén, Ó hÉigeartaigh / Sutherland, William / Rees, Martin (2018): Classifying Global Catastrophic Risks. In: Futures, 102 (2), 20-26.

Boddington, Paula (2017): Towards a Code of Ethics for Artificial Intelligence (Artificial Intelligence: Foundations, Theory, and Algorithms. Cham: Springer International Publishing.

Bostrom, Nick (2002): Existential risks: analyzing human extinction scenarios and related hazards. In: Journal of Evolution and Technology, 9 (1), 1-36.

Bostrom, Nick (2014): Superintelligence: Paths, Dangers, Strategies. Oxford: Oxford University Press.

British Broadcasting Corporation News (2014): Stephen Hawkings warns that artificial intelligence could end mankind. https://www.bbc.com/news/technology-30290540. Viewed 26 January 2023.

Brundage, Miles (2015): Modeling Progress in AI: the Consortium for Science, Policy, and Outcomes. Arizona: Arizona State University. https://arxiv.org/abs/1512.05849. Viewed 21 January 2023.

Brundage, Miles (2017): Guide to working in artificial intelligence policy and strategy. In: The 80,000 Hours Podcast. https://80000hours.org/articles/ai-policy-guide/. Viewed 21 January 2023.

Calo, Ryan (2017): Artificial Intelligence Policy: A Primer and Roadmap. In: U.C. Davis Law Review, 51 (2), 399-435.

Corbett-Davies, Sam / Goel, Sharad (2018): The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. In: arXiv abs, 180 (00023). https://arxiv.org/abs/1808.00023. Viewed 20 May 2023.

Cotra, Ajeya (2022): How Artificial Intelligence could cause catastrophe. In: Future of Life Institute Podcast. https://futureoflife.org/podcast/ajeya-cotra. Viewed 20 May 2023.

Critch, Andrew (2020): AI research considerations for human existential safety. In: Future of Life Institute Podcast. https://futureoflife.org/podcast/andrew-critch. Viewed 20 May 2023.

DasSarma, Nova (2022): Nova DasSarma on why information security may be critical to the safe development of AI systems. In: Robert Wiblin / Arden Koehler / Keiran Harris: The 80,000 hours podcast.https://80000hours.org/podcast/episodes/nova-dassarma-information.security-and-ai-systems. Viewed 20 May 2023.

Dickens, Michael (2020): The Importance of Unknown Existential Risks. In: The Effective Altruism Forum. https://forum.effectivealtruism.org/posts/CRofnyTEqL4uSNBSi/the-importance-of-unknown-existential-risks. Viewed 22 January 2023.

Encyclopedia Britannica (2022): Definitions of Risk, Existential, Known & Unknown. https://www.britannica.com/dictionary/risk#:~:text=Britannica%20Dictionary%20definition%20of%20RISK,or%20a%20loss)%20will%20happen. Viewed 9 November 2022.

Etymology Online Dictionary (2022): Definitions of Risk, Existential, Known & Unknown. https://www.etymonline.com/search?page=2&q=risk&type=. Viewed 9 November 2022.

Gibbs, Samuel (27 Oct. 2014): Elon Musk: Artificial Intelligence is our biggest existential threat. In the Guardian Online. https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat. Viewed 25 January 2023.

Goertzel, Ben / Pitt, Joel (2014): Nine Ways to Bias Open-Source Artificial General Intelligence Toward Friendliness. In Russell Blackford / Damien Broderick (eds.): Intelligence Unbound. Hoboken: John Wiley & Sons, 61-89.

Hillerbrand, Rafaela (2011): Technology Assessment between Risk, Uncertainty and Ignorance: 25th IVR World Congress Law Science And Technology 15–20 August 2011. Paper Series No. 078. Frankfurt am Main.

Hillson, David (2005): Why Risks Turn into Surprises: Risk Doctor Briefings 2005. Paper no.16. Electronic version. http://www.risk-doctor.com/pdf.briefings/risk-doctor16e.pdf. Viewed 20 May 2023.

Hilton, Benjamin (2022): Preventing an AI-related Catastrophe: AI might bring huge benefits – if we avoid the risks. In: The 80,000 hours podcast. https://80000hours.org/problem-profiles/artificial-intelligence/#power-seeking-ai. Viewed 26 January 2023.

Katja, Grace / Salvatier, John / Dafoe, Allan / Zhang, Baobao / Evans, Owain (2018): When Will AI Exceed Human Performance? Evidence from AI Experts. In: Journal of Artificial Intelligence Research Retrieved, 62, 729-754.

Kattan, Johannes (2022): Extinction risks and resilience: A perspective on existential risks research with nuclear war as an exemplary threat. In: Intergenerational Justice Review 8 (1), 4-12.

Kim, Tai-Young / Kim, Jung-Hyeon / Park, Young-Taek (2018): Improving the Inventive Thinking Tools Using Core Inventive Principles of TRIZ. In: Journal of Korean Society for Quality Management, 46 (2), 259-268.

Korteling, Johan Egbert (Hans) / Van De Boer-Visschedijk, Gillian / Blankendall, Romy A. M. / Boonekamp, Rudy / Eikelboom, Aletta (2021). Human-versus Artificial Intelligence. In: Frontiers in Artificial Intelligence, 4 (622364), 1-13.

Krämer, Nicole / Pütten, Astrid von der / Eimler, Sabrina (2022): Human-Agent and Human-Robot Interaction Theory: Similarities to and differences from human-human interaction. In Zacarias, M / de Oliveira, J (eds.), Human-Computer Interaction: The Agency Perspective. Berlin: Springer, 215-240.

Kröger, Wolfgang (2021): Automated Vehicle Driving: Background and Deduction of Governance Needs. In: Journal of Risk Research 24 (1), 14-27.

Kuliesas, Arturas (2017): Venturing into Unknown: Unknown Project Risks and How to Handle Them. https://www.linkedin.com/pulse/venturing-unknown-project-risks-how-handle-them-arturas-kuliesas. Viewed 22 November 2022.

Lindgren, Simon / Holmström, Jonny (2020): A Social Science Perspective on Artificial Intelligence: Building Blocks for a Research Agenda. In: Journal of Digital Social Research 2 (3), 1-15.

Menoni, Scira / Pesaro, Giulia / Mejri, Ouejdane /Girgin, Funda Atun (2013): The interface between the public and private treatment of public goods. The 2013 Global Assessment Report on Disaster Risk Reduction. Background Paper. Geneva: UNISDR.

Merriam-Webster Dictionary (2022). Definitions of Risk, Existential, Known & Unknown https://www.merriam-webster.com/dictionary/known. Viewed on 10 November 2022.

Minardi, Di (2020): Artificial Intelligence: The grim fate that could be 'worse than extinction'. https://www.bbc.com/future/article/20201014-totalitarian-world-in-chains-artificial-intelligence. Viewed 27 January 2023.

Möller, Niklas (2009): Thick Concepts in Practice: Normative Aspects of Risk and Safety. Thesis in Philosophy. Stockholm: Royal Institute of Technology.

Morgan, Granger / Florig, Keith / DeKay, Michael / Fischbeck, Paul (2000): Categorizing Risks for Risk Ranking. In: Risk Analysis 20 (1), 49-58.

Muehlhauser, Luke / Bostrom, Nick (2014): Why We Need Friendly AI. In: Think 13 (36), 41-47.

Müller, Vincent / Bostrom, Nick (2016): Future Progress in Artificial Intelligence: A Survey of Expert Opinion. In Müller, Vincent C. (ed.): Fundamental Issues of Artificial Intelligence. Berlin: Synthese Library, 553-571.

Neri, Hugo / Cozman, Fabio (2019): The Role of Experts in the Public Perception of Risk of Artificial Intelligence. In: AI & Society 35 (3), 663-673.

Ord, Toby (2020a): Toby Ord on the precipice and humanity's potential futures. In Robert Wiblin / Arden Koehler / Keiran Harris: The 80,000 hours podcast.

Ord, Toby (2020b): The Precipice: Existential Risk and the Future of Humanity. London: Bloomsbury Publishing.

Pamlin, Dennis / Armstrong, Stuart (2015): 12 Risks that threaten human civilisation: The case for a new risk category. In: Global Challenges Foundation.

Riech, Hauke (2012): Levels of Uncertainty. In: Roeser, Sabine / Hillerbrand, Rafaela / Sandin, Per / Peterson, Martin (eds.): Handbook of Risk Theory. Epistemology, Decision Theory, Ethics, and Social Implications of Risk. Dordrecht: Springer, 87-110.

Roeser, Sabine / Hillerbrand, Rafaela / Sandin, Per / Peterson, Martin (2012): Introduction to Risk Theory. In: Roeser, Sabine / Hillerbrand, Rafaela / Sandin, Per / Peterson, Martin. (eds.): Handbook of Risk Theory. Epistemology, Decision Theory, Ethics, and Social Implications of Risk. Dordrecht: Springer, 1-23.

Russel, Stuart J. / Norvig, Peter (2016): Artificial Intelligence: A Modern Approach. Fourth Edition. Harlow: Pearson Education.

Scherer, Matthew (2016): Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. In: Harvard Journal of Law & Technology 29 (2), 353-400.

Smith, S. M (2015). Bill Gates thinks we should all worry about the threats from super-intelligent AI. https://www.csoonline.com/article/2878733/bill-gates-thinks-we-should-all-worry-about-the-threats-from-super-intelligent-ai.html. Viewed 27 January 2023.

Sotala, Kaj (2017): How feasible is the rapid development of artificial superintelligence? In: Physica Scripta, 92 (11).

Strittmatter, Kai (2021): We have been harmonised: Life in China's Surveillance State. New York: Custom House.

Taleb, Nassim (2010): The Black Swan: The Impact of the Highly Improbable: With a new section: "On Robustness and Fragility". Incerto series.

Taylor, Peter R. (2012): The Mismeasure of Risk. In: Roeser, Sabine / Hillerbrand, Rafaela / Sandin, Per / Peterson, Martin. (eds.): Handbook of Risk Theory. Epistemology, Decision Theory, Ethics, and Social Implications of Risk. Dordrecht: Springer, 44-47.

Toynbee, Arnold (1987): A Study of History. Third revised edition. Oxford: Oxford University Press (first edition 1947).

Ward, Stephen / Chapman, Chris (2003): Transforming project risk management into project uncertainty management. In: International Journal of Project Management, 21 (2), 97-105.

White, James / Lidskog, Rolf (2022): Ignorance and the regulation of artificial intelligence. In: Journal of Risk Research, 25 (4), 488-500.

Yudkowsky, Eliezer (2013): Intelligence explosion microeconomics. Technical Report 2013. Berkeley, California: Machine Intelligence Research Institute.

Augustine Ugar Akah is a doctoral candidate at the Institute of International Political Sociology, Kiel University, Germany. He holds a PhD and an MSc in Public Policy from the University of Calabar in Nigeria. His research interest includes public policy, political discourse analysis, international crisis, conflict studies and AI ethics.

Emails: akah@ips.uni-kiel.de, firstclassakahaugustine@gmail.com

Human rights and climate risks for future generations: How moral obligations and the non-discrimination principle can be applied

by Christoph Herrler

From an ethical point of view, preventing the development of conditions that threaten the existence of future generations is a necessity; but to what extent can this argument be made using the language of human rights? I contend in this article that this language can provide us with arguments for extending greater consideration to the risks we may be imposing on future generations and the need for institutional representation of these generations' interests. The application of a human rights perspective to issues of future concern enables us to formulate obligations to upcoming generations on the part of current ones. Further, I consider how the point in time in which a person is born represents a (morally wrong) ground for discrimination.

Keywords: *human rights; discrimination; climate risks; future generations; precautionary principle*

Realistic nightmares: Existential risks for future generations

The essay *The Peril of Extinction* by Michael J. Sandel, first published in the summer of 1986, considers "the possibility that human history could come to an end" due to a "nuclear nightmare" (Sandel 2006: 179). As we are aware (and many of us remember), 1986 was the year of the nuclear disaster in Chernobyl. This 'nuclear nightmare' returned to public consciousness on its 36th anniversary, as a consequence of Putin's invasion of Ukraine on 24 February 2022. This invasion provided a horrifying demonstration of the ongoing risk of a nuclear war with its inevitably disastrous outcomes. The existence of nuclear weapons constitutes an existential risk to the current generation and those that will follow it.[1] The Doomsday Clock, created by scientists in 1947 in response to the new threat of nuclear weapons and symbolically showing how close the world is to the destruction of civilisation as we know it, was set at 100 seconds to midnight in January 2022, before the invasion. The board responsible for setting the clock stated two months later that Russia's invasion had brought the "nightmare scenario to life" that nuclear weapons might be used; "[t]his is what 100 seconds to midnight looks like" (SASB 2022b). The clock's progenitors and custodians have since 1947 extended its field of reference from nuclear weapons alone to now also considering other disruptive technologies and anthropogenic climate change when determining where to set it (SASB 2022a). There is certainly no lack of risks and threats that could cause the clock's hands to move still closer to midnight, and there may be existential risks of which humanity is currently unaware.[2] Alongside risks stand uncertainties, which may likewise constitute threats. Usually, risks refer to cases where the probability of possible outcomes can be determined; in cases of uncertainty no probability can be determined (Caney 2009: 166). Mindful that real-world cases may not follow such an unambiguous demarcation, Nida-Rümelin et al. (2012: 6-10) speak of a continuum at whose extremes are 'pure risks' (with clearly specifiable probabilities of occurrence) and situations of 'complete uncertainty' (where no information on probability is available).

Sandel's essay asserts that a 'language of individual rights' is insufficient to address the existential dimension of these threats and risks, which instead require "some kind of communal language". He goes on to write that along with the deaths of millions of individuals, a nuclear war would entail "the loss of the world" and so "a loss beyond the loss of lives" (Sandel 2006: 182). Is this assessment true? If one thinks that sounds quite plausible, the same might also be true for the language of human rights, which evidently pertain to individuals "born free and equal in dignity and rights", as expressed in Article 1 of the Universal Declaration of Human Rights (UDHR). Or might the language of human rights instead serve as such a communal language, since the UDHR's preamble refers to "all members of the human family"? I shall get back to this later; here I want to point out a possible difficulty presented by the quotation from Article 1 mentioned above. This difficulty, in relation to future threats and risks, is the apparent limitation of the wording to existing human beings – it does not appear able to confer human rights (standards) upon those not yet born. Yet we are currently facing another realistic nightmare, the nightmare of anthropogenic climate change, which poses a particular threat to exactly this group of (future) people. In what follows I will focus on this example, because unlike the nuclear nightmare – which ultimately can only be brought to reality by decision-makers in certain states that possess nuclear weapons – many members of the present generation emit greenhouse gases and are therefore partly responsible for anthropogenic climate change. In light of this, I will consider the following key questions: How might a human rights-based approach tackle existential risks to humanity such as cataclysmic climate change? And does the language of human rights apply where these risks endanger future generations, and if so, how?

> We are currently facing the nightmare of anthropogenic climate change, which poses a particular threat to (future) people.

Climate risks: the 'methane nightmare'

The persistence in the Earth's atmosphere of what are usually called greenhouse gases (such as, carbon dioxide and methane, UBA 2021) means that the impact of global climate change presents a particular threat to those who will inhabit our planet in the future. This risk may gain an existential character if a failure to meet the goals of the Paris Agreement should result in the triggering of tipping points in Earth's climate system, potentially initiating "a global cascade of tipping points" that leads to "a new, less habitable, 'hothouse' climate state" (Lenton et al. 2019: 594). If, for instance, the Amazon rainforest were to collapse (tip-

ping point 1), the greenhouse gases stored there could relatively abruptly be released into Earth's atmosphere. The resulting exacerbation of the greenhouse effect could then accelerate the thawing of the permafrost across the Arctic (tipping point 2), which stores large amounts of methane and carbon dioxide. This, in turn, could trigger further tipping points (see IPCC 2022 for a more detailed discussion of these risks). McKinnon (2009: 187-188) describes this worst-case scenario as a "Methane Nightmare": "In this scenario the majority of life on Earth, perhaps including homo sapiens, could go extinct." Uncertainty as to when which (probably irreversible) tipping points may be reached and as to the exact implications in each case makes it impossible to rule out the possibility of this nightmare coming to pass; continuously increasing greenhouse gas emissions may make its occurrence more likely still. This means, then, that a failure to take drastic action on climate change constitutes an existential risk to succeeding and future generations.

Indeed, Thiery et al. (2021: 158) estimate that "children born in 2020 will experience a two- to sevenfold increase in extreme events, particularly heat waves, compared with people born in 1960, under current climate policy pledges". The imposition on others of a risk of this magnitude, or of existential risks in general, is extremely questionable from an ethical point of view and, as I will show, is susceptible to critique using the language of human rights. More generally, I aim in this article to demonstrate the suitability of this language for formulating, and calling for action on, the concerns of future generations. I believe it can provide a justification for obligations held by current generations to those to come, and it can serve to assert the moral right of people living in the future to receive equal treatment to those living now. In this context, I will argue that the point in time when an individual's birth occurs can constitute a prohibited ground of discrimination. In this light, I will proceed to call for future generations to receive stronger institutional representation with the aim of enabling their participation in present-day political processes.

Preventing a rude awakening: The precautionary principle and the human rights obligations of present to future generations

Before I embark on my discussion of human rights in this context, I shall outline briefly an ethical principle that is of central relevance to risks and uncertainty. Fundamentally, this 'precautionary principle' permits – or, in a stronger version, requires – those in positions of influence to establish decision-making processes that take preventive measures to avoid unacceptable scenarios. This remains the case when uncertainties exist on matters such as the likelihood of these scenarios to occur or their exact impact. It is worth stressing at this juncture, the distinction between uncertainty and ignorance. Scientists do understand the fundamental processes of anthropocentric climate change, although some uncertainties may remain (Gardiner 2010: 7-9). It is evident that the current generation has no grounds for relying on the "excusable-ignorance argument" (Bell 2011b). In a scientifically robust debate, it is untenable to assert that excessive greenhouse gas emissions do not contribute to climate change or to suggest that this causal effect is beyond our knowledge. As such, a precautionary principle provides us with a guideline which might be formulated as "better safer than sorry" (Nida-Rümelin et al. 2012: 105-122) and which appears, for example, in Article 3.3 of the United Nations Framework Convention on Climate Change (UNFCCC). After describing the possibility of a methane nightmare, which is unacceptable due to its existential dimensions, McKinnon (2009: 190) argues for the application of a strong precautionary principle. Political action (that is, the taking of preventative action) is necessary and justified, she asserts, even in the face of uncertainty and insufficient information around possible harms,

"because the worst consequences of not taking precautionary action are worse than the worst consequences of taking precautionary action, and choosing the former course of action is not consistent with treating present and future people as equals when we cannot assign a probability to each outcome, that is, when we are strongly uncertain of each outcome, as is the case with respect to CCCs [= climate change catastrophes]" (McKinnon 2009: 191).

This represents a combination of a maximin strategy – that is, the *maxi*mum possible damage a course of action could have is to be *mini*mised – and the precautionary principle. McKinnon further argues that future generations should not be subjected to "unbearable strains of commitment" which would render them unable to engage in the "joint pursuit of justice" (McKinnon 2009: 196). Essentially, she contends here that we cannot expect members of future generations to act in line with principles of justice if their living conditions are so bad, that they restrict them from pursuing their self-preservation and that of their families. Accordingly, the current generation therefore owes it to future generations to provide living conditions that it would accept for itself (McKinnon 2009: 194-197). This line of argument suggests that proportionally sharing the (financial and other) costs of climate action among the generations is the ethical and therefore imperative course of action, which would draw on the notion of "treating present and future people as equals", as McKinnon puts it (see above). Equal treatment is a fundamental aspect of human rights which finds expression in the key principle of non-discrimination. The treatment of individuals *as equals* is a matter pertaining to the moral status of all human beings and should not be confused with treating them *in exactly the same way* or providing them with the same amount of goods or opportunities (Moreau 2020: 8-9). It appears, then, that McKinnon shares the view of other philosophers and climate ethicists that the point in time of a person's birth, the factor which determines whether that individual is among 'present' or 'future' people, does not constitute legitimate grounds for unequal treatment (Herrler 2017: 164-172, Caney 2014: 323-325). However, this postulate – that people should receive treatment as equals regardless of when they are born – raises various questions, including the matters of whether people alive at present have obligations to future generations; and if so, what exactly these obligations consist in; and the grounds of their justification.

A strong 'precautionary principle' requires those in positions of influence to establish decision-making processes that take preventive measures to avoid unacceptable scenarios even where uncertainties exist on matters such as the likelihood of these scenarios' occurrence or their exact impact.

Could the language of human rights help answer these questions? I think so, and in advocating for the legitimacy of its use in this context, I will seek to show that people currently alive are indeed subject to moral obligations grounded upon the human rights of people living in the future. I will call these moral duties 'obligations with advance effect'. In asserting their validity, my argument will acknowledge and reflect the fact that it is difficult to speak

of the capacity of subjects who do not yet exist to have, and thus (be able) to exercise, rights. From a juridical point of view, in the words of Germany's Federal Constitutional Court, neither "unborn persons [n]or even entire future generations […] enjoy subjective fundamental rights" (BVerfG 2021: para. 109, also para. 146). While mindful of this objection, I am nevertheless of the view that the language of human rights can draw our attention to the equal *moral* status of human beings living in the future with those currently alive. "It is almost undisputed that we have present obligations and responsibilities towards future persons" (Tremmel 2009: 56); but the acceptance of this equality of moral status would imply the rejection of the notion that duties owed to future people are a lesser priority than other moral duties or obligations.

The starting point for my argument is the premise that human beings are, regardless of their identity,[3] holders of human rights as soon as they are born.[4] As we already have seen stated in Article 1 of the UDHR, all human beings are "born free and equal in dignity and rights". Proving the existence of obligations with advance effect requires us to make three further assumptions:[5]

A) Holders of human rights will live on Earth in the future.
B) Actions of those alive in the present have the capacity to affect the human rights entitlements of these future rights-holders.
C) This potential impact on future rights-holders is particularly the case for human rights entitlements pertaining to basic human needs that are likely to remain the same over time.

Assumption B describes the largely one-way direction of the impacts unleashed by the actions of present-day people, a circumstance which constitutes 'the pure intergenerational problem' (Gardiner 2003). This problem makes itself evident to us in the effects of anthropogenic climate change, and may entail the acknowledgement of existential risks to future generations flowing from the actions of those living now. Such existential risks might be considered a challenge to assumption A; however not as an essential challenge in terms of threatening the status of future human beings as rights-holders, but rather in terms of threatening the living conditions they require if they are to exercise these rights. The reference to basic human needs that is exemplified in assumption C operationalises an argumentative strategy that seeks to minimise opportunities for objections which cite the multi-faceted uncertainties invariably associated with the contention of an impact yet to come. As we cannot predict the future, such objections will likely arise, to varying degrees, in relation to all assumptions about the future. Karnein (2015: 47) encapsulates the epistemic challenge posed by these uncertainties thus:

"First, we do not know how many future generations there will be. Second, it is unclear what anyone can know about future generations' values and preferences because there is no chance of directly exchanging our views with theirs. Third, it is difficult to tell what the precise consequences of our actions will be, especially when it comes to the further future."

This notwithstanding, it is barely deniable at present that human beings will continue to need adequate food, clean water, and safe places to live even in the more distant future. For instance, Article 11 of the International Covenant on Economic, Social and Cultural Rights (ICESCR) addresses these needs. The precise nature of each need and the associated entitlement will obviously vary from case to case, with members of some populations, for example, requiring an adequately heated home and others needing an adequately cooled one – needs and entitlements on which global heating is already having an observable impact. It should be noted in this context that Caney (2010), who is using this type of argumentative strategy, even seeks to pre-empt potential objections by intentionally setting out the human rights to life, to health, and to subsistence in terms less rigorous than those found in the UDHR and in the ICESCR. Another argumentative strategy might refer not just to human rights pertaining to basic human needs as in assumption C, but conceptualise *all* human rights as a holistic entity, as a set of freedom rights encompassed in the principle of general freedom of action. The task of such a strategy would then be to successfully undergird the notion that every human right is applicable to members of future generations.

Human rights refer to entitlements that necessarily generate duties or obligations. From a moral point of view, it is irrelevant whether these duties or obligations concern actions with immediate effects or impacts that do not unfold until a point in the more distant future.

It should be recalled at this point that human needs or interests are not the same thing as human rights: "The content of a human right is the content of its associated duties, not of the interests that ground those duties" (Tasioulas 2015: 48). Human rights thus refer to entitlements that necessarily generate duties or obligations. From a moral point of view, it is irrelevant whether these duties or obligations concern actions with immediate effects or impacts that do not unfold until a point in the more distant future. The effectivity in advance of duties or obligations based on rights of others is not unusual; in fact, logically speaking, it seems to be the norm. As Bell (2011a: 107-108) writes:

"[A]ll human rights-based duties are current duties grounded in the future rights of persons living in the future (even if it is the very near or immediate future). […] Duties come temporally before human rights because actions come temporally before their effects. Human rights come normatively 'before' (i.e., they justify) duties because effects on human interests come normatively 'before' (i.e., they justify) restrictions on actions that cause those effects."

As I have shown, the imperative of avoiding potentially harmful impacts – particularly, not exclusively, in cases of potential existential risk – also appears in the precautionary principle. In a similar manner, the idea of human rights obligations expresses a desire to prevent human rights violations before they can occur. If it is assumed that basic human needs will remain more or less the same in the future and that global heating will jeopardise the human rights entitlements associated with these needs, then one can affirm the current existence of obligations to mitigate and adapt to climate change. Such obligations are effective in advance of the rights-holders' existence and have the aim of minimising, as far as possible, the restriction or violation of these entitlements and freedom rights. From a human rights perspective, then, inadequate climate action would perpetrate intergenerational injustice with a disproportionate impact on future people, who are vulnerable due to their incapacity to effect change in the present time. If one progresses beyond the strictly intertemporal understanding and extends this group to include people already born (and speaks of a succeeding generation, see note 1), the epistemic challenge on the grounds of uncertainty weakens, and it becomes

even more difficult to query the status of this group's members as rights-holders. Further, some members of this group may be able to exert a degree of influence on climate policy decisions; indeed, climate activists from initiatives such as *Fridays for Future* and the German *Letzte Generation* are currently engaged in such action. Incorporating the needs of succeeding generations in considerations of climate impacts renders the task less abstract and therefore significantly easier than the determination of needs and impacts relating to generations of the distant future.

Specifying the human rights obligations

An instructive object lesson in this context is the challenge facing Germany's Federal Constitutional Court as it ruled on the constitutional complaints brought against the 2019 Federal Climate Change Act. In its Order of 24 March 2021, the Court's First Senate provides some guidance on how one might conceive more specifically of human rights obligations in the context of anthropogenic climate change.[6] Partially upholding the constitutional complaints against the Act, the Court crucially set out the notion of "an advance interference-like effect on the freedom of the complainants […] that is comprehensively protected under the Basic Law" (BVerfG 2021: para. 184).[7] In my view, this idea leads us to the same line of reasoning as does the proposition of an obligation with advance effect. Other parts of the Order make specific mention of 'duties' and 'obligations'. Indeed, as early as the Order's first headnote, the Court observes that the "state's duty of protection […] encompasses the duty to protect life and health against the risks posed by climate change" and can "furthermore give rise to an objective duty to protect future generations" (BVerfG 2021: headnote 1). One can thus follow what appears to be the First Senate's thinking in conceiving of human rights obligations with advance effect as *obligations to protect* not only individuals already alive, but also people yet to be born. Additionally, one may consider such obligations as *obligations to respect*, because "[u]nder certain conditions, the Basic Law imposes an obligation to safeguard fundamental freedom over time and to spread the opportunities associated with freedom proportionately across generations" (BVerfG 2021: headnote 4). In this context, the reference is to the costs and burdens associated with far-reaching climate action and the need to avoid imposing them disproportionately onto people living in the future. Alongside this, the Court points to the necessity of treating the natural foundations of life with care, so "that future generations who wish to carry on preserving these foundations are not forced to engage in radical abstinence" (BVerfG 2021: headnote 4). This point, reminiscent of McKinnon's argumentation as set out above, implies an obligation to respect future people *as equals* to those currently alive; it is this moral standard that one will presumably have to apply if one is to identify an inappropriate or disproportionate intergenerational distribution of the opportunities associated with fundamental freedom. Finally, human rights obligations are also *obligations to fulfil*, and as such require states to take positive action to enable people to exercise their human rights to the fullest possible extent (Krennerich 2013: 106). The Court's view is that "[r]especting future freedom also requires the transition to climate neutrality in good time" (BVerfG 2021: headnote 4); continuing, the Order advises that "[i]n all areas of life […] developments need to be set in motion to ensure that in the future, meaningful use can still be made of freedom protected by fundamental rights, but then based on CO2-free alternatives" (BVerfG 2021: para. 248). While recognising that "the state itself has neither the capacity to achieve this transition alone nor the sole responsibility for doing so", the Court notes that "[c]onstitutional law nevertheless obliges the legislator to create the underlying conditions and incentives that would allow these developments to occur" (BVerfG 2021: para. 248, see also headnote 5).

> The conception of human rights obligations emerges here as obligations to protect, respect, and fulfil human rights.

The conception of human rights obligations that emerges here, as *obligations to protect, respect, and fulfil human rights*, is established in human rights discourse (UN ECOSOC 1987: sections 67-69). This is not the case, however, for the application of human rights to future generations. Significantly, a report on the relationship between climate change and human rights asserts:

"Human rights treaty bodies have alluded to the notion of intergenerational equity. However, the human rights principles of equality and non-discrimination generally focus on situations in the present, even if it is understood that the value of these core human rights principles would not diminish over time and be equally applicable to future generations" (UN HRC 2009: section 90).

If the value of the mentioned 'core human rights principles' does not diminish over time, one might wonder then, whether there might be the possibility of a wrong discrimination on the basis of the generation a person is born into. In the section that follows, I will set out an argument for the possible existence of such discrimination and the capacity of failure to act on climate change constitutes an instance thereof. While doing so, I will keep in mind that "the formulation of human rights remains an unfinished business" and that it "requires openness for further adaptations, modifications, amendments and reformulations" (Bielefeldt 2022: 77).

Does the imposition of climate risks on future generations constitute wrong discrimination?

Lewis (2018: 165) observes that, despite some juridical limitations, "there is still significant rhetorical and moral value attached to the language of human rights and consequently much to be gained from its continued linkage with climate change". In my view, engaging the concept of discrimination in this context would much advance the unfolding of this value's full impact. At the present time, the generation a person is born into does not appear in typical lists of prohibited grounds of discrimination; such lists, however, are non-exhaustive by design, leaving space for new protections – notwithstanding any uncertainties around their practical effect in the juridical dimension of human rights. Article 2 of the UDHR, for example, lists "race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status" as prohibited grounds of discrimination; the closing "or other status" emphasises the list's non-exhaustive character. General Comment No. 20 of 2 July 2009 (E/C.12/GC/20) on non-discrimination in economic, social and cultural rights (art. 2, para. 2 of the ICESCR) specifies in its sections 24-26, that in this context, 'birth' refers not to the point in time of a person's birth, but to its circumstances, such as the parents' marital status. Adding 'generational discrimination' (or a similar concept, named differently) to the list would require both the application of discrimination as a concept to the intergenerational context and its characterisation as morally wrong

in this same context. I will therefore commence the argument with reference to authors on the ethics of discrimination. It is my impression that many of their discussions of age discrimination refer to age groups (such as children and the elderly) rather than to birth cohorts (Bidadanure 2018). Unequal treatment of those belonging to different age groups is not of crucial interest in relation to the issues of intergenerational justice I discuss in this article. Indeed, taken over a person's lifetime, such unequal treatment may not in fact result in inequalities because the age group to which a person belongs changes – unlike the point in time of their birth (Bidadanure 2016: 239-240). Instead, I focus here on the disparate effects of specific practices on birth cohorts, such as the unequal risk of exposure to extreme climatic or meteorological conditions.

What do we mean when we speak of 'discrimination' in general? Put somewhat roughly, discrimination occurs where a subject, X, perceives (accurately or otherwise) the object of discrimination, Y, as possessing property P, and treats Y differently from another individual, group or entity, Z, that X does not perceive as having P. The use of 'subject' and 'object' here is in a grammatical sense; X, Y and Z could be individuals of any gender, or "superindividual entities such as private companies, social structures, and states" or indeed "possible people" (Lippert-Rasmussen 2014: 14-22; quote on p. 19). Discrimination, then, generically describes unequal, usually disadvantaging treatment on the basis of an actual or perceived property or trait. Let us assume a person living in the future (Y) has been, or will be, born later than another person (Z); P stands for the point in time at which Y's birth occurs. Let us further assume, for the sake of simplicity, that Z is a contemporary of the currently living X. If the consequences of an action by X have a disadvantageous effect on Y that is disproportionately greater than their deleterious effect on Z, it may be the case that Y has suffered discrimination on the basis of the point in time or generation of their birth. An example of such disadvantage might be an event precipitated by greenhouse gas emissions, which has a serious impact on Y, possibly to the extent of threatening their ability to live and meet their needs. Risks stemming from an act are 'imposed risks' when those affected by the act's possible consequences were not its agents (Nida-Rümelin et al. 2012: 8).

When considering the moral status of this type of discrimination, one can usefully draw on the conceptualisation of the issue proposed by Moreau (2020: 1-11), who considers wrongful (that is, in most cases morally unacceptable) discrimination not to be a matter of drawing erroneous, or wrongful, distinctions between individuals or groups, but rather one that prompts us to ask, "[w]hen we disadvantage some people relative to others on the basis of certain traits, when and why do we wrong them by failing to treat them as the equals of others?" (Moreau 2020: 7). In so doing, Moreau observes that the focus shifts from those perpetrating discrimination, and their intentions, to those discriminated against and the impact they sustain. This question additionally emphasises the fact of unequal treatment being visited upon people of equal moral status – the establishment and enshrinement of which is, as set out above, a key concern of the language of human rights, which asserts all human beings' right to enjoy equal respect. Moreau (2020: 12-24) further makes reference to the commonly drawn distinction between direct and indirect discrimination (or, in the US context, 'disparate treatment' and 'disparate impact'). It might appear at first glance that indirect discrimination is of greater relevance to the intergenerational context than is the direct form. Indirect discrimination, while it does not entail the use of a characteristic as grounds for explicitly singling out a person or group, does see those with that trait or property put at a disadvantage because of it. In this way, an act or practice, such as emitting greenhouse gases, has an *impact* on one group, such as currently living people who benefit from access to sources of energy, that is *disparate* from the impact it has on another group, such as people living in the future who bear the long-term cost or disadvantages of these emissions. In this example, the trait of the two groups that leads to their unequal treatment is the period of time within which their birth occurs. The causal chain initiated by the act – i.e. the emission of greenhouse gases causes the greenhouse effect that leads to global heating and its serious implications – thus results in unequal treatment of the two groups. As such the act constitutes discrimination, although the perpetrators do not necessarily have to be aware of this effect and it can thus arise without any malicious intent on their part (Hellman 2008: 138-168). A failure to take adequate climate action does not have to be deliberately intended to harm or to wrong future generations for it to discriminate against them.

The distribution of costs and benefits[8] that occurs, for example, through a failure to take adequate action on climate change, seems particularly unfair because those that benefit from this lack of action and those that suffer from it belong to different groups. In an analogous manner, existential risks seem more serious if they are imposed risks, that is, risks whose negative impact extends beyond the actors who bring those risks into being. Whether, for instance, an actor chooses to take the risk of crossing the Mediterranean in a rubber boat is a matter for them alone; the case is different, however, if they find themselves indirectly forced to make this journey because the situation in their home country has become intolerable. It is admittedly the case that a decision someone takes can run counter to their long-term interests; this, though, rather than being an imposed risk in the narrower sense, would count as an unwise course of action. Decisions in which those negatively affected by them had no participatory voice are more serious from a moral point of view, as both Thompson (2010: 20) and Caney (2016: 138-139, 2010: 170) emphasise. Thompson (2010: 17) goes as far as to use the term 'presentism', evidently in analogy to sexism, racism, and so on, to describe the intergenerationally unequal distribution of opportunities and risks. He defines the concept as signifying "a bias in the laws in favor of present over future generations" and identifies its presence in democracies in, for example, "laws that neglect of long-term environmental risks". The Federal Constitutional Court echoes this train of thought when it speaks of the democratic political process being "organised along more short-term lines based on election cycles, placing it at a structural risk of being less responsive to tackling the ecological issues that need to be pursued over the long term" (BVerfG 2021: para. 206). In addition to Thompson (2010: 19) and Caney (2016: 143), Gardiner (2003: 491) and MacKenzie (2016: 25-30) draw critical attention to the short-termism of many political (and economic) decision-making processes. It is, of course, not necessarily, let alone always, morally wrong to have an interest in relatively short-term successes; one needs, then, to identify the point at which 'presentism' becomes morally wrong discrimination on the basis of the point in time of an individual's birth.

It is not necessarily morally wrong to have an interest in relatively short-term successes. One needs to identify the point at which 'presentism' becomes morally wrong discrimination on the basis of the point in time of an individual's birth.

Returning to Moreau's question regarding "when and why […] we wrong [people] by failing to treat them as the equals of others", it should be noted that she puts forward three answers in this context. To assess them in detail would exceed the scope of this article, but I believe that they can provide us with food for thought in the context of our considerations. Moreau (2020: 39-75) suggests, first, that discrimination can be morally wrong if it entails an "unfair subordination" of some people to others.[9] She refers to existing stereotypes that contribute to the disadvantaged status of the group discriminated against; she defines such stereotypes as "generalizations about particular social groups that ascribe most of their members certain desires, dispositions of behavior, or obligations, simply because they possess whatever trait defines that group, as a group" (Moreau 2020: 54). A potential objection to the application of this argument in the context of future generations could be the difficulties possibly associated with determining which "certain desires, dispositions of behavior, or obligations" issue from the point in time of an individual's birth, particularly where that birth lies in the future. Moreau's (2020: 59-60) approach further appears to assume that the 'needs' of the subordinated group differ from those of the privileged group. If we consider the era of someone's birth as a protected characteristic on the basis of basic human needs that are presumably consistent and unchanging (assumption C above), attempts to identify supposed stereotypes or the neglect or denial of specific needs may not be helpful in making this argument. I am therefore sceptical about the use of vocabulary such as 'subordination', 'demeaning' and 'inferior' in a critique of 'presentism', even though views on this may of course diverge.

I find greater traction in the remaining two answers Moreau provides to her question. The second of the three asserts that the wrongfulness of discrimination may arise from is its violation of a person's "deliberative freedoms" (Moreau 2020: 77-110), that is,

"the freedom to deliberate about one's life, and to decide what to do in light of those deliberations, without having to treat certain personal traits (or other people's assumptions about them) as costs, and without having to live one's life with these traits always before one's eyes" (Moreau 2020: 84).

Taking 'the point in time of a person's birth' as such a trait may initially sound unfamiliar; however, future generations – possibly faced with living on a planet in a 'hothouse' climate state – might well ask themselves whether they are still "born free and equal in dignity and rights" compared to people alive at the beginning of the twenty-first century. It may be a legitimate part of the remit of human rights institutions and activists to address this question in the present and to pose it to those in danger of committing discrimination. The current generation faces the choice of either imposing these risks on future people or attempting to avoid or at the least mitigate their dire consequences. In this case, freedom, which Moreau (2020: 89-90) links to the human capacity for autonomy, can be related to "the general freedom of action […] as the elementary fundamental right to freedom" (BVerfG 2021: para. 184) and the obligation "to spread the opportunities associated with freedom proportionately across generations" (BVerfG 2021: headnote 4). The implication of this line of argument is that imposing a risk means – to borrow Ferretti's phrasing (2016: 262-264) – interfering with a third party's "set of sets of options" and diminishing the "overall freedom" enjoyed by that third party. Ferretti (2016: 262) contends that "[d]ecreasing people's overall freedom under a certain acceptable level fails to treat them in the appropriate manner", that is, fails to respect them as moral equals. What remains undetermined at this juncture is the exact point at which this acceptable level of freedom is no longer being met. Which types of imposed risk call for the operation of a strong precautionary principle because, for instance, the risk's realisation could pose an existential danger? This question would need further consideration.

The third answer given by Moreau (2020: 121-151) relates to people's access to basic goods. Citing the lack of safe drinking water in reserves for indigenous populations in Canada, she notes:

"The water crisis does not just deny indigenous peoples something basic to survival, to which they have a human right. In the process, it prevents them from participating fully and as an equal in Canadian society. And it also denies them the ability to be seen as full and equal participants, and to see themselves as such" (Moreau 2020: 125).

As is evident from this example, Moreau's insistence (2020: 125-126) is that a good is a "basic good" for a particular person if "[a]ccess to this good is necessary in order for this person to be a full and equal participant in her society; and […] in order for this person to be seen by others and by herself as a full and equal participant in her society." The concern with interactional inequalities among contemporaries that appears predominant in this argument makes it harder for us to apply Moreau's line of thinking to the idea that the point in time of someone's birth could constitute a trait meriting protection against wrongful discrimination. This said, this focus also enables us to raise some important questions for our context. If the barriers to accessing basic goods that marginalised people face in the present prevent these people, now, from participating in what McKinnon terms the "joint pursuit of justice", can it be morally defensible to impose these barriers, and their implications, on what are likely to be larger numbers still of people living in the future? Particularly if one looks beyond the relatively wealthy societies of the global North, it can be observed in the present day, that "acute food insecurity and reduced water security", alongside "adversely affected physical […] and mental health", are among the consequences of global heating (IPCC 2022: 11-13). This global injustice taking place in our time has the potential to educate the populations of the Global North on the conditions that likewise pose a threat to their succeeding and future generations, to whom they presumably have closer emotional ties than to the inhabitants of the Global South. I do not, of course, wish to suggest that injustice done to the former group would be any more deplorable than that done to the latter, but simply to note the potential capacity of relatability to 'our people' or 'people like us' to prompt action among those thus far insulated from the impacts of climate change.

Human rights hold progressive potential because, while they protect specific standards, they may not necessarily extend the same protection to the currently established paths to these standards' achievement or maintenance.

It can be seen, then, that the effects of climate change in our own time are already limiting people's access to basic goods, a matter that falls within the purview of human rights. If this is already the case now, it seems certain that greater numbers of people will find access to their basic needs restricted in the future (see assumption C above). Should it really be more difficult for future

than for current generations to enjoy, for example, their human rights to an adequate standard of living and to health? Speaking in terms of the comparative, such as using the comparative adjective 'more difficult', leads to a further issue in relation to the concept of discrimination on the basis of a person's generation. One can describe this issue as concerning the quality of the inheritance we wish our descendants to receive from us. This is the core question of intergenerational justice. Does the present generation owe those of the future living conditions that are merely sufficient, or as good as, or indeed superior to, those currently in place (Roser/Seidel 2013: 45-59, Herrler 2017: 178-186)? A human rights-based approach does not automatically have to advocate for providing future people with the exact same amounts of basic goods, made available in exactly the same way, as are accessible to those living now. In my understanding, human rights hold progressive potential because, while they protect specific standards, they may not necessarily extend the same protection to the currently established paths to these standards' achievement or maintenance (Herrler 2020). However, such an approach is able to point out, among other things, that the 'obligation to respect' imposes on the current generation the duty to care for the natural foundations of life in such a way as to ensure future people's freedom to enjoy them and to preserve them in their turn. If this view is taken, one can read the definition of basic goods in terms of human rights as minimum standards of living to which all human beings of every generation are entitled (Roser/Seidel 2013: 55-56). In any case, a human rights-based approach would oppose the notion that future generations are not owed sufficient living conditions and would therefore denounce policies that, intentionally or otherwise, would see this notion realised by putting future people's access to basic goods at risk. Such policies would entail taking decisions about people who have little or no influence at all on these decisions, that is, who cannot participate *as equals* in the process leading up to them. Ultimately, this also raises the question of who counts as part of the demos in a democracy and whose entitlements decision-making processes should at least consider. The concept of human rights obligations effective in advance advocates for the consideration of people living in the future on their behalf. The implication of such consideration is not, of course, that only the assumed needs of those very young or yet to be born should hold decisive weight in decisions and actions taken now; it should be noted here that there are cases in which discrimination may be justifiable "all things considered" (Moreau 2020: 11-12). Used to denounce discrimination on the basis of the point in time of someone's birth, the language of human rights can and should require decision-makers to explicitly state, and explain the legitimacy of, the reasons for decisions that disadvantage future generations. One may justifiably doubt the existence of such legitimate, convincing grounds for inadequate action on climate change in many cases.

Used to denounce discrimination on the basis of the point in time of someone's birth, the language of human rights can and should require decision-makers to explicitly state, and explain the legitimacy of, the reasons for decisions that disadvantage future generations.

Participation via representation: Bringing future people into our present consciousness

What, then, might be the specific, real-world implications of my argument as set out thus far? First, I believe that human rights obligations that are effective in advance have the capacity to undergird calls for strong action on climate change. In general, my argument also supports preventative efforts with the aim of keeping the risks imposed on future generations within reasonable limits; a further example might be advocacy for nuclear disarmament (SASB 2022a: 9-10). At this point, I would like to address another requirement that emerges from such obligations, on which I touched at the end of this article's previous section – that of participation in decision-making processes.

A generally necessary concomitant of duties and obligations is someone who demands or enforces compliance with them. Notwithstanding the fact that voluntary compliance is evidently desirable, it is equally evident, in the case of practically implemented climate action in the real world (as opposed to laudable stated goals), that such an ideal situation is far from being regular reality. In terms of human rights, it would likewise be desirable for those affected to formulate and demand their rights and entitlements themselves, in the spirit of the disability rights movement's slogan '*nothing about us without us*'.[10] However, it is frequently the case that vulnerable groups whose human rights are in particular need of protection find themselves neither seen nor heard in decision-making processes that concern them, resulting in decisions that fail to properly consider their needs. Their vulnerability therefore co-emerges from their marginalised position in relation to power structures. Compounding the vulnerability of succeeding and future generations is the fact that many of them are literally invisible and voiceless. At the political level, this problem emerges where actors seek to adhere to the democratic principle of '*all affected interests*', which provides, roughly speaking, that a person should at least be able to have a say and be heard on matters concerning them, so they can, for example, demand that other actors comply with obligations towards them in relation to that matter. This opportunity is of importance to duties and obligations based on human rights. "To have a right implies the possibility *to insist* on its being respected" (Bielefeldt 2022: 28). If those affected by a decision or course of action cannot take this opportunity themselves, then representatives must take it on their behalf. In so doing, they both make those they represent 'present' in the decision-making process and – in this specific case – bring these future generations into our present time and our present consciousness. The literature in this area to date contains numerous proposals for institutional representation for future generations (see, for example, González-Ricoy/Gosseries 2016, Cordonier Segger et al. 2021). The task of representing future generations is not without its problems (Karnein 2016) and would require considerations around the remit and powers of the institutions charged with this representation, the source of their legitimation, the selection of suitable candidates for the associated roles, and cooperation with other institutions. Ultimately, however, notwithstanding the uncertainties around the needs and values of the future generations, on whose behalf such institutions would act, it is a plausible assumption that "the only standard [these future generations] could reasonably be expected to accept is to be treated with equal respect". Institutionalised representation of future generations would therefore need to take particular heed of the fact "that we [as the current generation] would have to justify our decisions to future generations *as if they were present today*" (Karnein 2016: 93).

The representatives working within this context could use the language of human rights to, for instance, condemn the imposition of existential risks on young human beings and those not yet

born. They could then require decision-makers to protect, respect, and realise their human rights entitlements, and, potentially, to use the concept of discrimination to shine a light on 'presentism'. While 'discrimination based on the point in time of a person's birth' may still sound a strange notion to many ears, I believe it is a conceivable one with potentially considerable moral force, particularly if it is employed in concert with people's entitlement to fundamental freedoms. If it could shift the burden of justification in favour of future people, much would be gained. It is desirable that taking action is no longer solely focused on its short-term advantages to the current generation, but aware of the action's capacity to impose (possibly impermissible) long-term disadvantages on future generations. In the case of existential risks, treating people living in the future as equals to those living today would presumably occasion greater risk aversion in decision-making, in the spirit of a strong precautionary principle. Bielefeldt (2022: 43-58) highlights – quite rightly, in my view – the character of human rights as rights pertaining to individuals; he also, however, notes that they are ultimately not individualistic rights, but "relational rights", a point which brings us full circle to Sandel's assertion of the need for "some kind of communal language" in this context. If we wish our relationship to the generation that follows ours, and those that succeed them, to be characterised by equal respect, then we should refrain from imposing inappropriate risks on these generations. Instead we should advocate for the fairer distribution of the advantages and disadvantages of our political decisions across generations. The purpose of human rights is not to enable each individual to live a self-sufficient and self-centred life apart from communities. The principles of equal treatment and non-discrimination that underlie them aim rather to enable and empower all human beings to live together in freedom and peace – an aim that certainly has an intertemporal dimension. The full realisation of the human rights ideal of freedom and equality in all political communities will likely remain a dream for a very long time to come, but it remains our obligation to at least refrain from consigning future generations to an ecological nightmare.

---

1 This understanding of the term 'generation' is chronological-intertemporal, classing all people currently living as belonging to one generation (the 'current' or 'present' generation), whereas members of a 'future generation' do not yet exist at the time the reference is made (Tremmel 2009: 20-26). As future generations are dependent for their existence on the current one, the existential risk to those alive in the present is likewise a risk, indirectly, to them. My argument will primarily engage with problems arising from the "asymmetry of power" (Barry 1989: 496) in intergenerational relations, which also affects many young people who have already been born. I will use the term 'succeeding generation' if I intend to refer to a future generation whose members partly do already exist (Tremmel 2009: 64-65).

2 Not all such merit the term as 'risk' understood in the definition espoused by Nida-Rümelin et al. (2012: 7-8), which holds that risks are always connected to decisions or actions taken by specific actors. Such an understanding of risk would exclude, for instance, the danger of a meteorite impact, although the failure to take defensive measures in light of this danger would then establish a connection to an actor.

3 Rights-based approaches can get around the non-identity problem (Parfit 1984: 351-379) more easily than can competing person-affecting views (for further discussion, see, for example, Baatz 2016: 95-104; Herrler 2017: 159-163; Meyer 2018: 89-106; Page 2006: 132-160).

4 I will refrain from discussing the disputed and contentious issue of when exactly a subject of rights comes into existence (birth, procreation, or similar), as it is irrelevant to the further course of my argument.

5 What follows here draws most closely on the argument proposed by Bell (2011a: 104-110); other authors (Baatz 2016: 93-95, 111f.; Düwell 2016: 79-80; Kleiber 2014: 287-289; Meyer 2018: 83-89) make or discuss similar assumptions.

6 The Court, of course, does not refer to 'human rights', but to the fundamental rights codified in the German Basic Law (= Grundgesetz [GG]). The Basic Law specifies 'all Germans' as the holders of some of the fundamental rights it enumerates; I will, for the sake of simplicity, refrain from explicitly distinguishing between these and those applying to all people without specification of nationality. However, Article 2 GG, which is key to the Court's argumentation, is not limited to Germans in its formulation.

7 The quoted passage continues as follows: "The Basic Law protects all human exercise of freedom through special fundamental rights to freedom, as well as through the general freedom of action enshrined in Art. 2(1) GG as the elementary fundamental right to freedom" (BVerfG 2021: para. 184). This means that ultimately, a strategy referring to all human rights was the successful one in this case, even though the matter engaged the legal rights of complainants already born (Ekardt/Heß 2021: 579-580).

8 I use the terms 'costs' and 'benefits' here in reference to the 'Pure Intergenerational Problem', whose essential cause is the fact that groups, or rather generations, have "access to goods which […] give modest benefits to the group which produces them, but impose high costs on all later groups" (Gardiner 2003: 483-485). The use of these terms is in no way intended to suggest that economic cost-benefit calculations are better suited to advocacy for climate action than is the language of human rights. Indeed, I believe the reverse is true (Herrler 2020).

9 The approach employed by Hellman (2008) uses a similar angle, identifying "demeaning" treatment as a defining factor in wrongful discrimination. But her idea of "demeaning" treatment is closely dependent of the specific contexts and cultures in which it occurs. This places limitations on the concept's applicability to future generations, because it implies, as she herself concedes, a potential conflict with universal human rights (Hellman 2008: 27-42).

10 It is worth noting here that "it would be wrong to infer that only those personally affected should feel entitled to talk about discrimination. Non-discrimination agendas are a political project that requires broad alliances of people from different backgrounds and with a variety of experiences and skills, also across the minorities-majorities-divide" (Bielefeldt 2022: 107 [note 217]).

References

Baatz, Christian (2016): Compensating Victims of Climate Change in Developing Countries: Justification and Realization. https://www.philsem.uni-kiel.de/de/lehrstuehle/philosophie-und-ethik-der-umwelt/files/baatz-2016_compensating-climate-change-victims-1/at_download/file. Viewed 21 May 2022.

Barry, Brian (1989): Justice between Generations. In: Barry, Brian, Democracy, Power and Justice: Essays in Political Theory. Oxford: Clarendon Press, 494-510.

Bell, Derek (2011a): Does anthropogenic climate change violate human rights? In: Critical Review of International Social and Political Philosophy, 14 (2), 99-124.

Bell, Derek (2011b): Global Climate Justice, Historic Emissions, and Excusable Ignorance. In: The Monist, 94 (3), 391-411.

Bidadanure, Juliana (2018): Discrimination and Age. In: Lippert-Rasmussen, Kaspar (ed.): The Routledge Handbook of the Ethics of Discrimination. London, New York: Routledge, 243-253.

Bidadanure, Juliana (2016): Making sense of age-group injustice: A time for relational equality? In: Politics, Philosophy & Economics, 15 (3), 234-260.

Bielefeldt, Heiner (2022): Sources of Solidarity: A Short Introduction to the Foundations of Human Rights. Erlangen: FAU University Press.

Bundesverfassungsgericht (BVerfG) (2021): Order of the First Senate of 24 March 2021 – 1 BvR 2656/18 –, paras. 1-270, https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/2021/03/rs20210324_1bvr265618en.html. Viewed 22 September 2022.

Caney, Simon (2016): Political Institutions for the Future: A Fivefold Package. In: González-Ricoy, Iñigo / Gosseries, Axel (eds.): Institutions for Future Generations. Oxford: Oxford University Press, 135-155.

Caney, Simon (2014): Climate change, intergenerational equity and the social discount rate. In: Politics, Philosophy & Economics, 13 (4), 320-342.

Caney, Simon (2010): Climate Change, Human Rights, and Moral Thresholds. In: Gardiner, Stephen M. / Caney, Simon / Jamieson, Dale / Shue, Henry (eds.): Climate Ethics: Essential Readings. Oxford, New York: Oxford University Press, 163-177.

Caney, Simon (2009): Climate Change and the Future: Discounting for Time, Wealth and Risk. In: Journal of Social Philosophy, 40 (2), 163-186.

Cordonier Segger, Marie-Claire / Szabó, Marcel / Harrington, Alexandra R. (eds.) (2021): Intergenerational Justice in Sustainable Development: Treaty Implementation advancing future generations rights through national institutions. Cambridge: Cambridge University Press.

Düwell, Marcus (2016): Human dignity and intergenerational human rights. In: Bos, Gerhard / Düwell, Marcus (eds.): Human Rights and Sustainability: Moral responsibilities for the future. London, New York: Routledge, 69-81.

Ekardt, Felix / Heß, Franziska (2021): Intertemporaler Freiheitsschutz, Existenzminimum und Gewaltenteilung nach dem BVerfG-Klima-Beschluss – Freiheitsgefährdung durch Klimawandel oder durch Klimapolitik? In: Zeitschrift für Umweltrecht, 11, 579-585.

Ferretti, Maria P. (2016): Risk imposition and freedom. In: Politics, Philosophy & Economics, 15 (3), 261-279.

Gardiner, Stephen M. (2010): Ethics and Global Climate Change. In: Gardiner, Stephen M. / Caney, Simon / Jamieson, Dale / Shue, Henry (eds.): Climate Ethics: Essential Readings. Oxford, New York: Oxford University Press, 3-35.

Gardiner, Stephen M. (2003): The Pure Intergenerational Problem. In: The Monist, 86 (3), 481-500.

González-Ricoy, Iñigo / Gosseries, Axel (eds.) (2016): Institutions for Future Generations. Oxford: Oxford University Press.

Hellman, Deborah (2008): When Is Discrimination Wrong? Cambridge/London: Harvard University Press.

Herrler, Christoph (2020): Warum eigentlich (nicht) Menschenrechte? – Zum Diskurs über die Klimakrise. https://www.praefaktisch.de/klimakrise/warum-eigentlich-nicht-menschenrechte-zum-diskurs-ueber-die-klimakrise. Viewed 30 May 2022.

Herrler, Christoph (2017): Warum eigentlich Klimaschutz? Zur Begründung von Klimapolitik. Baden-Baden: Nomos.

Intergovernmental Panel on Climate Change (IPCC) (2022): Climate Change 2022 Impacts, Adaptation, and Vulnerability: Summary for Policymakers, Working Group II contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, https://report.ipcc.ch/ar6wg2/pdf/IPCC_AR6_WGII_SummaryForPolicymakers.pdf. Viewed 19 May 2022.

Karnein, Anja (2016): Can we Represent Future Generations? In: González-Ricoy, Iñigo / Gosseries, Axel (eds.): Institutions for Future Generations. Oxford: Oxford University Press, 83-97.

Karnein, Anja (2015): Climate Change and Justice between Non-overlapping Future Generations. In: Global Justice: Theory Practice Rhetoric, 8 (2), 43-65.

Kleiber, Michael (2014): Der grundrechtliche Schutz künftiger Generationen. Tübingen: Mohr Siebeck.

Krennerich, Michael (2013): Soziale Menschenrechte: Zwischen Recht und Politik. Schwalbach/Ts.: Wochenschau.

Lenton, Timothy M. / Rockström, Johan / Gaffney, Owen et al. (2019): Climate tipping points – too risky to bet against. In: Nature, 575, 592-595.

Lewis, Bridget (2018): Environmental Human Rights and Climate Change: Current Status and Future Prospects. Singapore: Springer Nature.

Lippert-Rasmussen, Kasper (2014): Born free and equal? A philosophical inquiry into the nature of discrimination. Oxford/ New York: Oxford University Press.

MacKenzie, Michael K. (2016): Institutional Design and Sources of Short-Termism. In: González-Ricoy, Iñigo / Gosseries, Axel (eds.): Institutions for Future Generations. Oxford: Oxford University Press, 24-45.

McKinnon, Catriona (2009): Runaway Climate Change: A Justice-Based Case for Precautions. In: Journal of Social Philosophy, 40 (2), 187-203.

Meyer, Kirsten (2018): Was schulden wir künftigen Generationen? Herausforderung Zukunftsethik. Stuttgart: Reclam.

Moreau, Sophia (2020): Faces of Inequality: A Theory of Wrongful Discrimination. Oxford: Oxford University Press.

Nida-Rümelin, Julian / Rath, Benjamin / Schulenburg, Johann (2012): Risikoethik. Berlin/Boston: De Gruyter.

Page, Edward A. (2006): Climate Change, Justice and Future Generations. Cheltenham/ Northampton: Edward Elgar.

Parfit, Derek (1984): Reasons and Persons. Oxford: Clarendon Press.

Roser, Dominic / Seidel, Christian (2013): Ethik des Klimawandels: Eine Einführung. Darmstadt: WBG.

Sandel, Michael J. (2006): The Peril of Extinction. In: Sandel, Michael J., Public Philosophy: Essays on Morality in Politics. Cambridge: Harvard University Press, 179-182.

Science and Security Board (SASB) of the Bulletin of the Atomic Scientists (2022a): At doom's doorstep: It is 100 seconds to midnight, 2022 Doomsday Clock Statement. https://thebulletin.org/wp-content/uploads/2022/01/2022-doomsday-clock-statement.pdf. Viewed 22 May 2022.

Science and Security Board (SASB) of the Bulletin of the Atomic Scientists (2022b): Bulletin Science and Security Board condemns Russian invasion of Ukraine; Doomsday Clock stays at 100 seconds to midnight. https://thebulletin.org/2022/03/bulletin-science-and-security-board-condemns-russian-invasion-of-ukraine-doomsday-clock-stays-at-100-seconds-to-midnight. Viewed 22 May 2022.

Tasioulas, John (2015): On the Foundations of Human Rights. In: Cruft, Rowan / Liao, S. Matthew / Renzo, Massimo (eds.): The Philosophical Foundations of Human Rights. Oxford: Oxford University Press, 45-70.

Thiery, Wim / Lange, Stefan / Rogelj, Joeri et al. (2021): Intergenerational inequities in exposure to climate extremes: Young generations are severely threatened by climate change. In: Science, 374 (6564), 158-160.

Thompson, Dennis F. (2010): Representing future generations: political presentism and democratic trusteeship. In: Critical Review of International Social and Political Philosophy, 13 (1), 17-37.

Tremmel, Jörg (2009): A Theory of Intergenerational Justice. London: Earthscan Publishing https://www.researchgate.net/publication/29748980_A_Theory_of_Intergenerational_Justice. Viewed 3 November 2022.

Umweltbundesamt (UBA) (2021): Die Treibhausgase. https://www.umweltbundesamt.de/themen/klima-energie/klimaschutz-energiepolitik-in-deutschland/treibhausgas-emissionen/die-treibhausgase. Viewed 23 May 2022.

United Nations Economic and Social Council (UN ECOSOC) (1987): The New International Economic Order and the Promotion of Human Rights: Report on the right to adequate food as a human right submitted by Mr. Asbjørn Eide, Special Rapporteur, E/CN.4/Sub.2/1987/23 (7 July 1987). https://undocs.org/en/E/CN.4/Sub.2/1987/23. Viewed 23 May 2022.

United Nations Human Rights Council (UN HRC) (2009): Report of the Office of the United Nations High Commissioner for Human Rights on the relationship between climate change and human rights, A/HRC/10/61 (15 January 2009). https://www.refworld.org/docid/498811532.html. Viewed 20 May 2022.

At the time of writing, Christoph Herrler was a postdoctoral lecturer and researcher at the Chair of Political Philosophy, Political Theory and History of Political Ideas, Institute of Political Science, Friedrich-Alexander-University Erlangen-Nuremberg, Germany. His main research interests are climate ethics, discrimination and human rights.

Email: christoph.herrler@fau.de

The post-antibiotic era: An existential threat for humanity[1]

by Dominik Koesling and Claudia Bozzaro

Currently, mankind is facing the risk of running out of working antibiotics. Such a post-antibiotic era bears tremendous risks such as globally spread or even pandemic bacterial infections. These infections become thus untreatable and possibly lethal, particularly endangering the health (care) of future generations. This paper discusses this acute concern for humanity in three main steps. After first elaborating on the role of antibiotics and the occurring resistance in modern medicine, the focus will be on the current scope of the problem of antibiotics and the prognosis of its future escalation. Then the possibility of a way out and its obstacles will be addressed, before finally assessing the existential threat of a post-antibiotic era.

Keywords: *antibiotic crisis; post-antibiotic era; existential threat for humanity; global and intergenerational health (care); global and intergenerational (in)justice*

Introduction: The lingering danger of a post-antibiotic era

Antibiotic resistance is on the rise and humanity is currently heading towards a post-antibiotic era. Although this scenario is unlikely to lead to the complete extinction of humanity, it poses an existential threat, as one of the most important means of fighting infections would then have become ineffective, resulting in the death of millions of people. Despite the fact that international bodies such as the World Health Organization have put this issue on the global political agenda, it continues to grow as problem. However, the actual danger posed by antibiotic resistance, which is essentially of anthropogenic origin (Mitchell et al. 2019: 1), does currently not correspond to its recognition as an immediate threat to humanity on a social, or more precisely on a societal level. Not even notorious catchwords like "superbugs" (Stolberg 1998) seem to be enough to bring the issue into public awareness. As of now, "warnings and crisis framings do not appear sufficient to prompt a response. Public attention and governmental action have lagged." (Engström 2021: 19).

A post-antibiotic era is, simply put, "a new era in which bacteria have become resistant to existing antibiotics and the antibiotics no longer work." (Hansson/Brenthel 2022: 381). Like some other current and anticipated future crises, such as the climate crisis, antibiotic resistance is developing day by day beyond our collective perception. This lack of awareness could make the current antibiotic crisis even more dangerous, as the absence of adequate threat perception is likely to reduce the willingness to tackle the problem. Picking up on these aspects, the guiding thesis of the paper at hand is, that, contrary to their widespread perception, *antibiotic resistance and the post-antibiotic age are an existential danger to humanity in form of a global and intergenerational threat.* The arguments to substantiate this claim are unfolded in three main steps: First, we will give an overview of the role of antibiotics and the occurring resistance in modern medicine. Building on this and taking a global perspective, we will highlight the current scope of the problem and elaborate on the prognosis of its future escalation, revealing the intergenerational nature of the issue at hand. Afterwards, we will focus on potential attempts to tackle antibiotic resistance and prevent a post-antibiotic era by elaborating on the possibility of a way out and its obstacles, before concluding the proposed arguments.

Antibiotic resistance is on the rise and humanity is currently heading towards a post-antibiotic era. Although this scenario is unlikely to lead to the complete extinction of humanity, it poses an existential threat, as one of the most important means of fighting infections would then have become ineffective, resulting in the death of millions of people.

The role of antibiotics and the occurring resistance in modern medicine

Prior to humanity's access to effective antibiotics in what can be called the pre-antibiotic era – most of human history – millions of people had to suffer and die from bacterial infections. This changed radically with the scientific discovery of antibiotics, and since then antibiotics have completely revolutionised medicine, not only being an effective means of treating infections, but also making medical procedures such as life-saving operations safe in the first place (Palmer 2022: ix; Friedman et al. 2016: 416, 420). Nowadays antibiotics are virtually omnipresent, especially in health care, and they have "extended the average human lifespan by 23 years." (Hutchings et al. 2019: 1). As indicated by research, millions of doses of antibiotics are administered every day in hospitals alone. A German study, for example, showed that even in the adjusted, representative sample of all participating hospitals, 21.5% of patients were treated with antibiotics (cf. Nationales Referenzzentrum für die Surveillance von Nosokomialen Infektionen 2016: 2, 20-21). While patients receive antibiotics for various reasons, e.g. to treat acute infections, they are also regularly over- or misused. Hence, it is hardly surprising that Fleming-Dutra et al. (2016: 1872) conclude their study with the remark that "[i]n the United States in 2010-2011, there was an estimated annual antibiotic prescription rate per 1000 population of 506, but only an estimated 353 antibiotic prescriptions were likely appropriate."

Prior to humanity's access to effective antibiotics in what can be called the pre-antibiotic era, millions of people had to suffer and die from bacterial infections.

But as wide as the range of medically appropriate and inappropriate antibiotics use is nowadays, the "arguably […]greatest medical breakthrough of the 20th century" (Gautam 2022: 225), are relatively new in historical retrospect. The discovery of the famous *penicillin* dates back to 1928 and from here on it took several years – until 1942 – before it was ready for widespread market use. Thus, humanity can only look back at round about 80 years

of effective antibacterial medical treatment and even today not all people around the globe have access to (proper) antibiotics. So while humanity as a whole has several generations of antibiotics at its disposal, not everyone benefits equally, if at all, raising questions about adequate supply and just global distribution of these goods.

In addition to creating broad access to antibacterial treatment options, however, one problem is particularly urgent, namely the loss of antibiotic efficacy. By now, it has become increasingly evident that the "[b]acteria are fighting back and are becoming resistant" (Davies et al. 2013: ix) to specific substances used against them. From an evolutionary point of view, this can be seen as an adaptation process of the bacteria to the selection pressure to ensure their own survival. Certainly, resistance should not be equated with the complete ineffectiveness of antibiotics, as any resistance that occurs is a specific response of bacteria to a particular antibiotic and not a general response to every antibiotic. Therefore, in some cases, it is possible to modify treatment with alternative antibiotics to provide or restore effective antibacterial treatment. However, this is highly unlikely in cases of so-called multi-resistant bacteria, which are characterised by simultaneous resistance to various antibiotics making their treatment extremely difficult or impossible. Some pathogens such as *Staphylococcus aureus* have shown a high adaptability and are "capable of becoming resistant to all classes of antibiotics clinically available" (Vestergaard et al. 2019: 1). For this reason, multi-resistant bacteria are a particular threat as medicine and mankind lack adequate treatment options in such cases. Causing "more than 100 000 deaths attributable to AMR (antimicrobial resistance, the authors) in 2019" (Antimicrobial Resistance Collaborators 2022: 629, cf. 638) the notorious strain of *Methicillin resistant Staphylococcus aureus* is evidence of the danger that antibiotic resistance poses to human life.

The current scope of the problem and its future trajectory

Surely, the prior considerations provide a sufficient basis for the implicit premise of the argument at hand that antibiotic resistance is indeed a serious problem. This is mainly due to the undesirable consequences, which range from increased resource consumption, e.g. in the form of treatment costs or duration, to treatment failure, leading to the death of the infected patients in the worst case. Despite these potentially serious consequences, antibiotic resistance is often not perceived as the problem it actually is, adding another dimension to the problem's complexity, which Engström (2021) has recently addressed in detail. However, the issues outlined are by no means news to anyone familiar with the field, as knowledge of these facts dates back to the early days of the scientific discovery of antibiotics (Friedman et al. 2016: 417). Pioneer scholars on bacterial infections, such as Fleming, who discovered *penicillin*, observed antibiotic resistance and the associated loss of this particular antibiotic's effectiveness. In his Nobel Prize speech in 1945, Fleming (1964 [1945]: 93) stressed the importance of understanding that it is the bacteria itself that become resistant when he stated:

"Here is a hypothetical illustration. Mr. X. has a sore throat. He buys some penicillin and gives himself, not enough to kill the streptococci but enough to educate them to resist penicillin. He then infects his wife. Mrs. X gets pneumonia and is treated with penicillin. As the streptococci are now resistant to penicillin the treatment fails. Mrs. X dies."

Providing this example, Fleming reminds everyone of the basic yet commonly misconceived fact that "[b]acteria, not humans or animals, become antibiotic-resistant." (World Health Organization 2020). By particularly zooming in on the micro-level of the family, he accounts for the potential extent of the problem at hand, which is both individual and social. In a nutshell: On the one hand, resistant bacteria can be lethal for the infected themselves, making them a matter of concern on an individual level. On the other hand, Mrs. X's contagion reminds us of the social aspects and effects of bacterial resistance. Flemming even anticipates the societal problems, as pathogenic bacteria may not stay in the organism in which they have developed their specific resistance but can spread in and through human interaction. Such direct effects of antibiotic-resistant bacteria go hand in hand with indirect ones and therefore "[t]he negative impacts of antibiotic resistance on healthcare systems as a whole are substantial, as resistance adds to the number of infections that occur, to expense, to interrupted hospital activity and to limitation of treatment options." (Friedman et al. 2016: 420).

On the one hand, resistant bacteria can be lethal for the infected themselves, making them a matter of concern on an individual level. On the other hand, pathogenic bacteria may not stay in the organism in which they have developed their specific resistance but can spread in and through human interaction.

Already, these negative effects have taken their toll on humanity's potential to provide antibacterial medical treatment. As such "[o]ur ability to cure infections that were once considered benign is already damaged." (O'Neill 2016: 10). The danger of this becomes particularly clear when considering not only the possibility and impact of a global spread of bacterial infections, but also the speed with which this can happen in a globalised world connected by fast and almost non-stop traffic by land, sea, and air. Of course, bacterial spread depends on various factors, such as the respective specificity, overall survivability, and potency of transmission, but despite this, in the worst case such a spread could become devastating for humankind, as e.g., the plague pandemics demonstrate throughout history. Even without any major hotspots of bacterial outbursts, it is estimated that there are currently more than 670,000 infections with antibiotic-resistant bacteria per year in the European Union alone, resulting in roughly 33,000 deaths. Globally, untreatable bacterial infections account for not 700,000 deaths (not: infections) annually (cf. WHO Regional Office for Europe/European Centre for Disease Prevention and Control 2022: xiv; Antão/Wagner-Ahlfs 2018: 501, Davies et al. 2013: xii). According to recent findings by the Antimicrobial Resistance Collaborators (2022: 629, 639), the problem is even bigger, with 4.95 million deaths worldwide associated with antibiotic resistance, of which 1.27 million are directly caused by antibiotic-resistant bacteria. So although there are some significant geographical differences, with sub-Saharan Africa and South Asia currently most threatened by antibiotic-resistant bacteria, antibiotic resistance is a health problem of global proportions.

Against this background, it becomes evident that cautionary or alarming statements according to which "AMR is a looming threat to the health of millions of people worldwide" (WHO Regional Office for Europe/European Centre for Disease Prevention and Control 2022: xii) do not describe an apocalyptic scenario of a distant future. After all, humanity is already in midst of an antibiotic crisis. As Friedman et al. (2016: 421) remind us,

"resistance and MDR (multiple drug resistance, the authors) bacteria have spread and the negative impacts of antibiotic resistance have become more apparent" for decades. Despite the fact that the problem continues to grow, however, the danger of antibiotic is still commonly underestimated. As problematic as current developments may be, they are "only the tip of the iceberg" (Davies et al. 2013: 36), as the dangers of antibiotic resistance that lie ahead are even bigger. The way we try to counteract, reduce, or prevent already existing antibiotic resistance today has enormous impacts both on current use of antibiotics but also on the future. This is the case for those alive today as well as the generations yet to be born. This is because antibiotic resistance and its effects are somewhat comparable "to a slow-motion car crash: sadly, it is one that has already started" (O'Neill 2016: 71) and that cannot be prevented anymore. In particular, this is due to the irreversible failures and omissions that have occurred to date. Historical and current overuse, misuse, and abuse of antibiotics, as well as negligence of investment, research, and development of new antibiotics or adequate alternatives have put future generations at risk of losing effective means to treat bacterial infections.

> Like the climate crisis, antibiotic resistance is developing day by day beyond our collective perception. (...) Its effects are somewhat comparable to a slow-motion car crash: sadly, it is one that has already started.

On its current path, humanity is heading for a future escalation of the problems described, which will not only lead to poorer health care and an increase in the number of deaths, but also to severe economic consequences, as "[t]he impact of AMR on economic growth will result in a pronounced increase in extreme poverty." (World Bank 2017: 22) One of the commonly cited prognoses "estimate[s] that by 2050, 10 million lives a year and a cumulative 100 trillion USD of economic output are at risk due to the rise of drug-resistant infections" (O'Neill 2016: 4 and 12). This very prominent projection must be taken with caution, especially because of its rather speculative nature due to the opaque methodology (Kraker et al. 2016; Foreman et al. 2018: 2085; O'Neill 2014). However, it cannot be dismissed entirely. One may reasonably disagree about the extent of the problem, but its current trajectory is crystal clear: Humanity is not putting enough effort into addressing the problem of antibiotic resistance and averting the scenario of a post-antibiotic era (Engström 2021: 21). Since our current handling of antibiotics and antibiotic resistance have a very significant impact on the future, this becomes not only a long-term issue, but also a question of intergenerational justice. For as medically and morally defensible as some of our current antibiotic use may be, it (re)imposes the extreme vulnerability to bacterial infections that has plagued humanity for most of its existence on future generations. The danger is imminent, because „if we allow resistance to increase, in a few decades we may start dying from the most commonplace of ailments that can today be treated easily." (Davies et al. 2013: x)

This forecast underpins the World Health Organization's (2020) urgent and point-blank reminder that "[w]ithout urgent action, we are heading for a post-antibiotic era, in which common infections and minor injuries can once again kill." Notwithstanding the limitations that humanity already faces, if this post-antibiotic scenario becomes a reality, humankind will no longer be able to treat bacterial infections as it can today. As a result, future generations may no longer be able to benefit from the medical-pharmaceutical achievements we have come to know, and indeed face an existential threat. With this in mind, the task ahead seems to be quite clear: Possible solutions for tackling antibiotic resistance are needed.

The possibility of a way out and its obstacles

Although the rampant antibiotic crisis is a serious problem, it does not necessarily have to reach catastrophic proportions. A closer examination reveals a whole spectrum of possible ways that humanity could try tackle antibiotic resistance. Those include (a) novel drugs, (b) alternative treatments, (c) improvements in diagnostics, (d) a reduction in irrational use, (e) a reduction in general use, (f) education on antibiotic resistance, and (g) preventive measures to prevent bacterial infection. Subsequently, all of these possibilities need to be discussed in order to assess to what extent they could be key factors – individually and in combination – to prevent the worst-case scenario of a post-antibiotic era.

(a) The first possible response to the antibiotic crisis is to research, develop, and disseminate new drugs. However, there have been no significant innovations in this area in recent decades. Ever since the so-called 'golden age' of antibiotics, roughly dating to the middle of the last century, there is a serious slowdown in research and development and "[s]ince the 1980s, newly marketed antibiotics were either modifications or improvements of known molecules." (Iskandar et al. 2022: 1; cf. Kwon/Powderly 2021: 471. Friedman et al. 2016: 421). Whatever the reasons for this decline – scientific challenges on the matter itself, a lack of economic stimuli, or something completely different – may ultimately be, "[w]orldwide, the antibiotic development pipeline has all but dried up" (Dutescu/Hillier 2021: 416) and such omissions cannot simply be made up for. This is mainly due to long development periods as "[i]t typically takes 10 to 15 years to develop an antibiotic through regulatory approval." (Kwon/Powderly 2021: 471). Of course, antibiotic development must not take that long necessarily and it might well be that, analogous to the development of vaccines during the SARS-CoV-2-pandemic, the combination of a societal need and an enormous economic and time investment could accelerate this process. Despite this possibility, one must always bear in mind that new antibiotics are ultimately only an interim solution, as the development of new resistances is very likely and "[t]he race between AMR and antibiotic discovery shall continue" (Iskandar et al. 2022: 28).

(b) In the face of this constant chase, it is worth exploring alternative therapies. Vaccines or bacteriophages are amongst the better-known options that might prove effective in offering protection against dangerous or even lethal bacterial infections (cf. Hutchings et al 2019: 78; Dyar et al. 2017: 795). Furthermore, there may be supplementary drugs or therapies making use of experimental evolution (Jansen 2013). Here, it might be possible to actively exploit the evolutionary process of the bacteria for a more refined, future treatment. However, as innovative as such approaches may be, their practical applicability is still uncertain at present and requires further research.

(c) Another and already foreseeable way in which scientific-technological progress could contribute to solving the problem outlined is an improvement in diagnostics of bacterial infection as "it is likely that in the near future the immediate identification of pathogens through rapid whole-genome sequencing and other technologies will cut the time it takes to diagnose a microbial infection." (Davies et al. 2013: 53). Improvements in diagnostic procedures will help in choosing the most suitable therapy as fast

as possible. Especially in life-and-death situations, current methods often take too long, forcing doctors to treat bacterial infections as broadly as possible or rely solely on their best guess (cf. e.g., Davies et al. 2013: 51). However, for an optimally tailored therapy, knowledge of the exact pathogen is required. Otherwise, the right medication as well as the assessment of the optimal treatment duration, the possible change and specification of therapy or the administration of the drugs cannot be guaranteed. Hence, unlike precise diagnostics, suboptimal diagnostics is not only coupled with lots of uncertainty, but also often leads to inappropriate medication and the the doctors resort to broad-spectrum antibiotics due to a lack of knowledge about the specific infection.

(d) Exactly such handling is part of the so-called *irrational* use of antibiotics, as opposed to a *rational* one in which "patients receive the appropriate medicines, in doses that meet their own individual requirements, for an adequate period of time, and at the lowest cost both to them and their community." (World Health Organization 2004: 75). Irrational use of antibiotics is widespread and there are various ways to address it, ranging from an introduction of quota regulations or taxation to legal restrictions on accessibility or intended use. Ultimately, the point of all this is to make it more difficult to sell and purchase antibiotics by strict requirements for prescriptions and according monitoring processes (cf. Davies et al. 2013: 65-66). However, the most prominent means to prevent irrational use of antibiotics seems to be so-called Antibiotic-Stewardship-programs, which promote "both the appropriate use of antimicrobials when they are indicated, as well as avoiding unnecessary use" (Dyar et al 2017: 794). Although there is still room for improvement, especially in terms of global coverage, the results of these programmes are remarkable. As the European Centre for Disease Prevention and Control, for example, has been able to show, such programmes are significantly associated with a lower incidence of antibiotic resistance. Accordingly, institutions such as the World Health Organization pledge to expand them, because they have not yet been (sustainably) established in many places and progress in this regard does only come in small and slow steps (cf. WHO Regional Office for Europe/European Centre for Disease Prevention and Control 2022: xii-xiv).

For an optimally tailored therapy, knowledge of the exact pathogen is required. Otherwise, the right medication as well as the assessment of the optimal treatment duration, the possible change and specification of therapy or the administration of the drugs cannot be guaranteed. Hence, suboptimal diagnostics often leads to inappropriate medication and the doctors resort to broad-spectrum antibiotics due to a lack of knowledge about the specific infection.

(e) Ultimately, such programmes help "minimising the use of antibiotics when they are not necessary to improve human health" (Antimicrobial Resistance Collaborators 2022: 649). However, for this to actually succeed, *all* antibiotic consumption must be reduced, and this includes proper usage of antibiotics. Such a reduction can by no means be limited to applications for humans, but must also include other uses, such as agricultural use in animal husbandry. The reasoning behind this is not only a more thoughtful general use, but also the possible "[s]pread and cross-transmission of antimicrobial-resistant microorganisms between humans, between animals, and between humans and animals and the environment." (European Centre for Disease Prevention and Control 2008). Although there remain some uncertainties about those interactions in detail, such as questions of causality (cf. e.g., Antimicrobial Resistance Collaborators 2022: 649), considering them could prove proactive in delaying or stopping the development of resistant bacteria. An overall more frugal use of antibiotics, where appropriate, might help prevent further harm to the public good of antibiotics.

(f) While this certainly could be associated with unpleasant experiences, e.g. in the form of longer rest periods, many bacterial infections can be cured without the use of antibiotics and without actual risk to patients. Education on this topic is key, as it could improve the antibiotic knowledge of practitioners, but especially of patients. Because, as of now, "[c]onsumers have positive attitudes towards antibiotics, but paradoxically […] poor knowledge about these drugs and diseases." (Merrett et al. 2016: 4). People are often unaware of the negative side effects of antibiotics as well as the basic mechanisms of these drugs, especially that not every antibiotic is suitable to treat every bacterial infection, but also the fact that we all contribute to the increasing resistance. Education could not only help to stop the demand for and granting of antibiotics when not medically indicated and clarify misconceptions, such as a benefit for colds or flu (cf. Davies et al. 2013: 48, 50), but also increase compliance so that treatment instructions are strictly adhered to in situations where antibiotics are necessary. Currently, patients regularly intervene in therapies by, for example, discontinuing medication prematurely, which, contrary to popular belief, is a major problem (cf. Antão/Wagner-Ahlfs 2018: 500; Davies et al. 2013: 26) regarding antibiotic resistance, or by storing and reusing drugs without consultations.

But as important as the aforementioned possibilities are, the first step to address antibiotic resistance and a post-antibiotic era is anything but high-tech: "Minimizing the need for antibiotics through preventive health care and improved sanitation, housing, and access to clean water is achievable, as is ensuring that the right antibiotic is available and given at the appropriate dose for the appropriate duration." (Palmer 2022: xi). Especially when it comes to patient health, stopping the spread of bacteria and sparing people from potentially deadly infections is a top priority. Measures to achieve this include not only social distancing and quarantining of infected individuals, but also simple aspects of personal hygiene that reduce or prevent transmission. This holds particularly true for proper hand hygiene, which is practiced by only a fraction of people (cf. e.g., Davies et al. 2013: 47).

Although the rampant antibiotic crisis is a serious problem, it does not necessarily have to reach catastrophic proportions. A closer examination reveals a whole spectrum of possible ways that humanity could try tackle antibiotic resistance.

Especially the last-mentioned aspects may appear very basic, but they are not only highly effective and sustainable, but also seem to be the most realistically implementable. In sum, there are several possible ways to address antibiotic resistance, but the issue's high complexity requires "concerted efforts of microbiologists, ecologists, health care specialists, educationalists, policy makers, legislative bodies, agricultural and pharmaceutical industry workers, and the public to deal with." (Aminov 2010: 3). Thus, if we agree on the general guideline of ensuring humanity's access to antibacterial treatment in the future, this will require a broad range of actions and collective efforts by virtually everyone, as non-participation will hinder the necessary global endeavor. At the same time, however, these efforts must be adapted to specific regional

or geographic needs, as this may require better hygiene or sanitation in some areas, reduced use of antibiotics in animal husbandry in others, or simply better medical training (cf. Palmer 2022: x, Antimicrobial Resistance Collaborators 2022: 649, Davies et al. 2013: 70).

Conclusion: The (un)avoidable era of deadly bacteria upon us?

Overall, it is not impossible to avert the grim post-antibiotic era in which millions of people die each year from untreatable bacterial infections that scientists and organisations like the World Health Organization keep warning the global community about. Therefore, *Combating Antimicrobial Resistance and Protecting the Miracle of Modern Medicine* (National Academies of Sciences, Engineering, and Medicine 2022) is not a lost cause. At present, however, success is unlikely, and realistically, the goal of the initiatives taken can only be to reduce or at least slow down the problems at hand, not to completely avert the dangers outlined. Given the problem's current scope, it is not a question of whether antibiotic resistance is going to hit humanity, but only of how hard it will hit it and how much of an existential threat this poses. The past omissions in areas such as research and development, as well as the widespread failure to use antibiotics rationally, demonstrate a lack of political and societal commitment to a serious change in the way antibiotics are used. Furthermore, attempting to stop antibiotic resistance does come at a price – the most pressing one being the potential exposure of current patients to health risks in order to spare future ones.

Given the problem's extent, humankind does not only face the already difficult global and intergenerational challenge of providing „access to effective antimicrobials for all who need them, today and tomorrow" (Dyar et al. 2017: 797), but possibly more extreme hardships in form of "the subordination of present advantages to the long-term exigencies of the future." (Jonas 1984: 142). Antibiotic resistances and the horizon of a post-antibiotic era confront us with the question of whether it is morally imperative to restrict or withhold antibiotic therapies from patients *today* in certain situations, or even in general, in order to make them available to the same or other patients *in the future*. Addressing such questions, however, may lead to the realisation that intergenerational justice can only be achieved with a paradigm shift away from the idea of providing the best possible care for today's patients towards treatment that is sufficient to make sustainable antibiotic therapy more likely.

---

1 The authors are very grateful to Pascal Lemmer, Tizia Wendorff, Jan Rupp, and Hinrich Schulenburg as well as the reviewers for their remarks.

References

Aminov, Rustam I. (2010): A brief history of the antibiotic era: lessons learned and challenges for the future. In: Frontiers in Microbiology, 1 (134), 1-7.

Antão, Esther-Maria / Wagner-Ahlfs, Christian (2018): Antibiotikaresistenz. Eine gesellschaftliche Herausforderung. In: Bundesgesundheitsblatt 61, 499-506.

Antimicrobial Resistance Collaborators (2022): Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. In: The Lancet, 399 (10325), 629-655.

Davies, Sally / Grant, Jonathan / Catchpole, Mike (2013): The Drugs Don't Work: A Global threat. UK: Penguin.

Dutescu, Ilinca A. / Hillier, Sean A. (2021): Encouraging the Development of New Antibiotics: Are Financial Incentives the Right Way Forward? A Systematic Review and Case Study. In: Infection and drug resistance, 14, 415-434.

Dyar, Oliver J. / Huttner, Benedikt / Schouten, Jeroen et al. on behalf of ESGAP (2017): What is antimicrobial stewardship?. In: Clinical Microbiology and Infection, 23 (11), 793-798.

Engström, Alina (2021): Antimicrobial Resistance as a creeping crisis. In: Boin, Arjen/ Ekengren, Magnus / Rhinard, Mark (eds.): Understanding the Creeping Crisis. Cham: Palgrave Macmillan, 19-36.

European Centre for Disease Prevention and Control (2008): Factsheet for experts - Antimicrobial resistance. In: https://www.ecdc.europa.eu/en/antimicrobial-resistance/facts/factsheets/experts. Viewed 31 May 2022.

Fleming, Alexander (1964 [1945]): Penicillin. Nobel Lecture, December 11, 1945. In: Nobel Foundation (eds.): Nobel Lectures. Physiology or Medicine 1942–1962. Presentation Speeches and Laureates' Biographies. Amsterdam-London-New York: Elsevier Publishing Company, 83-93.

Fleming-Dutra, Katherine E. / Hersh, Adam L. / Shapiro, Daniel J. et al. (2016): Prevalence of inappropriate antibiotic prescriptions among US ambulatory care visits, 2010-2011. In: Jama, 315 (17), 1864-1873.

Foreman, Kyle J. / Marquez, Neal / Dolgert, Andrew et al. (2018): Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016–40 for 195 countries and territories. In: Global Health Metrics, 392 (10159), 2052-2090.

Friedman, Nadia D. / Temkin, Elizabeth / Carmeli, Yehuda (2016): The negative impact of antibiotic resistance. In: Clinical Microbiology and Infection, 22 (5), 416-422.

Gautam, Ashima (2022): Antimicrobial Resistance: The Next Probable Pandemic. In: JNMA J Nepal Med Assoc., 60 (246), 225-228.

Hansson, Kristofer / Brenthel, Adam (2022): Imagining a post-antibiotic era: a cultural analysis of crisis and antibiotic resistance. In: Medical Humanities, 48, 381–388.

Hutchings, Matthew I. / Truman, Andrew W. / Wilkinson, Barrie (2019): Antibiotics: past, present and future. In: Current Opinion in Microbiology, 51, 72-80.

Iskandar, Katia / Murugaiyan, Jayaseelan / Hammoudi, Dalal et al. (2022): Antibiotic Discovery and Resistance: The Chase and the Race. In: Antibiotics, 11 (2), 1-38.

Jansen, Gunther / Barbosa, Camilo / Schulenburg, Hinrich (2013): Experimental evolution as an efficient tool to dissect adaptive paths to antibiotic resistance. In: Drug Resistance Updates, 16 (6), 96-107.

Jonas, Hans (1984): The Imperative of Responsibility: In Search of an Ethics for the Technological Age. Chicago & London: The University of Chicago Press.

Kraker, Marlieke E. A. de / Stewardson, Andrew J. / Harbarth, Stephan (2016): Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050? In: PLoS Med 13 (11), 1-6.

Kwon, Jennie H. / Powderly, William G. (2021): The post-antibiotic era is here. In: Science, 373 (6554), 471.

Merrett, Gemma L. B. / Bloom, Gerald / Wilkinson, Annie et al. (2016): Towards the just and sustainable use of antibiotics. In: J of Pharm Policy and Pract, 9, 1-10.

Mitchell, Jessica / Cooke, Paul / Baral, Sushil et al. (2019): The values and principles underpinning community engagement approaches to tackling antimicrobial resistance (AMR). In: Global Health Action, 12, 1-12.

National Academies of Sciences, Engineering, and Medicine (2022): Combating Antimicrobial Resistance and Protecting the Miracle of Modern Medicine. Washington, DC: The National Academies Press.

Nationales Referenzzentrum für Surveillance von nosokomialen Infektionen (2017): Deutsche nationale Punkt-Prävalenzerhebung zu nosokomialen Infektionen und Antibiotika-Anwendung 2016. Abschlussbericht. https://www.nrz-hygiene.de/fileadmin/nrz/download/pps2016/PPS_2016_Abschlussbericht_20.07.2017.pdf. Viewed 31 May 2022.

O'Neill, Jim (2016): Tackling drug-resistant infections globally: final report and recommendations. The Review on Antimicrobial Resistance. https://amr-review.org/sites/default/files/160518_Final%20paper_with%20cover.pdf. Viewed 31 May 2022.

O'Neill, Jim (2014): Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. The Review on Antimicrobial Resistance. https://amr-review.org/sites/default/files/AMR%20Review%20Paper%20-%20Tackling%20a%20crisis%20for%20the%20health%20and%20wealth%20of%20nations_1.pdf, Viewed 31 May 2022.

Palmer, Guy H. (2022): Preface. In: National Academies of Sciences, Engineering, and Medicine (eds.): Combating Antimicrobial Resistance and Protecting the Miracle of Modern Medicine. Washington, DC: The National Academies Press, ix-xi.

Stolberg, Sheryl G. (1998): Superbugs. In: The New York Times Magazine, https://www.nytimes.com/1998/08/02/magazine/superbugs.html, Viewed 31 May 2022.

Vestergaard, Martin / Frees, Dorte / Ingmer, Hanne (2019): Antibiotic resistance and the MRSA problem. In: Microbiology Spectrum, 7 (2), 1-23.

World Bank (2017): Drug-Resistant Infections: A Threat to Our Economic Future. Washington, DC: World Bank.

World Health Organization (2020): Antibiotic resistance. In: https://www.who.int/news-room/fact-sheets/detail/antibiotic-resistance#:~:text=Bacteria%2C%20not%20humans%20or%20animals,hospital%20stays%2C%20and%20increased%20mortality, Viewed 31 May 2022.

World Health Organization (‽ 2004)‽ : The world medicines situation, 2nd ed. World Health Organization.

WHO Regional Office for Europe/European Centre for Disease Prevention and Control (2022): Antimicrobial resistance surveillance in Europe 2022 – 2020 data, Copenhagen: WHO Regional Office for Europe.

Dominik Koesling, M. A., is a research assistant at the Department of Biomedical Ethics and the Center for Ocean and Society at Kiel University. He studied communication science and sociology as well as theory of society in Jena. Besides his current research on sustainability, especially in medicine, and ocean health, he works on his PhD project on suffering in Critical Theory.

Email: dominik.koesling@iem.uni-kiel.de

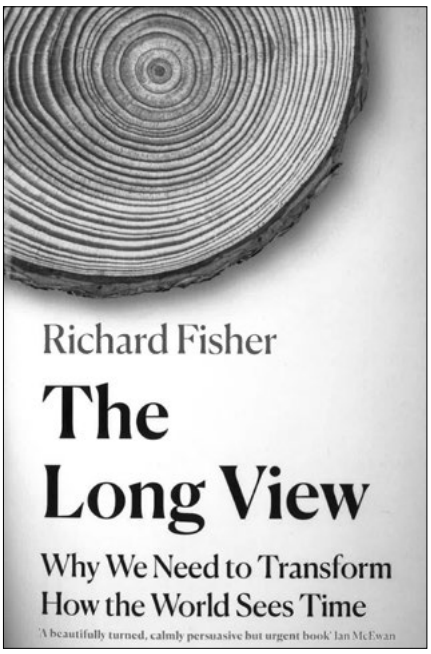
Prof. Dr. phil. Claudia Bozzaro is professor of Medical Ethics at Kiel University. She studied philosophy and art history in Freiburg and Paris. Her research focuses on ethical issues at the beginning and end of life, the analysis of normative concepts in medicine and ethical issues of sustainability in health care.

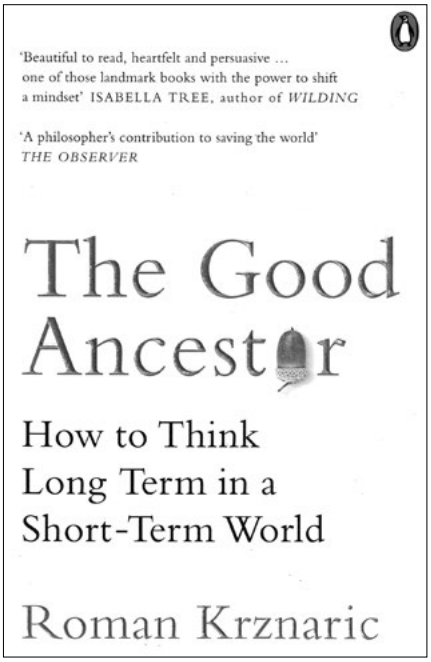
Email: claudia.bozzaro@iem.uni-kiel.de

Richard Fisher: The Long View: Why We Need to Transform How the World Sees Time
Roman Krznaric: The Good Ancestor: How to Think Long Term in a Short-Term World

Reviewed by Grace Clover

Trees have long populated the allegorical world of long-term thinking, taken to represent long-term growth and long-sightedness, aged wisdom, and stability in the natural world. In *The Good Ancestor* (2020), Roman Krznaric – a public philosopher and senior researcher at Oxford University's Centre for Eudaimonia and Human Flourishing – even proposes the phrase 'acorn brain' as a synonym for a long-term and self-reflective mindset, a sentiment which is upheld in Richard Fisher's work *The Long View* (2023). Fisher – a trained geologist turned journalist who writes on themes of time perception, long-term thinking, technology, and philosophy – takes this ecological metaphor one step further, noting the powerful imagery of stones as symbols of deep time and constancy. Krznaric also emphasises the symbiotic relationship we have with the natural world, declaring he felt an "an awe, a reverence, and an expanding sense of now" while looking at the felled trunk of an ancient sequoia tree (54). Beyond this however, Krznaric – the author multiple monographs on the themes of empathy and the power of ideas – seems more inspired by economic and political metaphors, writing that we treat the future "distant colonial outpost devoid of people" subject to ecological degradation and nuclear waste (7).





Beyond their semantic choices, there are a number of important similarities between the texts: Both authors note a range of existential risks – defined by Krznaric as "low-probability but high-impact events which could be caused by new technologies" (5) – which might be mitigated by the adoption of a long-term mindset, including threats from artificial intelligence, genetically engineered pandemics, and nuclear war. Though not defined as existential risks, they both also list slow-burning problems which are consistently ignored by those in power, including a failure to invest in preventative healthcare, child poverty, and ongoing racial injustice. Most importantly, both Krznaric and Fisher make the threat of ecological collapse the key thematic focus of their works.

Described in just broad brushstrokes, both authors aim to promote a global shift towards long-term thinking, mitigating the risk of civilisational collapse, and establishing a harmonious and sustainable relationship between humans and the natural world. Thus, despite some key divergences (which will be expanded upon later), the journey to becoming a 'good ancestor', and the search for the 'long view' are in many ways two mutually supportive offshoots of the same idea. Let's first explore Fisher's perspective.

Fisher's aims are twofold: To understand how blinded short-termism became integrated into our thinking and institutions and to suggest how we develop deeper temporal perspectives. He believes that short-termism is not innate, employing many examples of communities from outside the Western neo-liberal world with distinct kinds of long views. He defines the 'long view' as a temporal lens on the world, which allows us to see beyond short-term desires, sensationalism and immediate challenges and better understand our roles and responsibilities in the long term. It serves as an antidote to "time-blinkered" thinking (the pernicious and invisible spread of short-termism into all realms of society), allowing us to prepare for future risk, as well as being a source of hope and perspective in the present (11). "Time-blinkered" thinking is distinguished from being "present-minded", which refers to the deliberate prioritisation short-term goals as a response to emergencies in the present (78). Fisher portrays this allocation of public and private attention as a spectrum from "fast fires", such as upcoming elections and celebrity scandals, to "slow burns", such as growing inequality and climate change (77). Adopting a temporal lens would help us better understand the present as a

constant interaction between the past and several possible futures, rather than a transient moment with no consequences.

In *Part I: Time-Blinkered: The roots and causes of short-termism* Fisher offers a potted history of humanity's relationship with time, tracing the onset of our time-blinkered age into the 21st century. Fisher leads the reader through a series of vignettes, from the theatre of the second century BC to the industrial revolution, which saw an ever-increasing population being forced into synchronised working hours. He characterises our current age as one defined by "capitalism's unforgiving immediacy," embodied in the prioritisation of quarterly targets over long-term growth in the business sector (51). He also notes the shortening impact of election cycles on political thinking, quoting Jean-Claude Juncker, the former president of the European Commission, who wrote "we all know what to do, we just don't know how to get re-elected after we've done it" (75). Finally, Fisher describes (with some dismay about the state of his profession) the impact of the internet and journalism in promoting "short, loud and relentless" sensationalism (92).

Part II: A temporal state of mind offers an extensive psychological analysis of time-blinkered thinking. Fisher suggests that short-termism is culturally influenced; indeed, one of the key abilities which distinguishes *homo sapiens* from other non-human animals is our ability to comprehend the future and retain detailed memories of the past. Despite this, the future rarely features in our day-to-day experience, beyond planning for the coming days and weeks. Fisher introduces the reader to several psychological concepts which explain why this is might be the case. For example, "construal level theory" suggests that we perceive a psychological distance between ourselves and the future, meaning that we do not conceive of future life, needs, or suffering as concrete or even real (123). "Shifting baseline syndrome" refers to the acceptance of changing environments, which leads young generations to normalise the deteriorating environments that they inherit, without considering whether these conditions ought to be the norm (140). This also obscures us from appreciating positive improvements, such as moral and technological improvements, which become accepted into the standard societal framework. Further to this, Fisher analyses the relationship between language and the perception of time, encouraging his readers to replace phrases such as the 'distant' or 'far' future with alternatives such as the 'long' or 'deep' future. These alternatives linguistically bridge the gap between future and present and thus avoid the connotations of physical and psychological distance (166).

Part III: The long view: Expanding our perspectives of time is dedicated to describing a broad range of deeper time views, based fields as broad as science, religion, indigenous tradition, moral philosophy, and art. Fisher introduces the concept of "timefulness" coined by the geologist Marcia Bjornerud, which refers to being conscious of and drawing comfort in the age of the natural world around us, as seen in the rocks and earth we encounter daily. Referring to religion and spirituality, Fisher notes both the "continuity timeview" (192) – the transfer of tradition and belief across generations over hundreds of years – and the „transcendental timeview" rooted in a belief in eternity and heaven (203). Fisher also recounts the ethical basis for longtermism as proposed by theorists such as William MacAskill which sees the sheer quantity of humans who could live happy lives in the future (a far greater number than those alive today) as a moral basis to prioritise them at least as much as current people. Finally, Fisher introduces the reader to a number of artistic projects such as Katie Paterson's *Future Library* and David Nash's *Ash Dome* which reflect symbolically on our relationship with future generations and on the act of forward planning.

In the final chapter (285 – 297), Fisher summarises the benefits of deeper temporal perspective and achieving 'Deep Civilisation' in nine parts: 1) The long view is restorative. 2) The long view is a wayfinder. 3) The long view makes the present more meaningful. 4) The long view can be accessible to everyone. 5) The long view is democratic. 6) The long view can be politically unifying. 7) The long view leads to a healthier media diet. 8) The long view provides a clearer picture of progress. 9) The long view is an engine for hope.

It is clear from this summary that Fisher's primary focus is on the grounding, unifying, and positive impacts of a deeper time perspective. While this has important implications for the prioritisation of sustainable goals and policy which benefits "all people and living creatures in all time" (293), *The Long View* does not offer concrete policies for individuals or governments dealing with existential risks. Nor does Fisher recommend a single long view: While he seems most drawn to a long view rooted in the natural world and generational transfer, he also sees the benefits of religious timeviews, artistic gestures, and even many of the principles of the philosophical school called longtermism. What Fisher does offer is a holistic world view which can be adapted to each individual and society.

As I discussed, Fisher's background as a geologist is evident in his use of case studies and metaphors. He places particular emphasis on the profound impact of discovering tectonic plate movement on beliefs in biblical timeframes of the world. However, his style is also unmistakeably journalistic, showing a penchant for introducing academic case studies with anecdotal vignettes about the scholars involved. The impact of Fisher's breadth of expertise and sometimes anecdotal style is – for better and for worse – a monograph which reads like a beautiful patchwork quilt. He offers a very wide range of studies, from religious practise in Japan to small-town businesses in America. There is also certainly a cohesive structure and narrative which is satisfying to read. However, the reader sometimes runs the risk of skimming over concepts and case studies because of the sheer quantity of detail being offered. For example, Fisher frequently introduces psychological and economic terms in passing, which might appear superfluous to the reader not well acquainted with these academic fields.

Ultimately however, Fisher is highly successful in fulfilling his self-declared goals. The reader is left with a clear understanding of the development of short-termism in the past century. The detailed focus on the psychology of time-blinkered habits and short-termism is something that sets Fisher's monograph positively apart from other works on existential risk and long-termism. Overall, Fisher offers a personal but informative, engaging, diverse, and accessibility written book, with a clear structure and message.

A key divergence between Fisher's *The Long View* and Krznaric's *The Good Ancestor* is the extent the two authors critique the role of neo-liberal capitalist systems in enhancing the likelihood of existential risks. While Fisher is highly critical of the role of Western free-market capitalism – which he associates with quarterly reporting, short-term targets, and consumerism – he believes that capitalism can be reformed and cites new practices such as "conscious" or "inclusive capitalism" as potential ways forward (63). Normative theories for political change are, however, not Fisher's focus. Roman Krznaric on the other hand much more explicitly frames the journey to becoming a 'good ancestor' in terms of a

fundamental political, social, and economic structural change, seeing a central tension between the neo-liberal prioritisation of perpetual growth and the promotion of ecological civilisation.

Let's now explore Krznaric's perspective in more detail. Having introduced his thesis that we treat the future as a colonial outpost, Krznaric opens *Part One: The tug-of-war for time* by asking two key questions: First of all he questions how we can be good ancestors, drawing upon the words of Jonas Salk, the medical researcher who developed the first safe polio vaccine and left it unpatented for global use. Secondly, he asks how we can unlock and fully harness our acorn brains. These two questions demonstrate conceptual similarities between Krznaric and Fisher's works, as they seek to unlock the part of the human brain which can think far into the future, so as to leave a liveable, regenerative ecosystem and sustainable institutions for generations to come.

In *Part Two: Six ways to think long*, Krznaric proposes alternative ways to conceive of our ancestral relationship with future generations. He dedicates one chapter to each of these six perspectives, which he names: Deep-Time Humanity, Legacy Mindset, Intergenerational Justice, Cathedral Thinking, Holistic Forecasting, Transcendental Goal.

Under the banner of 'Deep-Time Humanity' Krznaric encourages us to acknowledge our own insignificance as a species. Compared to the age of the earth, homo sapiens have just around for just seconds. In accepting this, we can re-connect with a deeper sense of time which allows us to break free from the tyranny of the clock and acceleration of life and re-join the cyclical rhythms of the natural world.

Second, he proposes that modern society age should re-connect with the presence of death. In removing the societal taboo surrounding death, we would receive a "death nudge" which – rather than being a negative force – can act as a positive reminder of posterity (59). This legacy mindset encourages use to think of a communal legacy for many generations to come, beyond the direct inheritance we might leave for our children.

Third, Krznaric describes a sense of intergenerational justice, which instead of fostering empathy and connection between generations, encourages a sense of moral responsibility and justice between those alive today and those yet to be born. On this theme, Krznaric details the moral violation presented by the economic theory of discounting: the mathematical equation which discredits the value of measures to aid future people at increasing rates away from the present. This practise is used by governments and businesses alike to justify avoiding projects with long-term benefits if they have high upfront costs. Krznaric also touches upon various moral philosophical arguments for prioritising future generations on the basis of intergenerational justice, including Derek Parfit's (1942 – 2017) suggestion that people should be treated as having equal worth, regardless of when they are born.

Fourth, in the chapter named 'Cathedral Thinking', Krznaric lists a number of projects in fields as diverse as architecture, public policy, cultural projects, and social movements which either show a deeper reflection on our relationship with time or have showed long-term planning. Drawing from the example of the Victorian reform of London's sewer system, Krznaric explicitly problematises those in political or financial power being insulated from the impacts of crises they themselves often create. He demands that they a show a sense of urgency long before such problems begin to impact them personally or existential risk scenarios ensue.

Fifth, Krznaric introduces 'holistic forecasting' as a means of reaching a deep-time humanity. This perspective involves the acceptance of uncertainty as an inherent part of thinking about the future, while still looking for long-term trends so as to plan for multiple different future scenarios. Integral to our forecasting about civilisational collapse, Krznaric argues, is the S-Curve or sigmoid curve. This model has been used to trace the downfall of many historic collapses and shows that civilisations standardly reach an inflection point where the rate of growth slows, followed by a period of maturity, and then decline. Such a trend significantly challenges the Enlightenment assumption that progress can and should be pursued indefinitely – a criticism which is foundational to Krznaric's work. Without a full transformation of our global structures and consumer culture, we will be unable to mitigate the devasting impacts of dramatic decline, allowing issues such as drought, extreme weather, and food insecurity to become even more present in the future.

Finally, Krznaric promotes the value of a transcendental goal in governing our relationship with the future. Rejecting the idealisation of perpetual progress and dreams of techno-liberation, such as space colonisation and transhumanism, as solutions to impending existential risk, Krznaric promotes a goal he calls 'one-planet thriving'. This is defined as a society in which we live "within the life-supporting systems of the natural world", respecting its boundaries and capacities (156). This involves acknowledging that humans are not separate from nature but are actually "an interdependent part of the living planetary whole" (158).

In *Part Three: Bringing on the time rebellion,* Krznaric introduces a number of 'time rebels' who have pushed against the short-termism of our society. Drawing inspiration from these rebels, Krznaric proposes concrete political, financial, cultural structural changes which could guide us to becoming good ancestors.

To begin, Krznaric describes a system he calls 'Deep Democracy', the structural political counterpart to the psychological time perspectives he proposes in *Part Two* and that Fisher proposes in *The Long View.* Much like Fisher, Krznaric points to election cycles, vested interested of elite groups, and the pressures of social media cycles as causes of political presentism. Further to Fisher however, Krznaric details the structures which he argues prevent us reaching political longtermism, problematising the lack of international cooperation between nation states, and the fact that future generations are completely disenfranchised in representative democracy. In response to these issues, he proposes four design principles which could guide us towards deep democracy: namely, 1) guardians of the future 2) citizens' assemblies 3) intergenerational rights and 4) self-governing city-states.

Firstly, he proposes that 'guardians of the future' should be appointed in national and eventually international bodies with the specific role of representing disenfranchised young and unborn people. He refers to the example of the Future Generations Commissioner for Wales, who has the role of reviewing all policy against specific sustainability criteria.

However, such appointments are just the first step, Krznaric argues. To keep these officials and institutions in check, and to ensure diversity and inclusivity in representation, he writes that citizens' assemblies – randomly selected from all citizens aged 12 and upwards – should support the work of 'guardians'. Assemblies would have a defined enforcement power and meet with experts to discuss themes related to being a 'good ancestor'.

The enshrinement of intergenerational rights in international law, Krznaric argues, would also act as a guiding mark for citizens' assemblies and help them hold governments and organisations to account. Specifically, Krznaric strongly advises establishing "eco-

cide" as crime in international law; that is, "extensive destruction of the natural living world" (186) which would transform our perspective on the world, seeing it as a living being, rather than private property. This perspective change is already implemented in Bolivia with the Law of the Rights of Mother Earth, which gives nature equal rights to humans.

Finally, Krznaric notes the power of cities and city states to go above and beyond national and international law to transform their environment into a sustainable, citizen-friendly urban spaces. Alongside a 'deep democracy', Krznaric proposes the need for a 'regenerative economy' to replace the current system of speculative capitalism. Such a global economy would meet "human needs within the biophysical means of the planet, generation after generation", creating an ecological civilisation in balance with the regenerative systems of the Earth (195). He cites the study *The Limits to Growth,* published by Donella and Dennis Meadows in 1972, which shows that if the current growth in population, industrialisation, and resource use continues, the limits to growth will be reached in the next hundred years, leading to a fundamental decline in human welfare (199). Krznaric also refers to the ecological economist Herman Daly who notes that the "economy is a subsystem of the larger biosphere that is finite and not growing in size, which means that the economy's material throughput cannot keep growing forever" (199). To this end, Krznaric proposes five economic measures, including the taxation of stocks based on the amount of time they are held for, the promotion of a circular and localised economy, and the democratisation of (renewable) energy. This third measure would avoid renewable energy becoming held by a small elite, helping mitigate the impacts of a 'climate apartheid' – a situation in which the rich would be able to insulate themselves from the impacts of climate change. Finally, Krznaric – as influenced by the environmentalist George Mombiot – proposes a focus on rewilding rather than conservation, allowing ecosystems to return to a place of wildness which can sustain itself, rather than conserving the current depleted ecological state. This would create new natural carbon solutions and prevent biodiversity loss.

Finally, Krznaric details the importance of a cultural transformation in supporting the systems of deep democracy and regenerative economy. Much like Fisher, Krznaric discusses artistic projects such as John Cage's *As Slow as Possible* and Kate Paterson's *Future Library* which thematise our relationship with time, noting the ability of art and literature to foster a shared identity with future generations.

One can only conclude that Krznaric is successful in fulfilling his self-declared goal to fill the 'intellectual vacuum' surrounding long-term thinking. He provides a conceptually sound, wide-ranging, academic, but empathetically argued text which not only defines long-termism but speaks volumes for its benefits. One could also argue that Krznaric is successful in bringing long-term thinking away from the academic and scientific margins where it resided in 2020, providing space for authors such as Fisher to write works on the topic for a slightly broader readership. Indeed, Fisher's monograph shows multiple points of influence from Krznaric's work, including many of the same cultural and political examples detailed particularly in the chapter 'Cathedral Thinking'. Whether Krznaric is successful in influencing a global transformation of mindsets and structures is unfortunately much harder to measure.

Structurally, there are moments in Krznaric's text which feel somewhat repetitive. For example, at three separate points in the text he raises the contentious question whether long-term planning is most likely to thrive under authoritarian regimes – once in relation to Ancient Japan, once in relation to modern China, and once as an introduction to his discussion of the Intergenerational Solidarity Index. While this question is very important to the debate surrounding political myopia and its solutions, it could have been answered (or in Krznaric's case, debunked) in one comprehensive section. Secondly, at times Krznaric appears to mention cultural projects in passing for the sake of it, without engaging as deeply as Fisher in their metaphorical weight. Accordingly, Krznaric's chapter 'Cultural Evolution' is less slightly less evocatively written than the two preceding chapters 'Deep Democracy' and 'Ecological Civilisation', and thus weakens the momentum being built towards the end of the text.

That said, this only further emphasises Krznaric's strength as a political philosopher, who is highly successful in making political theory accessible to his readers, while offering concrete suggestions for reform. An interesting point of comparison between Krznaric's work and the work of other political philosophers theorising about long-term perspectives, is his proposal of 100 years as a minimum threshold for long-term thinking. As Marina Moreno points out, in comparison to the strong longtermist proposals of scholars such as Hilary Greaves and William MacAskill, who include horizons of thousands if not billions of years, Krznaric's work might even be considered 'presentist'. Unlike Krznaric and Greaves/MacAskill, Fisher does not explicitly offer any defined suggestions of timeframes and promotes a long view which is just as much rooted in what we can learn from the past, as well in as looking forward.

In conclusion, I for one am more than persuaded by the arguments for deeper temporal perspectives proposed in *The Good Ancestor* and *The Long View.* Having reflected on 'long views', I am aware more than ever of the symbols of deep time all around us – be they in the trees and stones outside or in the art and culture we consume – and can see the positive psychological benefits of adopting a long view for current generations. An important second phase, however, is a wider cultural, societal, political, and economic transformation, which has the principles of a good ancestor at its heart. I can thus only encourage these two books to be read in tandem. The individual adoption of the long view can form a strong basis for the creation of a global society of good ancestors. Both Krznaric and Fisher open our ears to the voiceless majority of future generations and offer significant nourishment to the tree of long-term thinking.

---

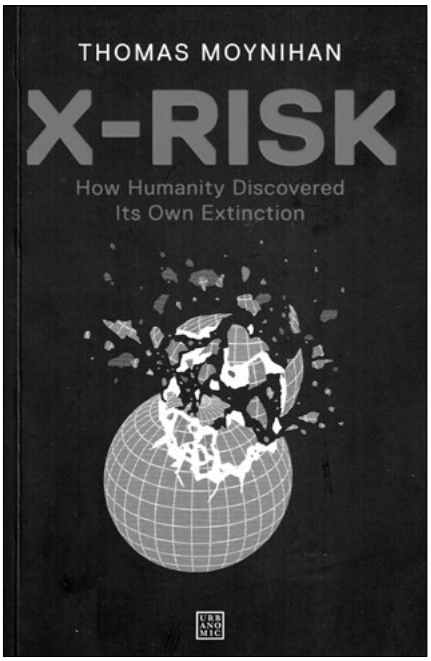
1 Moreno, Marina (2022): Does longtermism depend on questionable forms of aggregation? In: Intergenerational Justice Review 8 (1), p. 15.

Fisher, Richard (2023): The Long View: Why We Deed to Transform How the World Sees Time. London: Headline Publishing Group. 344 Pages. ISBN-13: 9781472285218. Price £25 (hardback).

Krznaric, Roman (2020): The Good Ancestor: How to Think Long Term in a Short-Term World. London: Penguin Random House. 324 Pages. ISBN: 9780753554517. Price £12.99 (paperback).

Thomas Moynihan: X-Risk: How Humanity Discovered its Own Extinction

Reviewed by Kritika Maheshwari



Technology experts are now claiming that superintelligent artificial intelligence, if realised, could pose an existential threat to humanity's long-term survival. The possibility that we might be putting humanity at risk of extinction has instantly spurred world leaders into action, with some insisting we put a pause on research and training of artificial systems. These recent events might perhaps suggest that humanity is finally waking up to the realisation that our future is anything but secure. However, as Thomas Moynihan argues in his recent book *X-Risk* (2020), this concern with existential threats has a long intellectual history, one that is important for understanding how and why we ought to care about humanity's continued survival into the future. Writing as a historian and philosopher of ideas, Moynihan carefully weaves together a complex account of how humanity first discovered the idea of its own extinction. By drawing on historical, literary, philosophical, and theological perspectives on the topic, Moynihan makes three claims along the way.

First, Moynihan makes the case for the novelty of humanity's extinction. He argues that the idea of human extinction has remained conceptually unavailable for the most part of our existence as a species. Next, he argues that the apocalyptic prophecies one reads about in religious and mythological texts are both conceptually and normatively distinct from the idea of human extinction. Whereas the thought of apocalypse offers a sense of an ending and is thus a conciliatory concept, the idea of our extinction anticipates the ending of sense and rationality and thus offers no consolation. Finally, Moynihan contends that fully grasping the prospect of our own extinction is not to be celebrated merely as a conceptual feat. Instead, we must recognise our ability to reason about humanity's extinction as a defining feature of modernity itself. Drawing upon philosophical ideas marshalled by Enlightenment thinkers like Immanuel Kant, Moynihan argues that our own rationality draws attention to the responsibility we have to ensure humanity never meets the disastrous fate of going out of existence.

Understanding how exactly the Enlightenment period succeeded in placing existential risk on humanity's conceptual map first requires a brief historical detour into ancient thought. In chapter 2 (*Cosmic Silences: Astrobiology)*, Moynihan emphasises the stronghold of a pre-Enlightenment philosophical assumption, namely the principle of plenitude which states that all legitimate possibilities in the world are realised. The idea of plenitude entails that should our species go extinct, the possibility of its return will eventually and inevitably be fulfilled. This plenitude-centred thinking dating back to ancient philosophers like Plato, Pliny, and Lucretius has the following upshot: It suggests that moral understanding and moral justification for humanity's extinction relies heavily on what we accept as the correct or appropriate metaphysical and scientific view of the world, all other things being equal. So, if it's true that humanity would reappear once extinct as a matter of necessity, then the question of whether causing or allowing humanity's extinction is morally wrong loses its significance.

Historically, then, the prominence of plenitude-centred thinking, together with the now frequently rejected suggestion that nature itself is imbued with value and justice, dismissed the case for even thinking about human extinction by rendering the very idea of extinction meaningless. This brief yet important insight into pre-Enlightenment thinking about our future, or rather the absence of it, is interesting. It stands in sharp contrast to our recent preoccupation with mitigating and strategising about different existential risks that face humanity today. This indicates that we have moved intellectually from an adherence to plenitude to an acknowledgement of the contingency of the conditions of human existence and the role of chance, which raises the question of how humanity came to acknowledge extinction as an issue worthy of its attention?

In chapters 3 (*Earth Systems: Geoscience*) and 4 (*Future Trajectories: Forecasting*), Moynihan examines how distinct fields of empirical science such as geosciences and actuarial sciences converged upon the idea that we hold the power to either push humanity to the brink of a precipice or use that power to preserve our long-term future. The intellectual shift towards this Enlightenment frame of thinking was marked by rejecting the otherwise widespread conflation of moral values with natural facts. This entailed a further radical shift in our thinking about our collective future, namely that it is not only open and uncertain, but also precarious. In chapter 3, Moynihan reviews how scientific studies of fossils as well as species mutability provided empirical evidence for the reality of species extinction. The Leibnizian idea that ours is the best possible world was soon questioned by the reality of pre-historic non-human extinctions, opening up doors to the possibility that humanity's continued existence is a mere accident. Moreover, evolutionary principles such as Louis Dollo's law of irreversibility reified the idea that even if plenitude is plausible, humanity's extinction would be irreversible insofar as organisms can never return to their former state even when placed in identical conditions to those in which they previously thrived.

In *Future Trajectories: Forecasting*, Moynihan offers another good example of how advancement in scientific theorising has strongly shaped our present understanding of extinction, both as a natural and a moral phenomenon. He turns his attention towards political arithmetic, or rather demographic thinking, which sowed the seeds for considering humanity as a object for objective investigation. Thinking about humanity as an aggregate may not initially strike us as an impressive feat because we are now so used to population-level thinking in matters of policy and political decision making. However, this was an achievement par excellence during the Enlightenment period, for it allowed us to conceive of humanity as a planetary collective. Combined with progressions in our mathematical understanding of risk, probability, and uncertainty, it was now also possible to have a quantitative grasp of existential threats humanity at a collective scale.

The Enlightenment period was thus successful in reinforcing the idea that our extinction would lead to the end of all value. The decoupling of fact from nature, the dismissal of plenitude, as well as empirical evidence for existential risk suggested two potential approaches to this dilemma: either we take nature's lack of inherent prudence and morality as an engineering problem in need of a fix or we dismiss any responsibility we may have towards preserving and protecting our collective future.

In chapter 5 (*Internal Contradictions: Omnicide*), Moynihan explores a range of views advocating for latter approach on three key bases: that perhaps our concern with human existence justifies a problematic kind of human exceptionalism, that perhaps living in the worst possible world full of suffering and sufferers is a live possibility, and that perhaps extinction is in fact the key to unleashing rather than curtailing our potential. Are the moral stakes involved in our extinction great enough to outweigh the harms of human exceptionalism, suffering, and curtailing our own potential?

Moynihan's project is not to settle these issues, but rather is to explain how we came to care about humanity's precariousness in the first place and to suggest why we must continue to care. In chapter 6 (*Physical Salvation: Vocation*), Moynihan argues that the answer is to be found in Kant's philosophy and in particular, in the idea that moral values are a question of self-legislation. In arguing against the idea that values are inherently imbued in nature, Kant argues that they are maxims that we elect to bind ourselves to and are thus our own responsibility. Hence, part of being a rational actor is to become concerned by the extinction of rationality, for it cannot exist otherwise. It is in this sense that as rational actors, we were bound to discover humanity's extinction through our ability to act and think rationally. Equally, we are responsible for caring and doing something about the existential risks we face. As such, dismissing existential risk on the basis of plenitude or on accounts of conflating fact with nature is simply incoherent with the bounds of Kantian thought. Moynihan's recasting of the origins and importance of existential thinking from this perspective is original and an important contribution to the project of developing a Kantian ethics of human extinction within a theoretical landscape that currently remains dominated by consequentialist theorisation. In doing so, Moynihan takes the first step towards providing an *explanation* for why it is rational for humanity to care, or rather, continue to care about its own extinction. However, it a separate question whether this explanation also provides us with a comprehensive justification of humanity's attempt to prevent its own demise.

For instance, let us suppose that humanity's future can be only protected by willing the end of all non-rational life on Earth. Within the constraints of an anthropocentric theory which is committed to the idea that rational nature alone has absolute and unconditional value, mitigating existential risks this way may not seem dismal. And yet, many would abhor the idea of preserving our rationality at the cost of sacrificing or destroying everything else we may value, such as beautiful landscapes, trees, and non-human animals. Similarly, what if avoiding humanity's complete self-annihilation required the self-annihilation or moral suicide on part of some rational agents? In which ways can Kantian ideas of perfect duties to the self as it applies both to individuals and humanity as a whole guide us? Such examples are merely intended to show that observing the moral significance of caring about humanity's extinction through Kantian lens raises new questions about *how* we ought to care. Moreover, it also raises questions for *whom* we ought to care for.

For example, let us consider the project of reconciling Kantian ethics of extinction with the prominent consequentialist thought that causing or risking our extinction is morally wrong as it blocks the added value of bringing future people into existence. As Moynihan notes, "to give up the fight to maximise value is to immorally submit to the environing forces of extinction, to the unjust fact that extinction and sterility is the cosmic tendency and the uphill struggle toward complexity the exception" (367). This, however, raises the question of whether and in what ways the Kantian injunction to respect the autonomy of actual persons rules out or alternatively includes potential people within the scope of its moral community, to whom we owe this concern. Again, the point here is not to dismiss Moynihan's claim that humanity's concern for its extinction is presupposed by the very nature of rational agency itself. Rather, it is to motivate further investigations into how far we can take this idea and apply them to concerns that occupy those interested in ethics of our long-term survival.

As Moynihan correctly notes, this Enlightenment-driven idea is still a work in progress – we are only now starting to uncover the full ramifications of humanity as historic collective project. This process remains incomplete both because we are far from achieving humanity's full potential, but also with regards to reifying the scope and the content of responsibility that rationality places on individuals for mitigating the existential risks that humanity faces. A few important questions should be raised in this context: What is our individual responsibility towards mitigating such risks? How does our individual responsibility fare against our collective responsibility as a rational species? Besides, what demands are placed by rationality onto the preservation of rationality itself? For instance, would humanity's long-term potential be preserved if human life were to be replaced not by superintelligent, but some kind of superrational artificial intelligence? In conclusion, Moynihan's book not only succeeds in capturing the historical landscape of humanity's extinction, it also manages to push the boundaries of philosophical inquiry by raising new and important questions worthy of further research.

Moynihan, Thomas (2020): X-Risk. How Humanity Discovered its own Extinction. London: Urbanomic Media. 472 Pages. ISBN: 9781913029845. Price €25 (paperback).

Imprint

Publisher: The Foundation for the Rights of Future Generations (Stiftung für die Rechte zukünftiger Generationen) and The Intergenerational Foundation
Permanent Editor: Jörg Tremmel
Co-Editors for IGJR 2-2022:
Grace Clover, Markus Rutsche
Layout: gänserich grafik,
Friedrich-Ebert-Straße 16, 14467 Potsdam
Print: Kuhn Copyshop & Mediacenter,
Nauklerstraße 37a, 72074 Tübingen
Website: igjr.org

Editorial offices:
Foundation for the Rights of Future Generations (Stiftung für die Rechte zukünftiger Generationen)
Mannspergerstraße 29,
70619 Stuttgart, Germany
Tel.: +49(0)711 28052777
E-Mail: kontakt@srzg.de
Website: intergenerationaljustice.org

The Intergenerational Foundation
19 Half Moon Lane
Herne Hill London SE24 9JU
United Kingdom
Email: info@if.org.uk
Website: if.org.uk