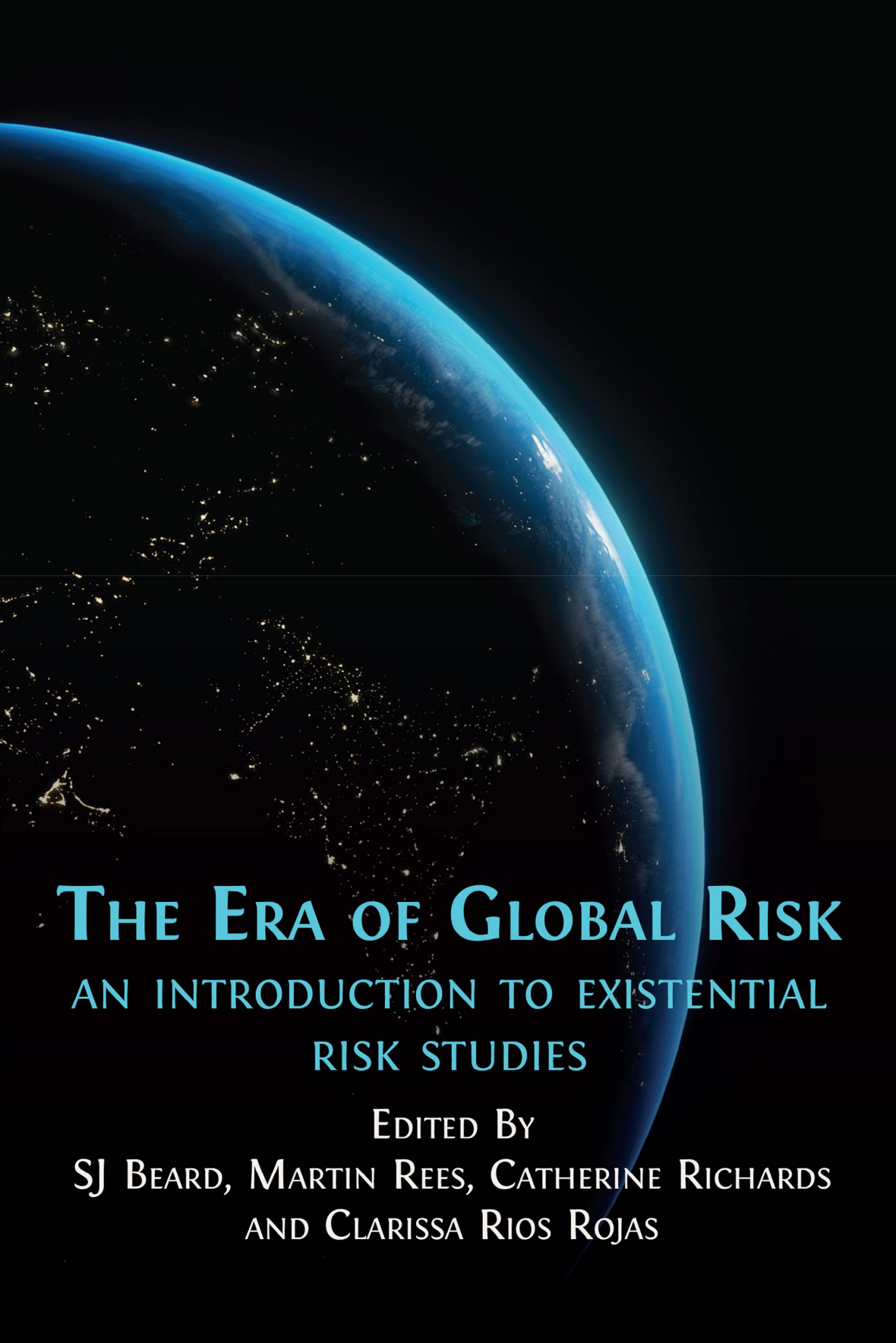

The Era of Global Risk

An Introduction to Existential Risk Studies

Edited By
SJ Beard, Martin Rees, Catherine Richards and Clarissa Rios Rojas

©2023 SJ Beard, Martin Rees, Catherine Richards, and Clarissa Rios Rojas. Copyright of individual chapters is maintained by the chapters' authors

This work is licensed under an Attribution-NonCommercial 4.0 International (CC BY-NC 4.0). This license allows you to share, copy, distribute and transmit the text; to adapt the text for non-commercial purposes of the text providing attribution is made to the authors (but not in any way that suggests that they endorse you or your use of the work). Attribution should include the following information:

SJ Beard, Martin Rees, Catherine Richards and Clarissa Rios Rojas (eds), *The Era of Global Risk: An Introduction to Existential Risk Studies*. Cambridge, UK: Open Book Publishers, 2023, https://doi.org/10.11647/OBP.0336

Copyright and permissions for the reuse of many of the images included in this publication differ from the above. This information is provided in the captions and in the list of illustrations. Every effort has been made to identify and contact copyright holders and any omission or error will be corrected if notification is made to the publisher.

Further details about CC BY-NC licenses are available at http://creativecommons.org/licenses/by-nc/4.0/

All external links were active at the time of publication unless otherwise stated and have been archived via the Internet Archive Wayback Machine at https://archive.org/web

Digital material and resources associated with this volume are available at https://doi.org/10.11647/OBP.0336#resources

ISBN Paperback: 978-1-80064-786-2
ISBN Hardback: 978-1-80064-787-9
ISBN Digital (PDF): 978-1-80064-788-6
ISBN Digital ebook (epub): 978-1-80064-789-3
ISBN XML: 978-1-80064-791-6
ISBN HTML: 978-1-80064-792-3
DOI: 10.11647/OBP.0336

Cover image: Anirudh, *Our Planet* (October 14, 2021), https://unsplash.com/photos/Xu4Pz7GI9JY. Cover design by Jeevanjot Kaur Nagpal.

1. A Brief History of Existential Risk and the People Who Worked to Mitigate It

SJ Beard and Rachel Bronson

Despite garnering significant academic, political, and public attention, the existential risks posed by nuclear weapons, environmental breakdown, and disruptive technologies continue to threaten human survival, and we may now be in a more perilous position than at any other time in history. For over 75 years we have been dragooned into unacceptable gambles by political and technological forces, and were lucky to survive thus far. However, this story has not just been about luck. Since the emergence of such risks, curious, compassionate, and courageous people (including many scientists) have sought to understand these terrifying forces, bring them to wider public attention, and work with every tool at their disposal to prevent human extinction and the collapse of civilisation. In this chapter, we seek to revisit the ups and downs of this perilous journey, using as our guide the shifting time of the Doomsday Clock, and to briefly tell the story of some of the people and organisations who sought to guide us safely through it. Understanding this history is vital, not only because these risks remain pressing, but also because it offers an opportunity for those currently working to reduce existential risk (and especially those in the nascent academic field of Existential Risk Studies) to learn from the successes and failures of the past. In particular, we show the importance of reinforcing key messages about risks and how to manage them, modelling extreme scenarios to understand them better, managing the pace of scientific

© 2023 SJ Beard and Rachel Bronson, CC BY-NC 4.0 https://doi.org/10.11647/OBP.0336.01

research, and placing its findings in the public domain. If we can learn these lessons and apply them rigorously, then history shows we can turn back the hands of the Doomsday Clock, and ensure that our future is no longer a hostage to our fortune.

The origins of our understanding of Existential Risk

People have speculated about the 'the end of the world' since the dawn of history—indeed, the oldest story that has been passed down may well be the Mesopotamian deluge myth, which tells of a flood that wiped out all but a few humans, and is familiar to most in the west through the biblical story of Noah.[1] However, such eschatological speculation has largely been bound up with religious beliefs and invariably ends with humanity continuing on Earth, in the afterlife, or via an eternal cosmic cycle of rebirth. Furthermore, as Martin Rees argued in his book *Our Final Century*:

> Throughout most of human history the worst disasters have been inflicted by environmental forces—floods, earthquakes, volcanos, and hurricanes—and by pestilence. But the greatest catastrophes of the twentieth century were directly induced by human agency.

Virtually everyone alive today is familiar (to some degree) with the anthropogenic risks that threaten global disaster, like nuclear war, climate change, and risks from disruptive technologies such as artificial intelligence (AI) and biotech. These risks are both naturalistic—in the sense that we understand how they could happen within the laws of nature—and absolute, in the sense that there may be no reprieve for humanity and no afterlife.

In fact, the very idea that humanity was vulnerable to going extinct in this way may be a relatively recent invention. It arose from the scientific discovery of prehistoric fossils and its implication of a 'deep past' during which evidence of extinctions was incontrovertible, our growing awareness that there is no great difference in kind between humans and other species, the spread of secular atheism, and the acceleration of social, scientific, and technological change.[2]

Perhaps the first group to fully express this change in thinking were authors of speculative fiction. For instance, Mary Shelley, one of the

founders of science fiction, wrote *The Last Man* in 1826,[3] which tells the story of Lionel, who witnesses the death of all other humans in the last few decades of the 21st century from a series of apocalyptic events, most notably a worldwide plague. The first mention of human extinction being caused by self-improving machines comes from Samuel Butler's 1863 *Darwin Among the Machines*, later reprinted as part of his novel *Erewhon*.[4, 5] Similarly, the first discussion of the existential risk posed by atomic weaponry is arguably found in H.G. Wells's *The World Set Free*,[6] while more recently, sci-fi authors have been among the first to explore how humanity may bring about its own demise through our harmful influence on planet Earth.[7]

Such writers not only captured the popular imagination but also directly influenced academic research. H.G. Wells's 1901 book *Anticipations of the Reaction of Mechanical and Scientific Progress Upon Human Life and Thought*,[8] for instance, is a foundational text for the academic discipline of Futures Studies, a subject of vital importance to our understanding of existential risk.[9] Wells also wrote at least two non-fiction, if not entirely serious, essays on the risk of human extinction, 'On Extinction' and 'The Extinction of Man',[10] while the first book-length non-fiction work to classify and explore the entire range of possible existential catastrophes was Isaac Asimov's *A Choice of Catastrophes*.[11]

Yet, while they have a vital role in raising awareness and exploring different futures, science-fiction authors are often the first to argue for the importance of the hard science on which they draw. Thus, the true foundation of the study of existential risk belongs to a group of pioneering scientists and philosophers working during and shortly after World War II, who became concerned about several overlapping trends and developments with the potential to significantly threaten humanity's future.

The threat of nuclear weapons

Worries about the risk of a global catastrophe first gained major scientific attention after World War II, with widespread concern about nuclear weapons and their potential to wipe humanity off the face of the Earth. The speed and violence with which nuclear technology evolved was breathtaking, even to those closely involved in its development.

As early as 1939, world-renowned scientists Albert Einstein and Leo Szilard penned a letter to US President Franklin D. Roosevelt about a breakthrough in nuclear technology that was so powerful, and could have such tremendous battlefield consequences, that a single nuclear bomb, "carried by boat and exploded in a port, might very well destroy the whole port", a possibility seen as too significant for the President to ignore. A mere six years later, one such bomb was used to destroy an entire city and its population, followed by another one. A few years after that, nuclear arsenals were capable of destroying civilisation as we know it.

The first scientific concern that nuclear weapons might have the potential to end humanity as a whole appears to have come from scientists involved in the first nuclear tests, and related to whether they might accidentally ignite the Earth's atmosphere, although these concerns were quickly dismissed.[12]

However, many who worked on the Manhattan Project continued to have severe reservations about the power of the weapons they helped to produce. After successfully performing the first controlled nuclear chain reaction at the University of Chicago in 1942, confirming its potential to release energy, the team of scientists working on the Manhattan Project dispersed, with some going off to Los Alamos and other research laboratories to develop nuclear weapons, while others stayed in Chicago to undertake their own research. Many of those who stayed were themselves immigrants to the United States and were keenly aware of the intertwining of science and politics. They, with the help of colleagues, began actively organising and engaging on how to keep the future of nuclear technology safe. For instance, they helped advance the Franck report in June 1945 that foreshadowed a dangerous and costly nuclear arms race, and argued against a surprise nuclear attack on Japan. This group went on to establish the *Bulletin of the Atomic Scientists of Chicago* (*The Bulletin*), whose first issue was published a mere four months after the atomic bombs were dropped on Hiroshima and Nagasaki. With support from the University of Chicago's President, Robert Hutchins, and colleagues in international law, political science, and other related fields, they helped kick off and support a global citizen-scientist movement that had a powerful effect on the creation of the global nuclear order.[13] Many of these same individuals were also

instrumental in establishing the Federation of Atomic (now American) Scientists, which was located in Washington DC to ensure proximity to key decision-makers whose views they hoped to sway. In contrast, *The Bulletin*'s headquarters in Chicago focused more on engaging and educating the public about the political and ethical challenges presented by the advancement of science, which they anticipated would only accelerate in the years to come. The founders believed that public pressure was key to political responsibility, and education was the best channel to ensure it.[14]

Two years after its founding, *The Bulletin* published Martyl Langsdorf's now iconic 'Doomsday Clock' to serve as the first cover of its new magazine. Over time, the Clock became globally recognised, in part because of its simplicity and bluntness. Married to a Manhattan Project scientist, Martyl was an artist who understood the urgency and desperation her husband and colleagues felt about managing nuclear technology. She created the Clock to convey their deep concern, as well as to draw attention to their belief that responsible citizens could prevent catastrophe by mobilising and engaging. The message of the Clock is clear: humans can prevent this clock from striking midnight. In that, it provides both a challenge and some hope.

In 1949 the USSR tested its first nuclear weapons, and in reaction to this, *The Bulletin*'s editor moved the hands of the Clock from seven to three minutes to midnight. In doing so, he activated the Clock, turning it from a static to a dynamic metaphor. The clock would evolve into a symbol that, according to Kennette Benedict, former Executive Director of *The Bulletin*:

> [warns] the public about how close we are to destroying our world with dangerous technologies of our own making. It is a metaphor, a reminder of the perils we must address if we are to survive on the planet.[15]

In 1953 the clock moved to two minutes to midnight, after the United States and USSR detonated the first thermonuclear weapons (H-bombs). This was the latest the clock was ever set in the 20th century (the furthest away it has been to midnight was 17 minutes in 1991, following the end of the Cold War). The Doomsday Clock, and its now annual setting, remains perhaps the most widely recognised and oft-cited symbol

of our existential predicament, as well as the most easily understood representation of our attempts to come to terms with it.

Another factor that increased both public and scientific concern about nuclear weapons was growing awareness of the risk that radioactive particles could contaminate the environment, with catastrophic effects. This theory was promoted by Hermann Muller, who discovered that radiation can induce genetic mutations and received the first post-war Nobel Prize in physiology for this work. Muller—along with Einstein, Bertrand Russell, and other prominent scientists of the day—later wrote the *Russell-Einstein Manifesto* in 1955, according to which:

> No one knows how widely such lethal radioactive particles might be diffused, but the best authorities are unanimous in saying that a war with H-bombs might possibly put an end to the human race... sudden only for a minority, but for the majority a slow torture of disease and disintegration.[16]

An important consequence of this manifesto was the establishment of the Pugwash Conferences on Science and World Affairs, which were initiated in 1957 by Russell and Joseph Rotblat, a physicist who also worked on the Manhattan Project. The Pugwash Conferences were vital for establishing communication channels at a time when Cold War tensions were at their highest, and the Conferences undertook vital background work to establish key non-proliferation treaties such as the 1963 Partial Test Ban Treaty, the 1968 Non-Proliferation Treaty, and the 1972 Biological Weapons Convention. Joseph Rotblat and the Pugwash Conferences were awarded the 1995 Nobel Peace Prize for their "efforts to diminish the part played by nuclear arms in international politics and, in the longer run, to eliminate such arms".[17] To this day, Pugwash remains the existential risk organisation with the widest global reach.[18]

Alongside these efforts of scientists, popular protest and resistance to the development, creation, and use of nuclear weapons was also vitally important. This resistance has taken many forms. For instance, in the UK, Bertrand Russell helped to establish both the Campaign for Nuclear Disarmament, a large and conventional pressure group, and the Committee of 100, a group set up specifically to perform acts of civil disobedience. He explained the need for both groups as follows:

> The Campaign for Nuclear Disarmament has done and is doing valuable and very successful work to make known the facts, but the press is becoming used to its doings and beginning to doubt their news value. It has therefore seemed to some of us necessary to supplement its campaign by such actions as the press is sure to report. There is another, and perhaps more important reason for the practice of civil disobedience in this time of utmost peril. There is a very widespread feeling that however bad their policies may be, there is nothing that private people can do about it. This is a complete mistake. If all those who disapprove of government policy were to join massive demonstrations of civil disobedience they could render government folly impossible and compel the so-called statesmen to acquiesce in measures that would make human survival possible.[19]

In founding this group, Russell established a justification for civil disobedience in the face of existential risk that remains to this day, most prominently in the Extinction Rebellion protests.

Russell was far from the only person to lead such a movement. Martin Luther King Jr and other civil rights leaders saw nuclear disarmament as essential and inextricably linked to the quest for social justice and racial equality. Many shared the view of Langston Hughes that American racism had played an important role in Harry Truman's decision to use nuclear weapons aggressively against Japanese people, and feared that they would be used selectively against non-whites in future.[20] There was also the argument that nuclear weapons were simply the latest, and most dangerous, manifestation of oppressive and destructive attitudes that marginalised groups had struggled against for centuries. As Dr King put it in his very final speech delivered at the Bishop Charles Mason Temple in Memphis on April 3rd 1968:

> Another reason that I'm happy to live in this period is that we have been forced to a point where we're going to have to grapple with the problems that men have been trying to grapple with through history, but the demands didn't force them to do it. Survival demands that we grapple with them. Men, for years now, have been talking about war and peace. But now, no longer can they just talk about it. It is no longer a choice between violence and nonviolence in this world; it's nonviolence or nonexistence.[21]

Also of great significance, but often overlooked, has been the resistance of indigenous peoples to nuclear colonialism. Indigenous lands and lives were often the first to be co-opted for the production of nuclear

weapons, from being displaced to make way for nuclear test sites to being hired at low wages to work in the mining and refining of uranium and other materials at significant costs to their own health.[22]

While concerned scientists and others did much to expose the risks from nuclear weapons as they developed, the nature of these risks meant that many people, including some of those who were responsible for creating these risks, were not aware of their immediacy. For instance, the 1963 Arkhipov incident, in which a Russian submarine commander wished to use nuclear weapons in retaliation for a perceived attack during the Cuban Missile Crisis before his junior officer overruled him, remained largely unknown until 2002.[23]

Policies can—and have—made a difference in reducing risk. For instance, Robert McNamara sent a memo to President John F. Kennedy in 1963 arguing that falling production costs made it realistic to expect at least eight new nuclear powers to emerge in the next ten years, while Kennedy himself predicted that perhaps 25 nuclear weapon states would emerge by the end of the 1970s.[24] Yet nuclear powers have emerged at a far slower pace. By some estimates, up to 56 states may have, at one time or another, possessed the capability to develop a nuclear weapons programme, yet the vast majority of these either chose not to engage in nuclear weapons activity or voluntarily terminated their programmes, with only ten states ever having developed nuclear weapons of their own—one of which (South Africa) has since disarmed.[25] In this, and other ways, the various initiatives we describe here have helped to make our world safer, though it remains far from safe. One 2013 study estimated the odds of a nuclear war having occurred between 1945 and 2011 as 61%;[26] however, in truth it is likely still too early to say what our chances really were and, indeed, we may never know.

Environmental breakdown and climate change

Concerns about risks to humanity from our own environmental impacts are nothing new. In *The Epochs of Nature* (1778), Georges-Louis Leclerc, the Comte de Buffon, wrote contemptuously of those who have "ravaged the land, starve it without making it fertile, destroy without building, use everything up without renewing anything".[27] In 1821, Charles Fourier wrote *The Material Deterioration of the Planet*, concerning

humanity's negative impacts on our environment and the harmful consequences for ourselves. While his theories do not match with our modern understanding of the planetary system, his concern that "we bring the axe and destruction, and the result is landslides, the denuding of mountain-sides, and the deterioration of the climate" still rings true today.[28] Similarly, Fredrich Engels noted how:

> In relation to nature, as to society, the present mode of production is predominantly concerned only about the immediate, the most tangible result. Then surprise is expressed that the more remote effects of actions directed to this end turn out to be quite different, are mostly quite the opposite in character.[29]

In 1896 Svante Arrhenius was the first to uncover the basic principles of anthropogenic climate change. As he later explained his findings to a general audience: "any doubling of the percentage of carbon dioxide in the air would raise the temperature of the earth's surface by 4°", while "the slight percentage of carbonic acid in the atmosphere may, by the advances of industry, be changed to a noticeable degree in the course of a few centuries."[30]

Yet prior to the mid-20th century, and the post-war 'great acceleration' of population, economic productivity, and environmental destruction, such concerns remained marginal. Some of the earliest general studies of the possibility for human extinction and the collapse of civilisation, including William Vogt's *Road to Survival*[31] and Fairfield Osborne's *Our Plundered Planet*,[32] linked this threat specifically to environmental harms such as soil erosion and pollution. Another pivotal early work was Rachel Carson's *Silent Spring*,[33] which not only echoed these earlier concerns, but increased their scientific rigour and added a crucial policy edge by raising public awareness about the danger from chemical pesticides, such as DDT. Carson was a marine biologist, nature writer, and pioneering conservationist who became concerned about the ecological effects of indiscriminate overuse of pesticides, which she called "biocides". As she wrote in *Silent Spring*:

> Along with the possibility of the extinction of mankind by nuclear war, the central problem of our age has ... become the contamination of man's total environment with such substances of incredible potential for harm—substances that accumulate in the tissues of plants and animals

> and even penetrate the germ cells to shatter or alter the very material of heredity upon which the shape of the future depends (Carson, 1962).

A further growth in public concern about environmental harms came in 1968, when two young biologists, Paul and Anne Ehrlich, were commissioned to write *The Population Bomb*,[34] which received widespread attention in both the academic and popular press. It warned about the catastrophic impacts of overpopulation, which the Ehrlichs claimed could lead to "hundreds of millions" of deaths from starvation. Early manifestations of this concern included the founding of organisations such as Friends of the Earth and Greenpeace (both in 1969) and the first Earth Day (April 22nd 1970), which saw 20 million Americans march in cities across the country. In 1972, the Club of Rome—an organisation of scientists, economists, diplomats, government officials, and other influencers from around the world—published *The Limits to Growth*,[35] which developed the first global systems models to investigate the long-run impacts of trends in population, consumption, environmental degradation, and technology.[36] Its conclusions were stark: "If the present growth trends in world population, industrialization, pollution, food production, and resource depletion continue unchanged, the limits to growth on this planet will be reached sometime within the next one hundred years". By 1978, *The Bulletin* weighed in with a cover story asking *Is Mankind Warming the Earth?*—to which its author answered, "Yes."[37]

One of the more optimistic pronouncements within the *Limits to Growth* report was that the advent of nuclear energy might mean that the atmospheric concentration of greenhouse gases may cease to rise, "one hopes before it has had any measurable ecological or climatological effect". History has not borne this prediction out. Nor did it take long for scientists to confirm that humanity's emission of greenhouse gases was already having deleterious climatic and ecological impacts, with grave implications for our future. Less than a decade later, James Hansen led a ground-breaking study in *Science*, showing that:

> the anthropogenic carbon dioxide warming should emerge from the noise level of natural climate variability by the end of the century, and there is a high probability of warming in the 1980s. Potential effects on climate in the 21st century include the creation of drought-prone regions in North America and central Asia as part of a shifting of climatic zones,

> erosion of the West Antarctic ice sheet with a consequent worldwide rise in sea level, and opening of the fabled Northwest Passage.[38]

Hansen would go on to greatly increase public awareness of the reality of climate change when he testified before the US Senate Committee on Energy and Natural Resources, although the fame of this event was as much due to attempts by the Office for Budgetary Responsibility to censor Hansen's explanation of climate models and what they indicated.[39] However, while Hansen's climate research from this period is especially well-known, he was only one part of a vast array of scientists arriving at the modern consensus about anthropogenic climate change, a consensus that was well-established by the 1980s and has only strengthened since.[40]

This consensus has helped us to achieve some impressive feats of global governance and policy change, including the formation of the Advisory Group on Greenhouse Gases in 1985, and its development into the Intergovernmental Panel on Climate Change in 1988, as well as the signing of the United Nations Framework Convention on Climate Change at the Earth Summit in Rio de Janeiro in 1992, and its subsequent amendments (such as the 1997 Kyoto Protocol and 2015 Paris Agreement). However, equally important have been the consistent efforts of a small number of vested interests to undermine this consensus and spread misinformation about climate change and other risks, limiting the impact of what might otherwise have been possible.[41] As with nuclear weapons, however, this is not a gap that scientists can bridge on their own. Support for public efforts of advocacy and resistance is also needed if progress is to be made, especially those of the people who will be worst affected by climate change, such as poor, landlocked, small island nations and marginalised and indigenous communities around the world.[42]

As public awareness of, and concern about, environmental risks facing humanity has grown, people have increasingly come to talk about nuclear weapons and climate change in the same breath as equal threats to the future of humanity. For instance, in 2007 *The Bulletin* updated their framing of the Doomsday Clock to formally recognise climate change as a factor in setting it.[43] Of course, the timescale of these risks remains quite different: a nuclear exchange could happen within minutes, while climate risk is slowly accumulating year after year. Similarly, responsibility for the world's nuclear weapons lies in

the hands, or at the fingers, of only a few global decision-makers, while we are all engaged in climate change and environmental destruction, even if to a very unequal extent.[44] However, the severity of these risks—both in terms of their potential to cause global catastrophes and their likelihood of doing so—are undoubtedly comparable, and we should also assess both risks in terms of whether or not the current level of global action being taken to combat them is proportional to this severity and the rising urgency of reducing it.

The convergence of nuclear and environmental risks in the nuclear winter hypothesis

Yet, in at least one instance, concerns about environmental and nuclear risks have intertwined, and it is an event that warrants special attention as it may have been the occasion when scientific concern about existential risk most directly influenced global policy—the publication of the 'nuclear winter' hypothesis. By the early 1980s, some scientists had become worried that the greatest threat posed by nuclear conflict was not radioactivity but the massive firestorms that could inject soot into the stratosphere, blocking incoming solar radiation and causing global agricultural failures, environmental catastrophes, and maybe even human extinction. The result would be what the atmospheric scientist Richard Turco called "the nuclear winter". This hypothesis arose at a time when nuclear tensions were increasing, with events such as stalled US-USSR nuclear negotiations, rising numbers of global conflicts, and the entry of India as a nuclear weapons state, causing the Doomsday Clock to tumble forward nine minutes in 12 years and reach three minutes to midnight by 1984.

This possibility was rendered more plausible because of a study published in 1980 by Luis and Walter Alvarez that hypothesised that the non-avian dinosaurs became extinct because an asteroid struck Earth.[45] This impact threw dust into the stratosphere, blocking out sunlight and compromising photosynthesis. This hypothesis threatened the then-dominant paradigm that there is a 'balance of nature' that enforces gradual changes on life and protects the Earth from catastrophes, which had reigned for more than a century.[46] When the hypothesis was subsequently corroborated by the discovery of the Chicxulub crater

on the Yucatan Peninsula, it prompted the scientific community to acknowledge that events in the past had caused globally catastrophic effects, and this raised the possibility that it might happen again, this time due to anthropogenic causes.[47]

One of the most prominent scientists who warned about nuclear winter was the cosmologist, planetary physicist, and exobiologist Carl Sagan. Sagan had gained significant scientific prominence through his research, especially in the search for extra-terrestrial life, and had a pre-eminent reputation as a science communicator through his books and TV programmes such as *Dragons in Eden* and *Cosmos*.[48] Sagan and four other scientists published an influential study modelling the possibility of nuclear winter in the journal *Science*.[49] However, he also pre-empted this publication with more popular works and media appearances to increase the potential impact of the research on politicians and the public. For instance, Sagan wrote a cover story for the October 30th edition of the popular Sunday news supplement *Parade* magazine arguing that, should a nuclear conflict occur:

> Many species of plants and animals would become extinct. Vast numbers of surviving humans would starve to death. The delicate ecological relations that bind together organisms on Earth in a fabric of mutual dependency would be torn, perhaps irreparably. There is little question that our global civilization would be destroyed. The human population would be reduced to prehistoric levels, or less. Life for any survivors would be extremely hard. And there seems to be a real possibility of the extinction of the human species.[50]

To further develop this point, Sagan joined forces with Paul Ehrlich to co-organise a two-day conference and co-author the 1984 book on the "long-term biological consequences of nuclear war," *The Cold and the Dark*.[51]

Both the research behind the nuclear winter hypothesis and the strategy for its promotion were controversial. While the hypothesis itself was plausible in many respects, the modelling used to support it was seen as quite primitive, even at the time, and debate still rages about how significant a cooling effect a limited nuclear war might have. However, the greater controversy was about the scientists' decisions to go public before their paper had been published, and harness the power of the media to get their idea across quickly. This may have been

inspired (in part) by Sagan's health issues; he is said to have dictated a letter opposing Ronald Reagan's Strategic Defense Initiative (popularly known as Star Wars) from his hospital bed while recovering from an operation. However, it may also have been influenced by the involvement of Ehrlich and his earlier popular successes.

The article, along with others, certainly had an impact.[52] The work of Sagan and his counterparts harnessed public attention and helped to spur opposition to the Strategic Defense Initiative, inspire a Soviet moratorium on nuclear weapons tests in 1985, and ensure both parties continued negotiating towards the Strategic Arms Reduction Treaty (START). Soviet Premier Mikhail Gorbachev told Ronald Reagan in 1988 that Sagan was "a major influence on ending [nuclear] proliferation".[53] This was brought about by a swathe of new arms control initiatives including the 1987 Intermediate-Range Nuclear Forces Treaty (which banned all Russian and US land-based ballistic missiles with ranges between 500 and 5,500 km and saw 2,692 nuclear missiles removed from service), the 1991 signing of START (which would eventually lead to the removal of around 80% of nuclear weapons), the 1991 Presidential Nuclear Directives (that removed the vast majority of nuclear weapons out of the European theatre), and the Nunn-Lugar Cooperative Threat Reduction program (which wound down and ended the USSR's secret, illegal biological weapons programme).

These developments, together with political changes around the world at the end of the Cold War, saw a rapid improvement in humanity's existential situation, as charted by a shift of the Doomsday Clock from three minutes to midnight in 1984 to 17 minutes to midnight in 1991, the safest we had been since the end of World War II. The work of so many people to understand and draw attention to existential risks over the 20th century had undoubtedly helped to bring about this favourable state. While political leaders may have driven these outcomes, scientists provided the technical foundation and helped support a global grassroots civic effort that improved the safety and security of what Sagan evocatively described as this "pale blue dot, the only home we've ever known".[54]

The road to 90 seconds to midnight: Growing risks, disruptive technologies, and the foundation of Existential Risk Studies

Yet this positive state of affairs was not to last. A mere 12 years later, in 2003, the Doomsday Clock was back to where it began in 1947, at seven minutes to midnight. That same year, in his book *Our Final Century*, the British Astronomer Royal Martin Rees predicted that: "The odds are no better than fifty-fifty that our present civilization on earth will survive to the end of the present century."[55] Sadly, subsequent events have very much borne out such pessimism and the Doomsday Clock has continued its slide and today stands at 90 seconds to midnight, closer than it was even at the height of the Cold War. How has this happened?

One factor that has undoubtedly contributed to this massive decline in fortune has been the continued emergence of new kinds of threats. The pages of *The Bulletin* have long considered the challenges posed by new disruptive technologies, and these have recently been adopted as a third contributing factor when setting the Doomsday Clock.[56] Scientists first identified technological threats to human survival decades ago, including those associated with AI,[57] biological weapons,[58] nanotechnology,[59] and high-energy physics experiments.[60, 61] As well as threats from specific technologies, many have also come to see that our future is increasingly imperilled by the convergence of disruptive technologies with existing nuclear and environmental threats. These discoveries in turn precipitated the gradual formation of a dedicated academic discipline for studying existential risk, such as the publication of *Our Final Century* and the establishment of new research institutes like the Future of Humanity Institute, the Future of Life Institute, the Global Catastrophic Risk Institute, and the Centre for the Study of Existential Risk.[62] Many of these centres originated out of a view that transformative technologies were necessary precisely because they were required to help humanity reach a more safe and beneficial future, by giving us greater control over our environment, our societies, and ourselves, but that we could only achieve this if we could find out how to develop them safely. However, this original vision has now been joined by many other ways of understanding existential risks and how to manage them, such as by applying lessons from disaster studies[63] and

security studies,[64] viewing risks from a global systems perspective,[65] placing risk mitigation in a broader policy context,[66] and identifying opportunities to increase global resilience.[67]

As the number and variety of threats facing humanity has multiplied, so has the seriousness of the challenges posed by nuclear and environmental risks. By 2015 *The Bulletin* had moved its Doomsday Clock from five to three minutes to midnight. There were three key issues behind this move. Firstly, deteriorating relations between the US and Russia—who together possess 90% of the world's nuclear arsenal—and the actual and threatened dismantling of many of the Cold War instruments designed to keep those arsenals safe, such as the successor to the START treaty (New START).[68] Secondly, every major nuclear state was investing massively in its nuclear weapons systems, including replacement, expansion, and modernisation.[69] Finally, the global architecture needed to address climate threats was nowhere in sight.

In 2016 *The Bulletin*'s science and security board identified two possible bright spots, with the potential to reverse some of these negative trends: the Iran nuclear deal and the Paris climate agreement. However, they also noted that both agreements still had to be implemented, and in 2017 were forced to conclude that the situation had significantly worsened, with both of these bright spots being dimmed by changes in US domestic politics and growing evidence of a global disparagement of expertise and a recklessness about nuclear language and leadership. They thus moved the clock to two and a half minutes to midnight, and in 2018 moved it further to two minutes to midnight due to the continuing deterioration of international diplomacy.

In 2020, *The Bulletin* moved the clock closer to midnight than it has ever been before, a decision also endorsed by researchers at the Centre for the Study of Existential Risk.[70] Above all, the new time reflected the sheer instability of the current global situation, and the failure of international institutions to respond to the ticking clock of existential risk, including the collapse of the Intermediate Nuclear Forces Treaty that had marked the beginning of the end of the Cold War itself. We are back to a global political situation that is akin to what it was during the Cold War, and yet the risks we face now are so much more numerous and complex. If we are to turn back the Clock, we need a new wave of foresight, creativity, engagement, and education to study and manage

these risks. Sadly, recent events such as the COVID-19 pandemic and Russian invasion of Ukraine indicate that the scale of this challenge may only be growing, and this was recognised when, just a few days before this chapter went to press, *The Bulletin* moved the clock further forward to 90 seconds to midnight.

Learning from the pioneers of existential risk

How can studying the history of existential risk help us succeed in mitigating it? For one thing, it helps those fighting for humanity to focus on our long-term goals and the need for lasting safety, rather than getting stuck passing from urgent crisis to urgent crisis with only transient respites (like the one experienced at the end of the Cold War).

There are also more specific things we can learn from this history. First, certain key messages require constant reinforcement, such as that global catastrophes are a real possibility against which we have no fundamental protection. There have been times in the past, such as during the Cuban Missile Crisis, when politicians took these messages to heart, such as by frequently reiterating that "a nuclear war can never be won and must never be fought", but evidence suggests that this does not last. Second, if we are to understand these catastrophes we will have to engage in some speculation and modelling, such as that which was used to develop the nuclear winter hypothesis, as it will be too costly to wait until we can observe disasters directly. Third, if we are to prevent these catastrophes then we need to place this science in the public domain and be willing to publicly advocate for it, even if that is (at times) difficult or controversial, and especially if it involves working across cultural, geographical, or political divides, as the Pugwash Conferences were able to. Fourth, scientists and their allies must be called on repeatedly to attest to the fact that our current pace of scientific and technological development carries risks as well as benefits. While we all sincerely wish for scientific advancement to be wholly beneficial, this rarely—if ever—is the case. And yet, it is not in the province of scientists alone to ensure the benefits outweigh the risks. It requires diplomacy, governance, and public engagement. Finally, we need to build closer links between scientists and the public, and recognise that grassroots efforts (and even civil disobedience) may be essential components for managing

existential risk in the current geopolitical situation—what Nick Bostrom refers to as our "semi-anarchic default condition".[71]

We need, therefore, to be both realistic and respectful in our evaluation of this pioneering work. Many of the theoretical frameworks within which scientists work tend to be useful only for linking discrete exogenous shocks with catastrophic effects; for instance, by considering a simple causal chain from nuclear conflict, to firestorms, to stratospheric soot injection, to famine. Questions about the exact size of such risks, or the timeframes within which they might be expected to occur, are not amenable to precise answers and, while understandable, the public's demand for such answers has encouraged some to make claims that have not been borne out by the subsequent facts. How to handle such tensions responsibly is a challenge facing all who work on existential risk, and one we have heard few address with the directness it deserves. Furthermore, such studies are hard to reconcile with the need for a science of existential risk that can move beyond direct drivers of risk to embrace complexity and chaos in all its forms. This is one area where the integrated approaches to the study of existential risk that have emerged in the field of Existential Risk Studies may be especially valuable.[72] However, even here, approaches will undoubtedly continue to be indebted to systems thinkers and complexity scientists.

The founding objectives of *The Bulletin*, which echo the hopes of many who have sought to work on existential risks, were two-fold: to make scientists aware of "the relationship between their own world of science and the national and international politics", and to educate the public about the need to manage the "dangerous presents from Pandora's box of modern science".[73] Focusing attention on threats to human existence is not for the faint of heart, and the founders of *The Bulletin* were well aware that their aims would only be realised by a long, sustained effort. Political solutions can appear inconsequential against such extreme outcomes. Technological advancement will bring, and has brought, such grand benefits that it is disquieting to advocate for ethical and political roadblocks to reduce potential—and sometimes merely theoretical—risks. While the concerned scientists who have helped to keep humanity safe and flourishing are often labelled as pessimistic, history shows us that this charge is unfair. To assume that scientific, social, and technological change only carries risks would

leave us unable to explain the indisputable benefits they have given to humanity, and their potential to do even more good in future. However, to focus only on these benefits could lead us to fatally ignore some of the very real challenges currently facing us.

For the last 75 years, the Doomsday Clock has served as one of the few popular images that is able both to break through the global public discourse and to define such difficult trade-offs. Still, knowing the risks is never enough; we must also figure out the space of safe and beneficial scientific development and create the restraints necessary to govern it. These are hard problems. However, we can see from the history of existential risk that when talented scientists work together with committed politicians and motivated activists, it is possible to do more than simply cross our fingers and hope for the best. We can weigh the dice in our favour; we can turn back the hands of the Clock.

Notes and References

1 There are many instances of this myth. Another prominent version can be found in Tablet XI of the Epic of Gilgamesh. Some have suggested that the prevalence of the deluge in Mesopotamian culture attests to folk memories of real catastrophic floods caused by prehistoric global warming at the end of the last glacial age.

2 For more on this history, see Moynihan, T., *X-risk: How Humanity Discovered Its Own Extinction*. MIT Press (2020); and Torres, E., *Human Extinction A History of the Science and Ethics of Annihilation*. Routledge (2023).

3 Shelley, M.W., *The Last Man.* Henry Colburn (1826).

4 Butler, S., 'Darwin among the machines', *The Press, June 13 1863* (1863); Butler, S., 'The book of the machines', *Erewhon: Or, Over the Range*. Trübner and Ballantyne (1872).

5 The first mention of autonomous machines causing human extinction due to 'value misalignment', the principle concern for many scholars of existential risk today, can be found in Jack Williamson's short story *With Folded Hands* (Williamson, J., 'With folded hands', *Astounding Science Fiction*. Street & Smith (1947)), implying that this facet of AI ethics predates even other canonical AI-related thought experiments, such as John Searle's *Chinese Room* (Searle, J.R., 'Minds, brains, and programs', *Behavioral and Brain Sciences, 3*(3) (1980), pp.417–57. https://

doi.org/10.1017/S0140525X00005756) or Alan Turing's *Imitation Game* (Turing, A.M., 'Computing machinery and intelligence', *Mind LIX* (236) (1950), pp.433–60. https://doi.org/10.1093/mind/LIX.236.433). A comprehensive record of when different kinds of AI-induced catastrophes were first described can be found at https://timelines.issarice.com/wiki/Timeline_of_AI_safety#Timeline.

6 Wells, H.G., *The World Set Free: A Story of Mankind*. E.P. Dutton (1914).

7 The earliest example of this genre (often known as 'cli-fi' or climate fiction) may be Ballard, J.G., *The Burning World*. Berkley Pub (1964).

8 Wells, H.G., *Anticipations of the Reaction of Mechanical and Scientific Progress Upon Human Life and Thought*. Chapman & Hall (1901).

9 Warren, W.W., 'H.G. Wells and the genesis of future studies', *World Future Society Bulletin, 17*(1) (1983), pp.25–29.

10 Wells, H.G., 'On extinction', *Chambers's Journal* (30 September 1893) and Wells, H.G., 'The extinction of man', *Certain Personal Matters: A Collection of Material, Mainly Autobiographical*. William Heinemann (1897).

11 Asimov, I., *A Choice of Catastrophes: The Disasters That Threaten Our World*. Simon & Schuster (1979).

12 Konopinski, E.J., C. Marvin, and E. Teller, *Ignition of the Atmosphere With Nuclear Bombs.* Los Alamos National Laboratory (1946). https://fas.org/sgp/othergov/doe/lanl/docs1/00329010.pdf

13 Kimball Smith, A., *A Peril and a Hope: The Scientists' Movement in America: 1945–47*. University of Chicago Press (1965).

14 One early success of these citizen-scientists was contributing to the establishment of the 'nuclear taboo' against use that has lasted to this day. In private conversations, Eisenhower's Secretary of State complained that the "stigma of immorality" prevented the US from using nuclear weapons. The next generation of citizen-scientists in the 60s and 70s built on this by taking the ideas of arms control into government and helping to achieve the first agreements with the Soviets. See Adler, Emanuel, 'The emergence of cooperation: National epistemic communities and the international evolution of the idea of nuclear arms control', *International Organization,* 46(1) (1992), pp.101–45; and Tannenwald, Nina, *The Nuclear Taboo: The United States and the Non-Use of Nuclear Weapons Since 1945*. Cambridge University Press (2007).

15 Benedict, J., 'Doomsday Clock', *Bulletin of the Atomic Scientists* (January 2018). https://thebulletin.org/2018/01/doomsday-clockwork/

16 The full text of this manifesto can be read at https://pugwash.org/1955/07/09/statement-manifesto/

17 Norwegian Nobel Committee, 'The Nobel Peace Prize 1995' [Press Release] (1995) https://www.nobelprize.org/prizes/peace/1995/press-release/

18 Many of the founders of Pugwash, including Einstein and Russell, had initially believed that establishing a world government might be necessary to prevent humanity eradicating itself as a result of the advent of nuclear weapons, by eradicating the divisions that were seen as the causes of conflict. However, the organisation itself stands as a lasting testament to the ability of individual scientists to work across ideological and geographical divides to make the world safer, in spite of its divisions. For more on this distinction between 'top-down' and 'bottom-up' approaches to tackling existential risk and the history of Pugwash and similar initiatives, see the chapter by Lalitha S. Sundaram: 'Existential Risk and Science Governance'.

19 Russell, B., 'Civil disobedience', *New Statesman*, 17 February 1961 (1961).

20 Williams, P., 'Physics made simple: The image of nuclear weapons in the writing of Langston Hughes', *Journal of Transatlantic Studies*, *6*(2) (2008), pp.131–41.

21 King Jr, M.L., 'I see the promised land', in J.M. Washington (ed.), *A Testament of Hope: The Essential Writings and Speeches of Martin Luther King, Jr.* HarperOne (1986), pp.279–86.

22 For further resources on nuclear colonialism and indigenous people's resistance to it, see https://www.nirs.org/resisting-nuclear-colonialism-on-indigenous-peoples-day/ and http://www.environmentandsociety.org/exhibitions/risk-and-militarization/nuclear-colonialism

23 Savranskaya, S.V., 'New sources on the role of Soviet submarines in the Cuban Missile Crisis', *Journal of Strategic Studies*, *28*(2) (2005), pp.233–59.

24 Yusuf, M., *Predicting Proliferation: The History of the Future of Nuclear Weapons*. Brookings Institution (2009).

25 van der Meer, S., 'Forgoing the nuclear option: States that could build nuclear weapons but chose not to do so', *Medicine, Conflict and Survival*, *30*(supplement 1) (2014), pp.27–34.

26 Lundgren, C., 'What are the odds? Assessing the probability of a nuclear war', *The Nonproliferation Review*, *20*(2) (2013), pp.361–74.

27 Leclerc, G.L., *The Epochs of Nature*. University of Chicago Press (2018).

28 Fourier, C., *Selections From the Works of Fourier*. Sonnenschein & Company (1901).

29 Engels, F., 'The part played by labour in the transition from ape to man', first published in *Die Neue Zeit 1895–06*. Published in English by Progress Publishers (1934), translated by Clemens Dutt.

30 Arrhenius, S., *Worlds in the Making: The Evolution of the Universe*. Harper (1908). Note, however, that both Fourier and Arrhenius were actually more concerned about the possibility of global cooling than global warming, and that such (now widely discredited) anxieties continued to receive significant scientific interest well into the 20th century. For further discussion of this history see Locher, F. and Fressoz, J.B. 'Modernity's frail climate: A climate history of environmental reflexivity', *Critical Inquiry, 38*(3) (2012), pp.579–98.

31 Vogt, W., *Road to Survival*. William Sloan Associates (1948).

32 Osborn, F., *Our Plundered Planet*. Little, Brown and Company (1948).

33 Carson, R., *Silent Spring*. Houghton Mifflin Company (1962).

34 Ehrlich, P.R., *The Population Bomb*. Ballantine Books (1968).

35 Meadows, D.H., D.L. Meadows, J. Randers, and W.W. Behrens, *The Limits to Growth*. Universe Books (1972).

36 While this was the earliest (and, to date, most prominent) research by the Club of Rome, it is far from the only attempt to model the risks and causes of societal collapse. For more on the techniques employed by the Club of Rome and other researchers, see the chapter by Sabin Roman: 'Understanding and Predicting Societal Collapse'.

37 Kellogg, W.K., "Is mankind warming the earth?" *Bulletin of the Atomic Scientists, 34*(2) (1978), pp.10–19 (1978), https://doi.org/10.1080/00963402.1978.11458464

38 Hansen, J., D. Johnson, A. Lacis, S. Lebedeff, P. Lee, D. Rind, and G. Russell. 'Climate impact of increasing atmospheric carbon dioxide', *Science*, *213*(4511), pp.957–66 (1981). Similar concerns were raised the following year in an internal report for ExxonMobil that predicted a one-degree average global temperature rise by 2022—a prediction that, regrettably, has now been shown to have been highly accurate. See Supran, G. and N. Oreskes, 'Assessing ExxonMobil's climate change communications (1977–2014)', *Environmental Research Letters, 12*(8) (2017), p.084019.

39 Rich, N., *Losing Earth: The Decade We Could Have Stopped Climate Change*. Pan Macmillan (2019).

40 However, note that, while this consensus has existed since the 1980s, the topic of climate change as a global catastrophic risk has received

surprisingly little attention from the scientific community, even as it has become a hot topic of public and political debate. A full assessment of the research currently existing in this space—most of which has only been conducted within the last decade—can be found in Luke Kemp's chapter 'Ecological Breakdown and Human Extinction'.

41 Oreskes, N. and E.M. Conway, *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues From Tobacco Smoke to Global Warming*. Bloomsbury (2011).

42 However, the importance of working with groups seeking racial, indigenous, and other forms of justice in reducing existential threats like nuclear weapons and climate change is not merely that they have (in fact) been important agents of change. As Natalie Jones sets out in her chapter 'Beyond "Error and Terror": Global Justice and Global Catastrophic Risk', a failure to understand issues of global injustice, inequality, corruption, and oppression can give us a flawed understanding of the risks facing humanity; however, privileged academics in elite universities are often poorly situated to fully understand this.

43 Not that this was an especially new departure for the magazine. *The Bulletin*'s first article to mention anthropogenic climate change came out in 1954 (albeit in a satire of different public attitudes towards fossil fuels and nuclear power) while in 1961 it published further articles on climate change as part of a three-volume special on *Man and His Habitat*. See Frisch, O.R., 'On the feasibility of coal-driven power stations', *Bulletin of the Atomic Scientists, 10*(6) (1954), p.224. https://doi.org/10.1080/00963402.1954.11453480 and Suess, H.E., 'Fuel residuals and climate', *Bulletin of the Atomic Scientists*, *17*(9) (1961), pp.374–75.

44 Hickel, J., 'Quantifying national responsibility for climate breakdown: An equality-based attribution approach for carbon dioxide emissions in excess of the planetary boundary', *The Lancet Planetary Health*, *4*(9), pp.e399-e404 (2020).

45 Alvarez, L.W., W. Alvarez, F. Asaro, and H.V. Michel, 'Extraterrestrial cause for the cretaceous-tertiary extinction', *Science 208*(4448), pp.1095–108 (1980), https://doi.org/10.1126/science.208.4448.1095.

46 Palmer, T., *Controversy Catastrophism and Evolution: The Ongoing Debate*. Springer Science & Business Media (2012).

47 Of course, it also drew scientific attention to the possibility of similar catastrophes occurring due to asteroid impacts or volcanic eruptions in the future, as discussed by Lara Mani, Doug Erwin, and Lindley Johnson in their chapter 'Natural Global Catastrophic Risks'.

48 Sagan, C., *The Dragons of Eden: Speculations on the Evolution of Human Intelligence*. Penguin Random House (1977); Sagan, C., A. Druyan, and S. Soter, *Cosmos: A Personal Voyage*. PBS (1980).

49 Turco, R.P., O.B. Toon, T.P. Ackerman, J.B. Pollack, and C. Sagan, 'Nuclear winter: Global consequences of multiple nuclear explosions', *Science 222*(4630), pp.1283–92 (1983), https://doi.org/10.1126/science.222.4630.1283.

50 Sagan, C., 'Would nuclear war be the end of the world?' *Parade* (October 30th 1983).

51 Ehrlich, P.R., C. Sagan, D. Kennedy, and W.O. Roberts, *The Cold and the Dark: The World After Nuclear War.* W.W. Norton & Company (1984). While Ehrlich was initially sceptical that a nuclear conflict could cause human extinction, his view eventually changed. In his words: "it was the consensus of our group that, under those conditions, we could not exclude the possibility that the scattered survivors simply would not be able to rebuild their populations, that they would, over a period of decades or even centuries, fade away. In other words, we could not exclude the possibility of a full-scale nuclear war entraining the extinction of Homo sapiens." Badash, L., *A Nuclear Winter's Tale: Science and Politics in the 1980s*. MIT Press (2009).

52 Although it is worth noting that other important work on nuclear war and human extinction of the time did not depend on the nuclear winter hypothesis, such as Jonathan Schell's highly influential article series and book *The Fate of the Earth* (Schell, J., 'The fate of the Earth', *The New Yorker* (February 1982).

53 Frances, R.M., 'When Carl Sagan warned the world about nuclear winter', *Smithsonian Magazine* (2017).

54 Sagan, C., *Pale Blue Dot: A Vision of the Human Future in Space*. Random House (1994).

55 Rees, M., *Our Final Century: Will Civilisation Survive the Twenty-First Century?* Random House (2003).

56 For instance, *The Bulletin* has been raising awareness of the risks from biotechnology since the 1940s, geoengineering since the 1960s, AI since the 1970s, and nanotech since the 2000s. See Thimann, K.V., 'The role of biologists in warfare', *Bulletin of the Atomic Scientists*, *3*(8), pp.211–12 (1947), https://doi.org/10.1080/00963402.1947.11459090; Landsberg, H.E., 'Climate made to order', *Bulletin of the Atomic Scientists, 17*(9), pp.370–74 (1961), https://doi.org/10.1080/00963402.1961.11454271; Weizenbaum, J., 'Once more—A computer revolution', *Bulletin of the Atomic Scientists, 34*(7) (1978), pp.12–19. https://doi.org/10.1080/00963402.1978.11458531;

and Kosal, Margaret E., 'Nanotech: Is small scary?' *Bulletin of the Atomic Scientists, 60*(5) (2004), pp.38–47., https://doi.org/10.1080/00963402.2004.11460819.

57 Good, I.J., 'Speculations concerning the first ultraintelligent machine', *Advances in Computers 6* (1966), pp. 31-88, https://doi.org/10.1016/S0065-2458(08)60418-0.

58 Lederberg, J., 'Biological warfare and the extinction of man', in a statement before the Subcommittee on National Security Policy and Scientific Developments, House Committee on Foreign Affairs (1969).

59 Drexler, K.E., *Engines of Creation: The Coming Era of Nanotechnology*. Anchor Press/Doubleday (1986).

60 Dar, A., A. De Rújula, and U. Heinz. 'Will relativistic heavy-ion colliders destroy our planet?', *Physics Letters B*, *470*(1–4), pp.142–48 (1999).

61 The history of thinking about AI as an existential risk is discussed in John Burden, Sam Clarke, and Jess Whittlestone's chapter: 'A History of AI Existential Safety', while the growing threat from biotechnology and developments in the life sciences is discussed in Kelsey Lane Warmbrod, Kobi Leins, and Nancy Connell's chapter: 'Biosecurity, Biosafety and Dual Use: Will Humanity Minimise Potential Harms in the Age of Biotechnology?'

62 And these have also been joined by a growing range of think tanks and policy bodies such as the Centre for Security and Emerging Technology, the Centre for Health Security, and the Nuclear Threat Initiative.

63 Liu, H.Y., K.C. Lauta, and M.M. Maas, 'Governing boring apocalypses: A new typology of existential vulnerabilities and exposures for existential risk research', *Futures*, *102* (2018), pp.6–19.

64 Sears, N.A., 'Existential security: Towards a security framework for the survival of humanity', *Global Policy*, *11*(2) (2020), pp.255–66.

65 Avin, S., B.C. Wintle, J. Weitzdörfer, S.S. Ó hÉigeartaigh, W.J. Sutherland, and M.J. Rees, 'Classifying global catastrophic risks', *Futures*, *102* (2018), pp.20–26.

66 Cotton-Barratt, O., M. Daniel, and A. Sandberg, 'Defence in depth against human extinction: Prevention, response, resilience, and why they all matter', *Global Policy*, *11*(3) (2020), pp.271–82.

67 Baum, S., 'Resilience to global catastrophe', in Benjamin D. Trump, Marie-Valentine Florin, and Igor Linkov (eds), *IRGC Resource Guide on Resilience (Vol. 2): Domains of Resilience for Complex Interconnected Systems*. EPFL International Risk Governance Center (2018).

68 New START was eventually renewed by the Biden Administration in 2021 before being suspended by Russia in 2023.

69 A particular concern within nuclear modernisation has been the combination of AI and nuclear weapons, as discussed by Matthijs Maas, Kayla Lucero-Matteucci, and Di Cooke's chapter: 'Military Artificial Intelligence as Contributor to Global Catastrophic Risk'.

70 See https://www.cser.ac.uk/news/doomsday-clock-2020/

71 Bostrom, N., 'The vulnerable world hypothesis', *Global Policy*, *10*(4) (2019), pp.455–76.

72 For some introductory works on these approaches, see Leslie, J.A., *The End of the World: The Science and Ethics of Human Extinction*. Routledge (1996); Bostrom, N., 'Existential risks: Analyzing human extinction scenarios and related hazards', *Journal of Evolution and Technology, 9*(1) (2002). https://www.nickbostrom.com/existential/risks.pdf; Bostrom, N., 'The vulnerable world hypothesis', *Global Policy*, *10*(4) (2019), pp.455–76; Walsh, Bryan, *End Times: A Brief Guide to the End of the World*. Hachette (2019); Ord, T., *The Precipice: Existential Risk and the Future of Humanity*. Hachette (2020); as well as the other chapters in this volume.

73 Rabinowitch, E. and M. Grodzins, 'Bulletin', *Bulletin of the Atomic Scientists, 1*(1), p. 1(1945), https://doi.org/10.1080/00963402.1945.11454590.