A view of Earth from space, showing the curvature of the planet and a bright light source (the sun) just behind the horizon, creating a lens flare effect. The background is a dark, starry space.

Centre for the Study of Existential Risk



UNIVERSITY OF
CAMBRIDGE

A report prepared for CSER supporters

JUNE 2022



The University of Cambridge extends its sincere thanks for your support of the activities of the Centre for the Study of Existential Risk (CSER).

Supported by your generosity, the work of CSER researchers is increasing our understanding of, and preparedness for, existential threats to our world.

Contents

An introduction from Seán Ó hÉigartaigh	3
1. Update	4
2. People	5
2.1 New Staff	5
2.2 Visiting Scholars	5
3. Events, Engagement and Outreach	6
3.1 Academic engagement	6
3.2 Policy Engagement	7
3.3 Public Engagement	7
3.4 Events	8
4. Publications	9
4.1 Papers	9
4.2 Reports	11
Contact	13

Navigation

Scroll through the document, or click on the relevant section in the table of contents to go directly to that section. To return to the contents list, click the page number at the bottom of the page.

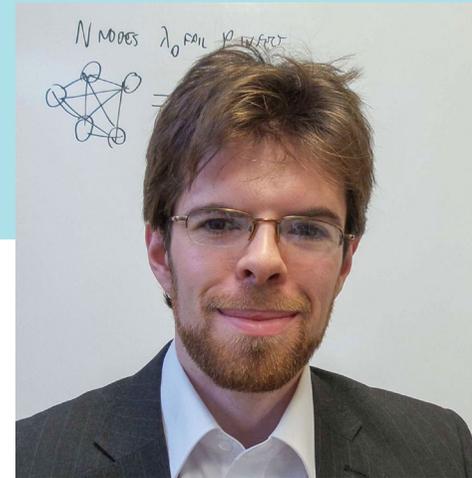


Click
to return to
contents

[Centre for the Study of Existential Risk Q1 2022](#)

An introduction from Seán Ó hÉigeartaigh

Executive Director, CSER



“

The Centre for the Study of Existential Risk (CSER) is an interdisciplinary research centre within the University of Cambridge dedicated to the study and mitigation of risks that could lead to civilizational collapse or human extinction. We are developing a science of global risk that we apply to Global Catastrophic Biological Risks, Extreme Risks and the Global Environment, and Risks from Artificial Intelligence. Our work is shaped around three main goals:

- **Understanding:** we study existential and global catastrophic risk.
- **Impact:** we develop collaborative strategies to reduce these risks.
- **Field-building:** we foster a global community of academics, technologists and policy-makers who share our goals.

Our last report covered October – December 2021. This report covers the period January – March 2022 and outlines our activities and future plans. Highlights of the last three months include:

- We **published five papers and four reports** – including one in *Nature Machine Learning* and another in partnership with the World Health Organization. These covered public health, planetary defence, global justice, the philosophy of Derek Parfit, AI in agriculture, AI risk management, the UK’s national resilience and future threats to global health, and biosafety.
- We hosted **three high profile public lectures**, with Andrew Leigh, Mark Lynas, and Paul and Scott Slovic, and two

interdisciplinary workshops, on scenarios for exploring and communicating extreme risks and future risks from the military applications of AI.

- Our **researchers presented their work at 10 workshops**, conferences, and talks hosted by, among others, the Center for War Studies, Oxford Earth Sciences, the International Institute for Applied Systems Analysis, Brunel University, and the Centre of Excellence for Development, Impact, and Learning.
- Our **work has also been featured in leading global media outlets**, including the BBC, Al Jazeera, *The Telegraph*, and Channel 4. This has included timely reaction and analysis around Russia’s invasion of Ukraine by our new Academic Programme Manager, Paul Ingram, who has been able to make use of his extensive knowledge and contact network in international security and disarmament.
- We **welcomed a new PA and administrative assistant** to support the work of our sustainable finance team as well as **three new visitors** working on x-risk communication, complex systems, and volcano risk.

1. Update

We continue to take advantage of the relaxation of COVID-19 restrictions to host more of our researchers in our office and begin to run public events again, while being sensitive to individual risks and needs. We are especially looking forward to hosting 80+ in person attendees, alongside a virtual audience of hundreds, to our next conference in April, and to hosting a group of 10 researchers at our Intensive Visitor Programme that will see the CSEER offices hosting a variety of talks and discussions in the two weeks immediately following this. We continue to plan towards our next stage of growth and the opportunities for new projects and research that more in-person working is making possible again.



2. People

2.1 New Staff

Over the last few months we have welcomed a new PA and administrative assistant to our sustainable finance team.

Matti Henderson (she/her) works as a personal assistant to Dr Ellen Quigley and as an administrative assistant within the Sustainable Finance team at CSER. Matti has a law degree from the University of Birmingham and is currently undertaking a MSc in Development Policy and Politics.



2.2 Visiting Scholars

We have also welcomed three new visitors to CSER during this period.

Georgiana Gilgallon (Gigi) is the Director of Communications at the Future of Life Institute, an independent non-profit that seeks to reduce large-scale, extreme risks from transformative technologies. Her focus in this role is improving media narratives around extreme threats to constructively influence society's perceptions of these risks, and thus increase its willingness to support mitigation efforts. As a Visiting Researcher at CSER, Gigi will work closely with Lara Mani, who is working to build an



empirical evidence base for communication techniques in the context of global risk. Gigi will be visiting from January 2022 to July 2022.

Arsène Pierrot-Valroff works on questions at the interplay between existential risks and complex systems studies; for instance, in a given network of societies, is a local civilizational collapse likely to trigger cascading failures. He has a background in mathematics (Ecole polytechnique, France), and a degree in astrophysics & aerospace engineering (ISAE-Supaéro, France). Arsène will be visiting from March 2022 to May 2022.



Mike Cassidy is a NERC research fellow specialising in volcano science at the University of Oxford. He has a background in Earth and Environmental sciences, with degrees from Bristol, Lancaster and Southampton. His work has taken him to volcanoes and research institutions around the world. At CSER he aims to investigate the global catastrophic impacts from large magnitude explosive eruptions with Lara Mani and others. Mark will be visiting from March 2022 to May 2022.



3. Events, Engagement and Outreach

3.1 Academic engagement

CSER researchers have presented our research at workshops, conferences, and talks hosted by the Center for War Studies, Oxford Earth Sciences, the International Institute for Applied Systems Analysis, Brunel University, and the Centre of Excellence for Development, Impact, and Learning, among others, as part of our ongoing efforts to engage a wide range of academics in the need to understand and mitigate the biggest risks facing humanity.

- 6 January: Matthijs Maas presented his paper “Into the Thick of It: Mapping the Emerging Landscape of Military AI Strategic Partnerships” at the AutoNorms / Center for War Studies (SDU) conference
- 24–28 January: Sabin Roman attended the Winter Workshops on Complex Systems at Saline Royale, Arc-et-Senans where he presented a tutorial on “Modelling the long-term evolution of societies”
- 1 February: Matthijs Maas took part in a panel on “Visions for the Future: Reflections on how to regulate cyberspace and AI in the future” at the CECS conference on Digitalization in Constitutional Law
- 1 February: Matthijs Maas presented a Conference paper on “Triaging the Technology Triad: Disruptive AI, Regulatory Gaps and Value Change” at the Conference ‘Workshop Machines of Change: Robots, AI and Value Change’
- 9 February: Lara Mani gave a talk at **Oxford Earth Sciences** on her research into volcanoes
- 8–10 February: Sabin Roman participated at the RISK-KAN workshop on ‘Understanding and Modelling Complex Risks in Coupled Human-Environment Systems’ where he presented his work on “Long-term feedback mechanisms underlying societal collapse”
- 25 February: Matthijs Maas gave a talk on “Artificial intelligence and international law: Charting a Regime Architecture for AI Governance” at the Asser Institute online winter academy 2022
- 21 March: SJ Beard and Alex McLaughlin spoke at the conference [Extinction: the Social, Ecological, Political, and Cultural implications of Extinction from the Microbial to the Planetary](#) hosted by Brunel University
- February – March: Sabin Roman conducted an extended visit to the **International Institute for Applied Systems Analysis** and the Vienna Complexity Hub in Vienna
- April 6: Lara Mani and Tom Hobson spoke at a webinar hosted by the Centre of Excellence for Development, Impact, and Learning on [The collaborative exploration of alternative futures: A different approach to Theories of Change](#)

3.2 Policy Engagement

Our researchers continue to engage with policymakers at a variety of levels, including the UK government and civil service, US National Institute of Standards and Technology, and World Health Organization.

- 24 January: Haydn Belfield, Matthijs Maas, Shahar Avin, and Seán Ó hÉigeartaigh responded to the **National Institute of Standards and Technology's** request for comments on its Concept Paper for AI Risk Management
- 7 February: Seán Ó hÉigeartaigh co-authored a briefing on Extreme Risks and UK National Resilience Strategy with Toby Ord (of the Future of Humanity Institute) and Angus Mercer (of the Centre for Long Term Resilience)
- 24 February: Haydn Belfield spoke at a Cambridge University Science Policy Exchange [workshop](#) on AI alignment, ethics, safety and governance
- 7 March: SJ Beard co-lead a training session for the UK Civil Service Reform Champions on Exploring Long Term Resilience, in partnership with the Centre for Long Term Resilience
- 17 March: Luke Kemp was lead author of a report by the **World Health Organization** setting out the results of its global health horizon scan, which he also helped to coordinate
- 29 March: Tom Hobson co-ordinated a joint submission with CSER researchers and colleagues to The Cabinet Office Enquiry on The Biological Security Strategy

3.3 Public Engagement

Our work has been showcased by leading global media outlets, including the BBC, Al Jazeera, *The Telegraph*, and Channel 4, and our researchers have also engaged with a range of public outreach talks and events.

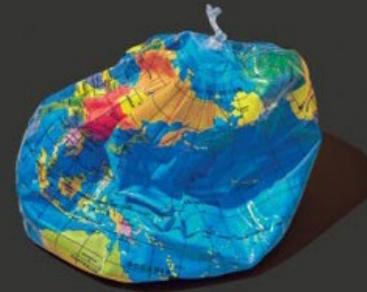
- 20 January: SJ Beard published [How to Read the Doomsday Clock](#) on **BBC Future**
- 21 January: Paul Ingram appeared on [Al Jazeera](#) TV to discuss the setting of the Doomsday Clock
- 24 January: SJ Beard gave a talk at the Cambridge Union on [Climate Change and Rights For Future Generations](#) hosted by the Cambridge University Liberal Association
- 6 February: SJ Beard appeared on the *21st Talks* Podcast to discuss [The Heart of Effective Altruism](#), the ethics and psychology of responding to existential risk
- 10 February: SJ Beard appeared on the BBC Radio 3 programme *Free Thinking* to discuss existential risk and the doomsday clock
- 25 February: Paul Ingram published **a response to Russia's invasion of Ukraine** and its relation to global catastrophic risk in [METRO](#)
- 5 March: Martin Rees spoke at a public symposium hosted by the Centre for Academic Research & Training in Anthropogeny on [Potential Utopian and Dystopian Futures](#)
- 5 March: Paul Ingram contributed to an article “Downing Street must ‘dust off Cold War plans and prepare for Russian nuclear strike’” in *The Telegraph*
- 19 March: Paul Ingram appeared on the programme *What If Putin Goes Nuclear?* with John Snow on Channel 4
- 20 March: Paul Ingram appeared on GB News's *The Discussion* to talk about the war in Ukraine

3.4 Events

We have hosted three public lectures, by Andrew Leigh, Mark Lynas, and Paul and Scott Slovic, which have each attracted hundreds of on-line attendees. We have also held two highly successful workshops on scenarios as tools to study and communicate about global catastrophic and existential risks and future risks from the military applications of AI.

- 17 January: Andrew Leigh gave a public lecture at CSER on [What's the Worst That Could Happen?](#), discussing his new book on existential risk and extreme politics
- 1 February: **Mark Lynas** gave a public lecture at CSER, [Don't Look Up: Is Climate Change and Extinction Level Event?](#), which was attended by more than 300 people (video [here](#))
- 1 and 2 March: Lara Mani hosted a highly successful virtual [workshop](#) titled "Scenarios as tools for exploring and communicating high-impact low-probability risks"
- 15 March: CSER hosted a public lecture by **Paul Slovic** and **Scott Slovic** on [The More Who Die, the Less We Care](#) (video [here](#))
- 24 March: The CSER Defence AI & Risk Workshop, organised with the **MoD's Defence Digital** team and Centre for Long-term resilience, brought together military and defence experts from the MoD, technical AI/ML safety experts, and academic and industry experts in extreme risks and military technology, to explore the principles and processes of assurance needed for the UK to address national security threats from AI-enabled adversaries and AI ethics, safety and security risks relating to its defence systems

WHAT'S THE WORST THAT COULD HAPPEN?



Existential Risk and Extreme Politics

Andrew Leigh

4. Publications

4.1 Papers

Several CSER researchers contributed to a solution scan on societal options to reduce the spread of respiratory viruses during the initial stages of the COVID-19 pandemic, and this has now been published as a paper in the *Journal of Biosafety and Biosecurity*.

[A solution scan of societal options to reduce transmission and spread of respiratory viruses: SARS-CoV-2 as a case study](#) in the *Journal of Biosafety and Biosecurity* 31 December 2021 [William Sutherland](#), Nigel G. Taylor, David C. Aldridge, Philip Martin, [Catherine Rhodes](#), [Gorm Shackelford](#), [S. J. Beard](#), [Haydn Belfield](#), Andrew J. Bladon, Cameron Brick, Alec P. Christie, Andrew P. Dobson, Harriet Downey, Amelia S.C. Hood, Fangyuan Hua, Alice C. Hughes, Rebecca M. Jarvis, Douglas MacFarlane, Silviu O. Petrovan

Societal biosecurity – measures built into everyday society to minimise risks from pests and diseases – is an important aspect of managing epidemics and pandemics. We aimed to identify societal options for reducing the transmission and spread of respiratory viruses. We used SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) as a case study to meet the immediate need to manage the COVID-19 pandemic and eventually transition to more normal societal conditions, and to catalogue options for managing similar pandemics in the future. We used a ‘solution scanning’ approach. We read the literature; consulted psychology, public health, medical, and solution

scanning experts; crowd-sourced options using social media; and collated comments on a preprint. Here, we present a list of 519 possible measures to reduce SARS-CoV-2 transmission and spread. We provide a long list of options for policymakers and businesses to consider when designing biosecurity plans to combat SARS-CoV-2 and similar pathogens in the future. We also developed an online application to help with this process. We encourage testing of actions, documentation of outcomes, revisions to the current list, and the addition of further options.

Lara Mani and fellow participants at the 7th Planetary Defence Conference published a [paper](#) in the *International Journal of Disaster Risk Science* on lessons in preparing for a Near-Earth Object (NEO) impact. The paper draws on a hypothetical NEO impact scenario that allowed the planetary defence community to discuss potential responses to a globally catastrophic impact event.

[When It Strikes, Are We Ready? Lessons Identified at the 7th Planetary Defense Conference in Preparing for a Near-Earth Object Impact Scenario](#) in the *International Journal of Disaster Risk Science* 7 January 2022 by Shirish Ravan, Tom De Groot, [Lara Mani](#), Einar Bjorgo, Richard Moissl, Jose Miguel Roncero, Katherine Rowan, David Schuld, Leviticus A. Lewis, Romana Kofler

Near-Earth object (NEO) impact is one of the examples of a high impact and low probability (HILP) event, same as the COVID-19 pandemic the world has faced since the beginning of 2020. The 7th Planetary Defense Conference held by the International

Academy of Astronautics (IAA) in April 2021 included an exercise on a hypothetical NEO impact event, allowing the planetary defense community to discuss potential responses. Over the span of the four-day conference this exercise connected disaster response and management professionals to participate in a series of panels, providing feedback and perspective on the unfolding crisis scenario. The hypothetical but realistic asteroid threat scenario illustrated how such a short-warning threat might evolve. The scenario utilised during the conference indicates a need to prepare now for what might come in the future, because even with advance notice, preparation time might be minimal. This scenario chose Europe for the impact, which may likely cope with such a disaster, through the Union Civil Protection Mechanism (UCPM) and other solidarity and support mechanisms within the European Union (EU), as well as with potential support from international partners. This short article raises concern about other areas in the world on how they may access NEO impact information and cope with such disasters. It also provides an idea on vast scale of such disaster vis-à-vis the current capacity of response systems to cope with a larger event in Europe or elsewhere. This scenario showed that planetary defense is a global endeavor. Constant engagement of the planetary defense and disaster response communities is essential in order to keep the world safe from potential disasters caused by NEO impacts.

Alex McLaughlin published a paper in the *European Journal of Political Theory* arguing that we ought to reject the integrationist attempt to unify our theorising between different domains of global politics in relation to justice issues such as climate change. The paper shows that integrationism is either trivial or it ignores the distinctive commitments held by its opponents and goes on to argue that the relevant disagreement is actually about the role of practices for political philosophy.

[Integrationism, practice-dependence and global justice](#) in the *European Journal of Political Theory* 17 January 2022 by [Alex McLaughlin](#)

An increasingly popular approach to global justice claims we should be 'integrationist,' where integrationism represents an attempt to unify our theorising between different domains of global politics. These political theorists have argued that we cannot identify plausible principles in one domain, such as climate justice, which are not sensitive to general moral concerns. This paper argues we ought to reject the concept of integrationism. It shows that integrationism is either trivial, or it obscures relevant disagreement by ignoring the distinctive methodological and substantive commitments held by its opponents. The paper then argues that the relevant disagreement is actually about the role of practices for political philosophy and, as such, should be framed in terms of the distinction between practice-dependent and practice-independent theory. Finally, Alex provides his own account of that distinction, identifying a practice-dependent claim that those concerned about the narrowness of prominent accounts of global justice should target.

SJ Beard and Patrick Kaczmarek published a [chapter](#) on the philosophy of Derek Parfit in his festschrift *Ethics and Existence*. In it, they argue that across the span of his work Parfit provided all the necessary components for his Theory X, a population axiology that was free of undesirable implications, and they show how these can be combined together and applied to the global challenges he argued matter most.

[On Theory X and What Matters Most](#) in *Ethics and Existence: The Legacy of Derek Parfit* ed. Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan (OUP) 24 January 2022 by [S. J. Beard](#), [Patrick Kaczmarek](#)

One of Derek Parfit's greatest legacies was the search for Theory X, a theory of population ethics that avoided all the implausible

conclusions and paradoxes that have dogged the field since its inception: the Absurd Conclusion, the Repugnant Conclusion, the Non-Identity Problem, and the Mere Addition Paradox. In recent years, it has been argued that this search is doomed to failure and no satisfactory population axiology is possible. This chapter reviews Parfit's life's work in the field and argues that he provided all the necessary components for a Theory X. It then shows how these components can be combined together and applied to the global challenges Parfit argued matter most: preventing human extinction, managing catastrophic risks, and eradicating global poverty and suffering. Finally, it identifies a number of challenges facing his theory and suggests how these may be overcome.

Asaf Tzachor, Shahar Avin, and Seán Ó hÉigeartaigh co-authored a perspective article in *Nature Machine Intelligence* on the risks and benefits of AI in global agriculture. The paper considers systemic risk factors of AI in agriculture and reviews risks relating to interoperability, reliability, and relevance of agricultural data; unintended socio-ecological consequences from machine learning models optimised for yields; and safety and security concerns. As a response, the authors suggest risk-mitigation measures.

[Responsible artificial intelligence in agriculture requires systemic understanding of risks and externalities in *Nature Machine Intelligence* 23 February 2022 by **Asaf Tzachor**, Medha Devare, Brian King, **Shahar Avin**, **Seán Ó hÉigeartaigh**](#)

Global agriculture is poised to benefit from the rapid advance and diffusion of artificial intelligence (AI) technologies. AI in agriculture could improve crop management and agricultural productivity through plant phenotyping, rapid diagnosis of plant disease, efficient application of agrochemicals and assistance for growers with location-relevant agronomic advice. However, the ramifications of machine learning (ML) models, expert systems and autonomous machines for farms, farmers and food security

are poorly understood and under-appreciated. Here, we consider systemic risk factors of AI in agriculture. Namely, we review risks relating to interoperability, reliability and relevance of agricultural data, unintended socio-ecological consequences resulting from ML models optimised for yields, and safety and security concerns associated with deployment of ML platforms at scale. As a response, we suggest risk-mitigation measures, including inviting rural anthropologists and applied ecologists into the technology design process, applying frameworks for responsible and human-centred innovation, setting data cooperatives for improved data transparency and ownership rights, and initial deployment of agricultural AI in digital sandboxes.

4.2 Reports

Haydn Belfield, Matthijs Maas, Shahar Avin, and Seán Ó hÉigeartaigh wrote a response to the National Institute of Standards and Technology's (NIST) request for comments on its Concept Paper for AI Risk Management.

[Response to NIST AI Risk Management Framework Concept Paper by **Haydn Belfield**, **Matthijs Maas**, **Shahar Avin**, **Seán Ó hÉigeartaigh** Published on 24 January 2022](#)

Summary of the response: We welcome the NIST AI Risk Management Framework Concept Paper, and commend NIST on your ongoing work to address risks in the design, development, use, and evaluation of AI products, services, and systems. We emphasise that the approach set out in the Concept Paper is the right one. With this framework, NIST can demonstrate global leadership and shape global standards. It will reduce harm, encourage adoption, and provide business certainty. **Our key piece of feedback is to keep this proposed framework and not water it down.**

Seán Ó hÉigearthaigh joined Toby Ord and Angus Mercer in writing a joint paper the UK's national resilience strategy and suggesting steps to improve resilience.

Extreme Risks and UK National Resilience Strategy by Toby Ord, **Seán Ó hÉigearthaigh**, **Angus Mercer** Published on 07 February 2022

A new joint paper by Future of Humanity Institute, CSER, and The Centre for Long-Term Resilience, discusses the UK's national resilience strategy and suggests steps to improve resilience. The report recommends a two-step process for boosting resilience based around domestic and international policy proposals.

Luke Kemp was lead author of a report by the World Health Organization setting out the results of its global health horizon scan. Dr Kemp was previously taken on as a technical consultant to run the horizon scan.

Emerging trends and technologies: a horizon scan for global public health by **Luke Kemp**, Kai Ilchmann, Soatiana Rajatonirina, Anna Laura Ross Published on 17 March 2022.

Luke's publication presents the findings of a global horizon scan on emerging technologies and trends relevant to global public health conducted in 2020 and 2021. We identified 15 new and emerging technologies and scientific advances that may have a significant impact on global health over the next two decades. WHO strives to remain "ahead of the curve" in relevant areas of research, science and technology in order to proactively identify, anticipate and shape issues that hold promise for prevention, diagnosis and treatment. The Global Health Foresight function was established in the WHO Science Division for this purpose and to assist Member States in building "futures thinking" into their strategic health planning. The horizon scan was organised

by the Emerging Technologies, Research Prioritisation and Support Unit of the Department, in particular by Luke Kemp (consultant), Kai Ilchmann, Soatiana Rajatonirina and Anna Laura Ross.

Tom Hobson co-ordinated a joint submission with CSER researchers and colleagues to The Cabinet Office Enquiry on The Biological Security Strategy.

Submission of Evidence to The Cabinet Office Enquiry on The Biological Security Strategy by **Tom Hobson**, **Lalitha S. Sundaram**, David Aldridge, Alec Christie, Brett Edwards, Malcolm Dando, Silviu Petrovan, Lijun Shang, **William Sutherland**, **Seán Ó hÉigearthaigh**, Rebecca Smith. Published on 29 March 2022

The authors are academic researchers in the field of biosecurity, bioethics, biodiversity, technology governance and policy. They have a long-standing interest in policy and practice of biological security in the UK and globally. In this submission they draw upon our collected expertise in technology and innovation governance, horizon scanning, expert elicitation, foresight methods and biosecurity education, in addition to knowledge of present and historical challenges and opportunities of biosecurity.



Contact

Elizabeth Brent

Associate Director — Arts and Humanities

University of Cambridge Development and

Alumni Relations

1 Quayside, Bridge Street

Cambridge

CB5 8AB

Elizabeth.Brent@admin.cam.ac.uk

+44 (0)1223 762891

www.cam.ac.uk/yourscambridge

Dr S.J. Beard

Academic Programme Manager and

Senior Research Associate

Centre for the Study of Existential Risk

16 Mill Lane

Cambridge

CB2 1SB

sjb316@cam.ac.uk

(+44) 01223 766838

www.cser.ac.uk



**UNIVERSITY OF
CAMBRIDGE**

**Dear World...
Yours, Cambridge**

The campaign for the University
and Colleges of Cambridge