

**Impact case study (REF3b)**

|   |
|---|
| <b>Institution:</b> University of Oxford  |
| <b>Unit of Assessment:</b> 32   |
| <b>Title of case study:</b> Predicting and Understanding Risks to Our Future Life   |
| <p><b>1. Summary of the impact</b> (indicative maximum 100 words)</p> <p>For about a decade, Professor Nick Bostrom and others have been pursuing research on what he calls ‘existential risk’: this research deals with basic threats both to the quality of our future life and indeed to our having any future life at all. This work has had considerable impact on policy. Professor Bostrom has been invited to play a large number of advisory and consultation roles, to government departments and major insurance companies among many others. His work has also attracted a huge amount of attention among the wider public. He has been invited to give prestigious public lectures, and he has given many interviews on his ideas to the media – thereby contributing to the public awareness of the huge risks at stake.</p>  |
| <p><b>2. Underpinning research</b> (indicative maximum 500 words)</p> <p>Professor Nick Bostrom and others have been pursuing research on what he calls ‘existential risk’ for about a decade. The period since 2006, when he became Director of the Future Of Humanity Institute in the Oxford Martin School, has seen significant developments in this research. Indeed, one of the founding aims of the Institute was precisely to bring interdisciplinary expertise to bear on the research in an attempt to tackle questions of crucial importance to humanity that had previously received remarkably little attention within academia.</p> <p>Professor Bostrom defines an ‘existential risk’ as a certain sort of global catastrophic risk: one that threatens the premature extinction of Earth-originating intelligent life or the permanent and drastic destruction of its potential for desirable future development. His research includes investigation both of potential causes of such risks and of biases and other flaws in how we think about them.</p> <p>In the 2005 article ‘How Unlikely is a Doomsday Catastrophe?’, which Professor Bostrom co-authored with Max Tegmark, the authors consider such extreme existential risks as a breakdown of a metastable vacuum state triggered by a microscopic black hole. They argue that many previous bounds on the frequency of such events give a false sense of security due to a failure to take into account the fact that observers are by definition in places lucky enough to have avoided destruction: the authors derive their own new upper bound. In his 2011 article ‘Information Hazards’ Professor Bostrom explores a quite different type of risk. By an ‘information hazard’ he means a risk that arises from the dissemination or the potential dissemination of information that may directly or indirectly cause harm. He argues that such hazards are often much more subtle than physical threats, and that it is therefore easy for us to overlook them, though they can be just as serious as physical threats. In ‘Anthropic Shadow’, Professor Bostrom joins forces with Dr Milan Ćirković and Dr Anders Sandberg to explore limitations on our ability to assess risks. In particular, they focus on an observation selection effect—what they call the ‘anthropic shadow’—that prevents us from observing certain sorts of catastrophes in our recent geological and evolutionary past. The significance of this is that it creates a kind of second-order risk, namely the risk of our significantly underestimating the risk of any type of catastrophe that lies within the shadow.</p> <p>In his 2009 article ‘The Future of Humanity’, Professor Bostrom explains more generally the importance of our considering the risks that attend our own future. He first raises some very broad issues about the nature of predictability, drawing attention to ways in which the long-term future can be easier to predict than the short-term future, then looks in particular at the question of whether our long-term future can be expected to be ‘post-human’, in the sense, roughly, that it can be expected to involve changes of degree in us profound enough to constitute changes of kind. He considers ways in which the assumptions that we make about these matters shape the decisions that we make, both in our personal lives and in public policy, and he highlights the very unfortunate consequences that sometimes accrue.</p> |

## Impact case study (REF3b)

Professor Bostrom and Dr Milan Ćirković are also joint editors of the 2008 collection *Global Catastrophic Risks*, to which they contribute an introduction. This collection explores a wide range of existential risks currently facing us, including those associated with natural catastrophes, nuclear war, the use of biological weapons, terrorism, global warming, advanced nanotechnology, artificial intelligence, and social collapse.

Professor Bostrom has been Director of the Future of Humanity Institute and of the Programme on the Impacts of Future Technology since their inception as part of the Oxford Martin School in 2005, having for the two previous years been a postdoctoral research fellow in the Faculty of Philosophy. Milan Ćirković has been a Research Associate at the Future of Humanity Institute since 2008. Anders Sandberg has been a James Martin Research Fellow at the Future of Humanities Institute and the Programme on the Impacts of Future Technology since 2005.

### 3. References to the research (indicative maximum of six references)

- Max Tegmark and Nick Bostrom, 'Astrophysics: How Unlikely is a Doomsday Catastrophe?' in *Nature* **238** (2005), 754. (An expanded version of this article appears on Professor Bostrom's personal website: <http://www.nickbostrom.com>.) doi:10.1038/438754a
- Nick Bostrom and Milan Ćirković (eds), *Global Catastrophic Risks* (Oxford: Oxford University Press, 2008). Can be provided on request.
- Nick Bostrom, 'The Future of Humanity', in Jan-Kyrre Berg Olsen, Evan Selinger and Soren Riis Aldershot (eds), *New Waves in Philosophy of Technology* (Basingstoke: Macmillan, 2009). Can be provided on request.
- Nick Bostrom, Milan Ćirković and Anders Sandberg, 'Anthropic Shadow: Observation Selection Effects and Human Extinction', in *Risk Analysis* **30** (2010), 1495-506. doi: 10.1111/j.1539-6924.2010.01460.x
- Nick Bostrom, 'Information Hazards: A Typology of Potential Harms From Knowledge', in *Review of Contemporary Philosophy* **10** (2011), 80-100.  
<http://www.cceol.com/asp/getdocument.aspx?logid=5&id=6abb8f35-3e01-4bbc-a6fc-1b27a78c4be4>

The quality of this research is evidenced in each case by the place of publication. The peer-reviewed journals and publishing houses concerned do not publish work that is not of internationally recognized quality. The third article listed, 'Anthropic Shadow', won the award for the best paper of the year by the journal editors.

### 4. Details of the impact (indicative maximum 750 words)

Professor Bostrom's work has led to his being invited to play a large number of advisory and consultation roles, to government departments, major insurance companies, and other institutions, and this in turn has led to his work's having a large impact on policy. In 2009, he advised Booz Allen, a strategy and technology consulting firm. That same year, he advised a joint meeting of the Royal Society and the International Council of Life Sciences on biomedical risk: the Royal Society published a report from that meeting which was subsequently circulated to the Presidential Commission for the Study of Bioethical Issues. In the transcript of the meeting of the Commission at which the report was circulated, and during which Professor Bostrom was several times cited, the concluding summary states that '[this] was a day... for the Commission that was rich in input... [We] heard a lot... about value and values... [including] the value of understanding risk... [to] many things that we have been deliberating as a Commission.'<sup>[i]</sup> Professor Bostrom served as an expert member of the World Economic Forum's Global Agenda Council on Catastrophic Risks from 2010 to 2011, and the council went on to produce a series of proposals, including the proposal that a 'transnational... mechanism that can identify and work to prevent, anticipate and prepare for catastrophic risks' be established.<sup>[ii]</sup> In 2010, he advised: the US State Department's Global Futures Forum; Amlin Insurance (the Aggregate Modelling Division); BAE Systems (the Global Combat Systems, Land and Armaments Divisions); and the Science for Humanity 'Global Risk Register', which uses expert advice in an attempt to achieve an appropriate management of risks

## Impact case study (REF3b)

and controls, and which Baroness Susan Greenfield, Trustee of Science for Humanity, describes as destined to have 'a positive impact on the economy as well as humanity'<sup>[iii]</sup>. In 2011, he advised the MacArthur Foundation. And that same year he advised Digital Sky Technologies concerning talking points for a roundtable discussion between its Chief Executive Officer, Yuri Milner, and the leaders of the G8 countries for the 37<sup>th</sup> Annual G8 Leaders Summit in Deauville. The advice that he gave Amlin Insurance impressed them to such an extent that they subsequently contributed £900,000 towards funding a collaborative research project between them and the Future of Humanity Institute to investigate systemic risk.

Professor Bostrom has marshalled his ideas to give practical advice in several other ways. He is on the Advisory Board of the Singularity Institute, which aims to bring rational analysis and rational strategy to the challenges facing humanity and which holds an annual summit to coordinate and educate those concerned with these challenges.<sup>[iv]</sup> He co-founded the Institute for Ethics and Emerging Technology, whose mission is to be a centre for voices arguing for a responsible approach to the most powerful emerging technologies.<sup>[v]</sup> He also founded the Existential Risk Reduction Career Network, a community whose members discuss the strengths and weaknesses of different careers in terms of existential risk, share advice on job applications and career advancement, help one another to find interviews, and suchlike.

In July 2008, Professor Bostrom organised a conference entitled 'Global Catastrophic Risks', concerned with the various ideas that were eventually published in that same year in the collection of the same name.<sup>[vi]</sup> Several of the conference's participants had significant non-academic profiles. They included: Joseph Cirincione, President of the Ploughshare Fund, a foundation dedicated to the reduction of nuclear proliferation; Chris Phoenix and Mike Tredor, co-founders of the Centre for Responsible Nanotechnology, an organisation whose mission is to raise awareness of, and to expedite the thorough examination of, the benefits and dangers of advanced nanotechnology, and to assist in the creation and implementation of the responsible use of such technology; and Sir Crispin Tickell, Director of the Policy Foresight Programme at the James Martin School. There was extensive coverage of this conference in *New Scientist*, *Reason Online*, *The Scotsman*, *Earth and Sky*, and CNN.

A significant additional part of the impact of Professor Bostrom's research derives from the way in which he has drawn the issues of existential risk to people's attention. (For instance, his work has inspired the setting up of an existential risk research centre in Cambridge, for which he serves as an external advisor.) He has disseminated his ideas widely through the media in a way that has attracted considerable public debate. Among the highly distinguished public lectures and other public presentations that he has given on these issues since 2008 are the following:

- a keynote address to the Global Catastrophic Risks Conference in 2008;
- the opening presentation at the Policy Foresight Programme Workshop in 2008;
- a discussion with Sir Martin Rees at the Science Foo Camp, organised by *Nature*, Google, and O'Reilly Media in 2008;
- the opening keynote address to the Guardian Activate Summit in 2009;
- a talk to the Chancellor's Court of Benefactors in 2009;
- a talk to the Future Scenarios Seminar run jointly by the International Centre for Community Development, the European Centre for Jewish Leadership, and the American Jewish Joint Distribution Committee in 2010;
- a panel contribution to the Leaders of Change Summit in the Istanbul World Political Forum in 2011;
- a keynote address at a Global Futures Forum workshop on transformational technologies, partly under the auspices of the US Department of State and the Italian Intelligence Community, in 2012.

Since 2008, Professor Bostrom has also been interviewed on the issues of existential risk by (among others) the Discovery Channel for Canadian television, *New Scientist*, *The Scotsman*, *Epoca* (the premier Brazilian news weekly), *Kultur Zeit* for German television, *Time*, *Die Zeit*, the History Channel and Bloomberg News for American television, *Time Magazine*, Bayerischer Rundfunk for German radio, ABC, *i* (a Portuguese newspaper), *World Affairs Monthly*, *Il Sole 24*

## Impact case study (REF3b)

Ore (an Italian monthly magazine), Utbildningsradion (a Swedish Educational Broadcasting Company), *The Times Online Magazine*, *The Guardian*, *Science & Vie*, *Scientific American*<sup>[vii]</sup>, *Bloomberg Businessweek*,<sup>[viii]</sup> *The Economist*, the BBC (television and radio), *Aeon* (a digital magazine with an online blog which in this case attracted very many enthusiastic comments) and Huffpost Line<sup>[ix]</sup> (an online journal that reaches 27 million monthly viewers) as well as in the documentary film *Transhuman* by Titus Nachbauer<sup>[xi]</sup>. An interview for *The Atlantic*,<sup>[xii]</sup> despite being over 5,000 words long, received over half a million reads, was shared over 6,000 times in social media (Google+, Facebook, LinkedIn, and Twitter) and made the front cover of Reddit. An interview for *Aeon*<sup>[xiii]</sup> was shared nearly 4,500 times in social media. An article on the BBC News website in April 2013,<sup>[xiv]</sup> following on from interviews for Radios 4 and 5, was the most read story of the day, with 1.2 million readers for that day, and, was shared over 13,000 times in social media.<sup>[1]</sup>

### 5. Sources to corroborate the impact (indicative maximum of 10 references)

#### Testimony

[1] Written testimony from Education Correspondent, BBC News

#### Other evidence sources

[i] The report of the Royal Society can be found at:

[http://royalsociety.org/uploadedFiles/Royal\\_Society\\_Content/policy/publications/2009/7860.pdf](http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/publications/2009/7860.pdf), and the transcript of the meeting at which it was circulated to the Presidential Commission is at: <http://bioethics.gov/cms/node/175>.

[ii] The proposals of the World Economic Forum's Global Agenda Council on Catastrophic Risks can be found at: <https://members.weforum.org/pdf/globalagenda2010.pdf>. The material quoted is on p. 189.

[iii] The page of the website of the Global Risk Register, on which there is a video link to Professor Bostrom's discussion of catastrophic risk and from which the quotation from Baroness Susan Greenfield is taken is: <https://www.globalriskregister.org/grr-news-press.html>

[iv] The website for the Singularity Institute is: <http://singinst.org/>

[v] The website for the Institute for Ethics and Emerging Technology and for the Existential Risk Reduction Career Network are: <http://www.ieet.org/> and <http://www.xrisknetwork.com/>.

[vi] The website for the conference 'Global Catastrophic Risks' is: <http://www.global-catastrophic-risks.com/programme.html>

Particularly significant media presentations include the following:

[vii] <http://www.scientificamerican.com/blog/post.cfm?id=technogenic-disasters-a-deadly-new-2011-07-06.html>

[viii] [www.businessweek.com/magazine/guardians-of-the-apocalypse-12152011.html](http://www.businessweek.com/magazine/guardians-of-the-apocalypse-12152011.html)

[ix] <http://www.economist.com/blogs/democracyinamerica/2011/05/futurology>

[x] <http://live.huffingtonpost.com/r/segment/extinction-level/512149972b8c2a4572000687>

[xi] [www.transhumandoc.com](http://www.transhumandoc.com)

[xii] <http://www.theatlantic.com/technology/archive/2012/03/were-underestimating-the-risk-of-human-extinction/253821/>

[xiii] <http://www.aeonmagazine.com/world-views/ross-andersen-human-extinction/>

[xiv] <http://www.bbc.co.uk/news/business-22002530>