



Quarterly Report – April 2017

Contents

1. Overview:	1
2. Recent appointments:	1
3. Recent workshops and conference:	2
4. Recent media coverage and public engagement:	4
5. Upcoming activities:	4
6. Funding	5
Appendix – Publications:	5

1. Overview:

We've had a busy Winter break and Lent term, in which we've:

- Hosted our first conference, the Cambridge Conference on Catastrophic Risk (CCCR), which took place on December 12-14th at Clare College (<http://cser.org/cccr2016/>).
- Organised a series of workshops on artificial intelligence risk scenarios, biorisks, geoengineering, and decision theory. These workshops have played important roles in both advancing our research projects, and in establishing strong links to relevant research and policy communities. The February AI and March 'extreme biorisk' workshop in particular allowed us to explore and identify research priorities in global-scale biorisk (natural and engineered).
- Co-organised the Darwin "Extremes" lecture series (organised by CSER's Julius Weitzdörfer) and a public lecture on climate change and existential risk (Professor Veerabhadran Ramanathan).
- Been featured extensively in the media, including on the front page of Wired UK.
- Published and submitted a number of papers as our research projects begin to bear fruit. This includes papers on geoengineering, on problems with current scientific structures, on biosecurity, on classifying global catastrophic risks, and several papers on decision theory.
- Made a new appointment to our team: Haydn Belfield (Academic Project Manager).
- Prepared to move office.
- Advanced research on our main project, Managing Extreme Technological Risks, and given talks on our research at a range of conferences and events in and outside Cambridge.
- Received a grant to expand our work on bio-risk and regulation, governance and responsible innovation in biotechnology.

Below, we give an update on our recent appointments, workshops, progress, and planned activities.

2. Recent appointments:

Our latest appointment has been for a second Academic Project Manager role, complementing Seán and Catherine's work, with a particular focus on outreach, communications, and building our policy and industry networks. This will also free up more time for Sean and Catherine to work with the



postdocs on CSER's current research projects and future proposals (with the aim of securing post-2018 CSER funding). Haydn Belfield has recently worked as a Senior Parliamentary Researcher and as a Policy Associate for the Global Priorities Project, where his work included co-writing a report on emerging technological risks for the Finnish Ministry of Foreign Affairs. He joined CSER in January 2017. New funding (two years) was secured for this post.

CSER has also welcomed seven new Research Affiliates. This affiliation recognises the contribution they have made to the success of the Centre. They are: Seth Baum, Ryan Carey, Nikita Chiu, Martina Kunz, William MacAskill, Ellen Quigley and Beth Barnes (see bios here: <http://cser.org/about/who-we-are/>).

3. Recent workshops and conference:

- December: **NIPS 2016 – Machine Learning and the Law Symposium**. Advances in AI mean that predictions and algorithms are already used in a wide variety of situations under regulatory or legal control. This includes such diverse areas as i.e. driving autonomous cars and predicting the likelihood of reoffending once released from prison. At this symposium, also sponsored by LCFI and Clifford Chance, key themes such as privacy, liability, and regulation of data were explored to inform the community of current and ongoing legislation as well as to help form better policies going forward. The Thirtieth Annual Conference on Neural Information Processing Systems - NIPS - is a multi-track machine learning and computational neuroscience conference. **Output:** Industry and academic contacts – this is the second well-attended NIPS symposium CSER has organised.
- December: **Cambridge Conference on Catastrophic Risk**. The conference was fully booked (110 participants) within one week after tickets were released. The call for papers resulted in 62 abstracts, and 19 were selected for presentation at the conference – alongside 11 high profile keynotes. The conference was divided into three main sessions: one on artificial intelligence, one on depreciation of earth systems and one on bio-risk. **Output:** Extended our network, increased our brand visibility and helped foster the community. Videos are now available online through <http://cser.org/cccr2016/> and Adrian Currie is editing a special issue of *Futures*.
- January: **“Black Sky” Infrastructure and Societal Resilience Workshop** This workshop at the Royal Society focussed on emerging black sky hazards – hazards that could severely disrupt the normal functioning of critical infrastructure (electrical, communications, etc.) in multiple regions for long durations. **Output:** the presentations are available [here](#).
- February: **AI and Bad Actors workshop (Shahar Avin, Sean O hEigeartaigh)** This workshop was run through the Strategic AI Research Centre (SAIRC - a joint FHI-CSER grant). It was co-organised by Shahar Avin but held at FHI. It brought together international industry and research leaders to discuss the risks of bad actors using AI and machine learning maliciously. Experts from cybersecurity, AI governance, AI safety, counter-terrorism and law enforcement discussed automated hacking, the use of AI for targeted propaganda, the role of autonomous and semi-autonomous weapons systems, and the political challenges posed by the ownership and regulation of advanced AI systems. **Output:** See next bullet point.
- February: **Envisioning and Addressing Adverse AI Outcomes (Shahar Avin, Sean O hEigeartaigh, Jaan Tallinn)** Held at Arizona State University, at Lawrence Krauss' 'Origins Project' – but CSER contributed extensively to the planning, organisation, research materials and chairing. It grouped participants into 'red-teams' and 'blue-teams' to explore various risky scenarios. **Output:** More and deeper contacts in the research, industry and governance communities. The outputs of the



two AI workshops will be consolidated into a research agenda for the field over the coming months and made available to the relevant research and policy communities.

- February: **Biological Extinction** Workshop at the Vatican on biodiversity and the Sixth Mass Extinction, organised by Partha Dasgupta and attended by Martin Rees. **Output:** Media coverage and contacts, also the proceedings are available [here](#).
- March: **Artificial Intelligence, Decision Theory, and Severe Uncertainty (Huw Price and Yang Liu)**. This workshop featured talks on rationality and future discounting, risk avoidance, and deciding with confidence. Yang and Huw's paper, Heart of DARCness, was presented.
- March: **Eighth Review Conference of the Biological Weapons Convention: Where Next? (Catherine Rhodes)**. This workshop brought together academics and decision-makers from international organisations to discuss a perceived lack of progress at the 2016 Review Conference, and to consider next steps. **Output:** Extensive contacts, and increased recognition of CSER, within the bioweapons community. Each session of the workshop aimed to identify a few 'next steps' for action that can be taken by participants to promote progress, particularly in the lead up to the 2017 Meeting of States Parties. Report in preparation.
- March: **Developing a Research Agenda and Methodologies for Extreme Bio-Risks (Lalitha Sundaram)**. The workshop asked industry leaders, technology investors and research leaders to discuss the most pressing research questions in 'extreme biorisk'. It built on our horizon-scan on emerging issues in bio-engineering. **Output:** It will inform work by Piers Millett (FHI) Lalitha Sundaram (CSER) and others to develop a research agenda for extreme bio-risks. This will aid us in identifying the key research questions and topics that need to be studied to understand biorisk on the global scale; in particular issues that are plausible but fall outside of the scope of existing biosecurity, biosafety and biotechnology governance programmes.
- April: **Geoengineering, ethics and society (Simon Beard, Adrian Currie)**. This workshop explored the history, governance and ethics of geoengineering research. It specifically focused on solar radiation management (SRM) – a controversial proposed technique which involves spraying sulphates into the atmosphere to increase the Earth's albedo, and therefore prevent temperatures from rising. **Output:** A collaborative paper looking at the risks and benefits of SRM research, development and deployment from an existential risk reduction perspective is planned.

Other activities:

- **Public lectures:** We suspended our regular lecture series in Lent term to avoid clashing with the Darwin 'Extremes' lecture series, organised by CSER researcher Julius Weitzdörfer. The series featured CSER-relevant talks on extreme climate change (Dr Emily Shuckburgh), engineering for extreme risk (Professor Nicholas Taleb, New York) and extreme politics (Professor Matthew Goodwin, Kent). The Darwin series is the best-attended lecture series held in Cambridge. We co-hosted one additional lecture, on "Climate change: morphing into an existential threat" by Professor Veerabhadran Ramanathan, in collaboration with the Centre for Science and Policy and Christ's College (<http://www.csap.cam.ac.uk/events/climate-change-lecture-7/>).
- **Policy and standards in AI:** We've been working with the Centre for the Future of Intelligence on policy and industry engagement in artificial intelligence, including with industry leaders at the Asilomar 2017 Beneficial AI Conference, and with policy-makers at the European Commission, UK Parliament's Science and Technology Committee and the IEEE Standards Association (the latter has resulted in the following report to which Seán Ó hÉigeartaigh, Shahar Avin, and Huw Price have contributed: http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf).
- **Working with Cambridge students:** We have continued to support Cambridge's excellent 'Future of Sentience' student society, established last year by our research affiliate Beth Barnes; in Lent



term we hosted a weekly research discussion and work session with them that was well-attended, and have involved them in a joint project to develop an online 'wiki' of knowledge on global catastrophic risks.

- **Academic engagement:** The CSER team has presented work at APA Pacific, the Geneva Centre for Security Policy, Exeter University, Columbia University, the UCLA Garrick Institute for the Risk Sciences, the BBC Festival of Ideas, MIT, the Finnish Embassy, University of Bristol, and has been represented at SynBioBeta, and Skoll World Forum on Social Entrepreneurship amongst other events.

4. Recent media coverage and public engagement:

- CSER was featured as a front page [cover story](#) for WIRED UK.
- CSER and Sean were featured in [XL Semenal](#), Spain's largest magazine, following the publication of his book chapter [Technological Wild Cards: Existential Risk and a Changing Humanity](#).
- Tatsuya's [paper](#) on the risks of English being the lingua franca of research was featured in [the Economist](#).
- Shahar wrote a piece in Aeon: [Science funding is a gamble so let's give out money by lottery](#) that has been shared over 400 times on Facebook.
- Martin was featured in the [Times Higher Education](#) Supplement, [Huffington Post](#) and [The Conversation](#) – which got picked up by many national papers.
- Huw wrote a follow-up [Aeon piece](#) to his much-discussed [cold fusion essay](#).
- Simon has been selected as a '[New Generation Thinker](#)' by the BBC and the Arts and Humanities Research Council – he will now have the opportunity to create radio and TV programmes on his research.

5. Upcoming activities:

Easter Term:

- **April 28th-30th: Risk and the Culture of Science (Adrian Currie).** The initial aim is to bring communities with interests and expertise relevant to risk, scientific 'mavericks', and the culture of science together, to make connections and identify common goals. This is intended to be the first step in a practical project: the design and implementation of a 'safe space': the maverick room. CSER's motivation is the study of new technological risks, but we intend the results to have relevance beyond that focus. [Event link](#).
- **Research visit** for four weeks by Heather Douglas, Chair in Science and Society, University of Waterloo, who will give our Blavatnik public lecture on 29th April on 'Responsibility and Inequality in a Risky World'. [Event link](#).
- Additional research visits by Dr Tamiko Nakamura (attorney; nuclear law), Professor Kristian Lauta (law and existential risk; Copenhagen); Seth Baum (Executive Director, Global Catastrophic Risk Institute).
- **May 2nd:** Professor Terrence Sejnowski, the President of the NIPS Foundation, will give a public lecture on deep learning on 2nd May. Over 400 people have already claimed tickets. [Event link](#).
- **May 8th:** Professor J. Doyne Farmer, Director of the Complexity Economics program at the Institute for New Economic Thinking at the Oxford Martin School, will give a public lecture on 'Collective awareness: A vision of a new economics and how it could reduce risk' on 8th May. [Event link](#).



- **May 8th-10th: Climate Ethics and Climate Economics: Risk, Uncertainty and Catastrophe Scenarios (Simon Beard).** This is the fifth of six ESRC-funded workshops exploring issues where the ethics and economics of climate change intersect. This one will address questions such as: How should we act when we believe that there is a chance of a catastrophe, but cannot make reliable probability estimates? How much should we worry about worst-case scenarios? What should we do when experts disagree about whether catastrophe is possible? [Event link.](#)
- **May 10th:** Professor Hilary Greaves will give a public lecture on ‘Overpopulation: A driver of climate change?’ on 10th May. [Event link.](#)
- **May 30th:** We will collaborate with the Royal Society and the Centre for the Future of Intelligence on a workshop at the Royal Society on lessons in public engagement surrounding transformative technologies.

Summer:

- **June 21st CSER post-doc research showcase:** we will organise a day-long event on 21st June for the post-docs to present their work to an expert group.
- **July 12th-14th CFI conference:** Our sister centre, the Centre for the Future of Intelligence, will hold its first Conference from July 12th-14th, with multiple CSER researchers participating.
- **July 28th-30th Conference on Decision Theory and the Future of Machine Intelligence (DTFMI)** which will be a joint effort between CSER, CFI and our partner at MCMP headed by Yang Liu. This is expected to be a long run conference series for years to come, but the inaugural meeting in 2017 will be held at Cambridge and financially supported by CSER (through the Templeton grant). [Event link.](#)

6. Funding

We are tremendously grateful to the Schmidt Sciences for their grant to support research on biorisk and responsible development in synthetic biology. The fund will be used to support the research of Dr Lalitha Sundaram as well as some group projects.

Appendix – Publications:

In addition to the papers below, a further dozen or so papers (on topics including problems with current scientific structures, classifying global catastrophic risks, and decision theory) have been submitted or are at the 'revise and resubmit' stage with journals.

Published papers and chapters are available here: <http://cser.org/publications/>

Date	Name	Deliverable	Type
2017-04	Adrian Currie	Paper published: <i>Accelerating the Carbon Cycle: The Ethics of Enhanced Weathering in Biology Letters.</i>	Published
2017-03	Adrian Currie	Paper published: <i>In Defence of Story-telling in Studies in History & Philosophy of Science</i>	Published
2017-02	Catherine Rhodes	Book chapter accepted (forthcoming): <i>Scientific Freedom and Responsibility in the Biosecurity Context in Scientific Freedom:</i>	Accepted



		An Anthology (Manchester University Press).	
2017-01	Bonnie Wintle	Paper published: <i>Metaresearch for Evaluating Reproducibility in Ecology and Evolution</i> in BioScience.	Published
2017-01	Sean O'hEigeartaigh	Book chapter published: <i>Technological Wild Cards: Existential Risk and a Changing Humanity</i> In <i>The Next Step: Augmented Humans and Exponential Life</i> (BBVA)	Published
2016-12	Tatsuya Amano	Paper published: <i>Languages are still a major barrier to global science</i> in PLOS Biology.	Published

We would like to thank the following generous benefactors for their support:

- Templeton World Charity Foundation
- Grantham Foundation for the Protection of the Environment
- Hauser-Raspe Foundation
- Blavatnik Foundation
- Kenneth Miller Trust
- Schmidt Sciences
- Libra Foundation
- Musk Foundation
- Milner Foundation
- Jaan Tallinn
- Other generous philanthropists