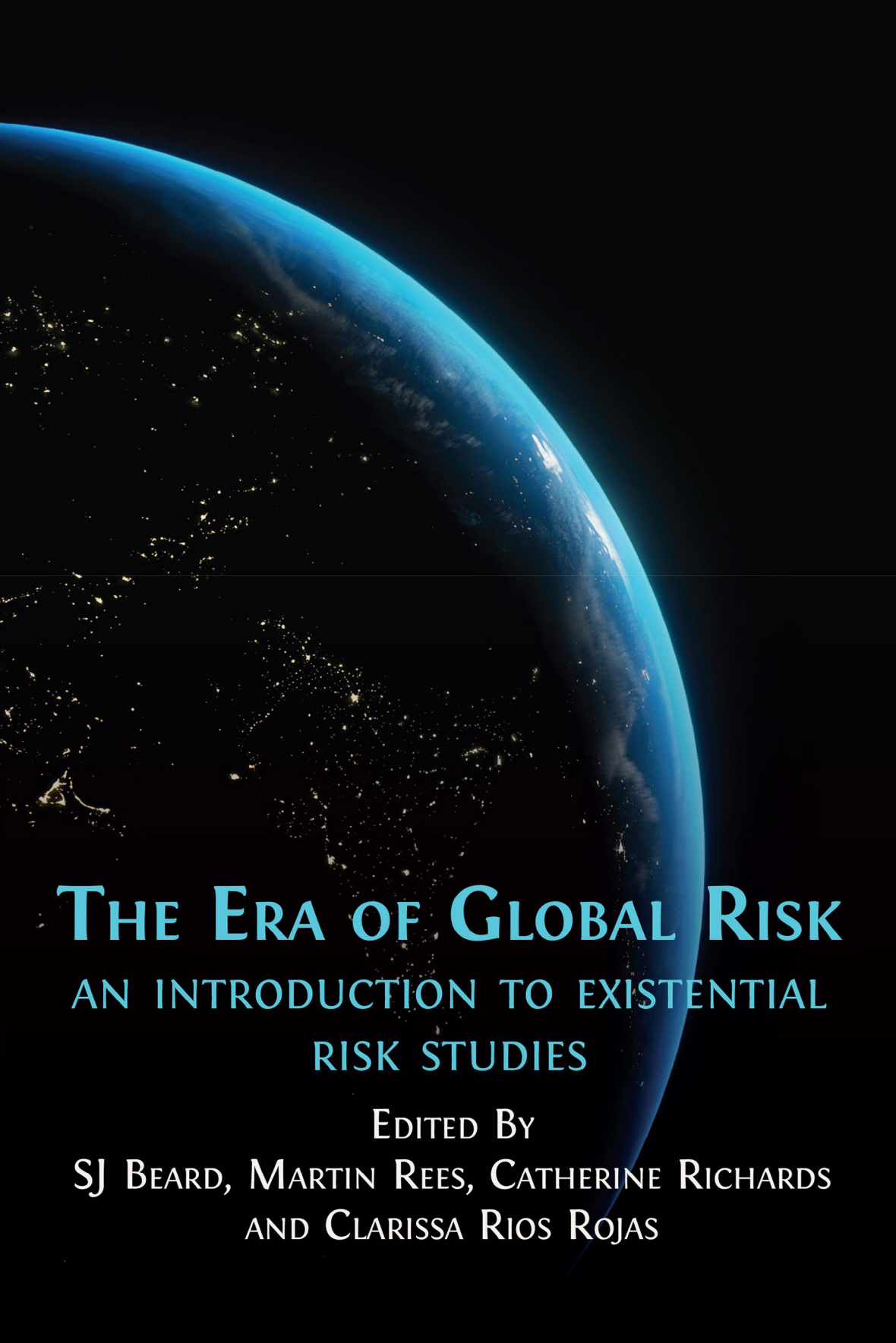

The Era of Global Risk

An Introduction to Existential Risk Studies

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

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

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.

Afterword

SJ Beard

This book was written as part of the 'Science of Global Risk' programme at the Centre for the Study of Existential Risk. The aim of this programme is to develop, implement, and refine a model systematic approach to addressing how global catastrophic and existential risk can best be identified, understood, managed, and mitigated. Many of the chapters it contains originated in sessions of an international conference hosted as part of this programme in 2020, which sought to survey the diversity of work already being undertaken in this field and position the centre's distinctive research in a constructive dialogue with a far wider community of researchers, decision-makers, and activists who were concerned about global risk.

The first five chapters of this volume developed arguments in relation to the study of global risk as an open and engaged field of academic study. Chapter 1 described how this study has long roots stretching back into the 20th century and beyond and has already made a number of significant contributions to making the world safer. Chapter 2 surveyed a range of methods that have been used to model the risks and causes of social and environmental collapse, and the directions in which these are currently being developed. Chapter 3 considered different approaches to the governance of risky scientific research and technological development, arguing for the importance of bottom-up—as well as more traditional top-down—approaches to this. Chapter 4 looked at how groups and individuals are contributing to the current level of global risk, and showed how the present reality of profound global injustice plays an important role in this causation. Finally, Chapter 5 looked at the increasingly important issue of diversity and inclusion,

 https://doi.org/10.11647/OBP.0336.11

and argued that within the field of global risk this is not only a pressing issue of justice and equity, but also of safety and resilience. Together, these chapters show how, even as it deals with unprecedented extreme future risks, our science can remain rooted in the real world and the realities of the here and now.

Building on these foundations, the remaining five chapters looked at a range of risk drivers (in 'Science of Global Risk' we often prefer to talk about individual risk drivers rather than individual risks, for the reasons I set out below). Chapter 6 looked at 'natural' risk drivers such as volcanoes and asteroids, and argued that even if these have anthropogenic causes, the current level of risk that we face is significantly determined by human-made vulnerabilities and exposures to such hazards. Chapter 7 considered more traditional anthropogenic environmental risk drivers, such as climate change, and argued that these undoubtedly contribute to the level of global risk, but that the traditional framing they have received in global risk research may be doing more harm than good with respect to our aim of understanding and addressing this. Chapter 8 provided a survey of recent developments in biotechnology with the potential to contribute to global risk, both positively and negatively, and argued that by enhancing existing science governance tools and mechanisms and introducing new risk management frameworks, these can be rendered far safer and more beneficial. Chapter 9 provided a contemporary history of thinking in the rapidly advancing field of AI safety, highlighting the shift away from safety concerns anchored only in the theoretical capabilities of AGI systems and towards thinking about the many challenges of aligning transformative AI with human society and human values. Finally, Chapter 10 considered the military use of AI, and argued that in recent years too much attention in this field may have been paid to the role of lethal autonomous weapons, and not enough to the connection of AI to nuclear weapons.

In their own way, each of these chapters shows us how this developing science is changing the way we understand and manage risks for the better by broadening the perspective of global risk researchers away from the most immediate hazards towards complex risk assessment, which also accounts for key vulnerabilities, exposures, and risk cascades, while still remaining relevant, and action guiding, to the most urgent problems facing humanity.

Together, then, these chapters highlight a field undergoing significant transformation and development. In previous work I have argued that this represents a natural result of previous developments in this crucial area of research.[1] While scholarship on global catastrophic and existential risk in the first decade of the 21st century was often marked by shared ethical and epistemological assumptions (such as commitments to utilitarianism, transhumanism, and Bayesian reasoning) that formed a coherent 'first wave' of research, a growing awareness of the scale of the challenges currently facing humanity—and a desire to harness additional resources to resolving these—led to a period of rapid expansion through the century's second decade. This saw the establishment of many new centres of research (including the Centre for the Study of Existential Risk itself) and the growth and development of existing centres through the generous support of committed funders. However, this in turn brought about a 'second wave' characterised by a far weaker set of assumptions and a growing desire to make research engaging for a wider audience, and to highlight the most urgent research and mitigation priorities. Yet, this second wave of research could never have been an end in itself, and as the field has continued to mature, its diversification and growth have brought in new research paradigms that may not only be based around different founding assumptions than those of previous waves of research, but also seek to articulate distinctive alternatives to the work that was already produced by them. This includes highlighting the importance of bottom-up as well as top-down governance mechanisms, global injustice as well as utilitarian efficiency, natural and environmental risk drivers as well as technological ones, and the importance of engaging with researchers who may not be personally committed to reducing global risks but whose work can nevertheless aid in this common endeavour.

Some key aspects of this emerging 'third wave' of research include a move from studying the aetiology of specific existential and global catastrophic risks (plural) to studying the interconnected drivers of existential and global catastrophic risk (singular); identifying the conditions and contexts within which risk is emerging and through which it can be managed; and working with experts from other disciplines, traditions, and cultures to achieve these aims. It is thus typified by the diversity of viewpoints on issues like how to classify

risks, what the best methods for studying them are, and how to evaluate different possible outcomes. However, in exploring this greater diversity, scholars who contribute to this third wave have often coalesced around the complexity of risk as a phenomenon that emerges from systems characterised by non-linear changes and feedback loops. This marks a shift away from focusing on the direct impacts of existential hazards to considering humanity's vulnerabilities and exposure to indirect and cascading impacts as well.

These are also the themes that come up time and again in the preceding chapters, from arguing that AI risk needs to be understood via the interrelation of AI with human systems or that asteroids and volcanoes are far more dangerous than they might be because of the design of critical infrastructure, to appealing to the value of engaging with existing models for studying social and environmental collapse or the need to understand individual motivations through the lens of global injustice. However, more than this, these themes can also be seen in how these issues sit in dialogue with each other, drawing on and developing different aspects of science governance, international security, environmental breakdown, and other complex phenomena to produce a systemic approach to identifying, understanding, managing, and mitigating global risk. Above all, I would argue that the sheer disciplinary diversity present in not just the authors and editors who have contributed to this volume (ranging from technical AI experts, engineers, and life scientists, to complex system modellers and volcanologists, to economists, lawyers, and philosophers, to the director of the Bulletin of Atomic Scientists and NASA's Planetary Defense Officer) but even more so to the almost astronomical diversity of works cited by them, encompassing almost every field of human knowledge in one form or another.

Of course, my way of thinking about the history of the science of global risk is by no means universally shared, and others have articulated different understandings of the field and its development.[2] However, to take the long view of our own long-termism—and begin to see this field as something that has a history, and even a sociology, behind it, as well as to recognise that the field is undergoing processes of development and change in sometimes contradictory directions—is itself a new and important idea that can help us to understand the work that we do and

how to improve it. And while I would not wish to foist such views on anyone, I sincerely hope that reading this book will have helped anyone interested in global risk to have improved not only their understanding of this risk, but also of how it is studied, how it can be reduced, and what role they might be able to play in achieving this goal.

Notes and References

1 Beard, S.J. and Phil Torres, *Ripples on the Great Sea of Life: A Brief History of Existential Risk Studies* (2020).

2 E.g. Moynihan, Thomas, *X-Risk: How Humanity Discovered Its Own Extinction*. MIT Press (2020); Cremer, Carla Zoe, and Luke Kemp, 'Democratising Risk: In Search of a Methodology to Study Existential Risk', *arXiv preprint arXiv:2201.11214* (2021).