

A conversation with Seán Ó hÉigeartaigh on 24 April 2014

Participants

- Dr. Seán Ó hÉigeartaigh—Senior Academic Manager for the Future of Humanity Institute and its constituent programmes (FHI core research, Programme on the Impacts of Future Technology, FHI-Amlin Collaboration on Systemic Risk) , and academic project manager of the Cambridge Center for the Study of Existential Risk (CSER)
- Nick Beckstead—Research Fellow, Future of Humanity Institute at Oxford University; Board of Trustees, Centre for Effective Altruism

Background

The Future of Humanity Institute and the Oxford Martin Programme on the Impacts of Future Technology are closely related projects, both directed by Nick Bostrom and both housed in the same location. For the purposes of this conversation, I refer to both as FHI.

Summary

Purpose of the call: I organized this call to learn about FHI’s room for more funding, CSER’s room for more funding, and CSER’s current state of operations.

Why this person: Sean has responsibility for managing much of the day-to-day operations of FHI and helping to get CSER operating, as well as for mid- to long- strategy and development of the FHI and its constituent programmes (in conjunction with the Director).

Regarding FHI, we discussed FHI’s most important achievements over the last few years, FHI’s current funding situation, how much funding FHI could productively use, what FHI would do with different levels of additional funding, and what FHI does to report on its output. FHI would use small amounts of funding (around £60,000) to expand its reserves and do other small projects. Larger amounts of funding would be used to hire “core operations” staff (such as a project manager, research assistants, and administrative staff), free up researchers operating under constraints from funders, extend researcher contracts, and hire additional researchers.

Regarding CSER, we discussed what is currently being done at CSER, CSER’s current funding situation, how much additional funding CSER could productively use, and what CSER would do with different levels of additional funding. CSER might use small amounts of funding (around £60,000) to supplement a recent 3 year donation to the level required to fund a postdoc, to hire project assistance to help in getting the project fully off the ground, and to fund more activities. Larger amounts of funding could be used to follow an already-drafted proposal for research.

Note: I, Nick Beckstead, work for FHI. I have tried to be objective, but I am likely to have biases in favor of asking easy questions of FHI and CSER. I believe that most of my questions are highly relevant and pretty purely factual, so I see this as less of a drawback than it may appear at first. But people reading these notes should keep my potential biases in mind.

FHI questions

What have FHI's most important achievements been over the last few years?

In the last few years, the main achievements include:

1. Nick Bostrom has written a book on superintelligence, which will be published by Oxford University Press, and has published the paper "Existential Risk Prevention as Global Priority" in *Global Policy*.
2. Bostrom, Anders Sandberg, Stuart Armstrong, Daniel Dewey, Seán Ó hÉigearthaigh and Carl Shulman have made important incremental research progress on AI risks and trajectories, and have also raised awareness of AI risk in academia and industry; the latter has led to the active involvement of Stuart Russell, among others.
3. Toby Ord and some other FHI staff have been advising governments and other policymakers on a variety of issues related to the future of technology and global health. This has resulted in the "Unprecedented Technological Risks" report, well-received and shared within the UK government. Sean Ó hÉigearthaigh has also been advising and establishing connections with policymakers and strategists in EU governments.
4. Very generally, FHI has helped create legitimacy for and interest in the idea of existential risk reduction. This has led to the field growing extensively, and the establishment of two similar high-profile centres at Cambridge and MIT, both advised and supported by FHI members and associates.

Further back, some standout achievements include:

1. In 2008, publishing the edited volume *Global Catastrophic Risks* with OUP, and holding the Global Catastrophic Risks Conference in Oxford.
2. In 2008, publishing *Whole Brain Emulation: A Roadmap* as an FHI technical report.
3. In 2010, publishing "Probing the Improbable" in the *Journal of Risk Research*.
4. In 2005, Bostrom and Tegmark's "Is a doomsday catastrophe likely?" in *Nature*.
5. Other published work on AI risk, and a large body of work on technological risk, technological prediction, the Fermi paradox, human enhancement, brain emulation and the Simulation Argument (arguably defining and leading the latter 3 fields).

Room for more funding

Pure finance issues

What is FHI's annual budget?

In the last year, FHI had an annual budget of about £700,000. This directly funds 12 staff members (9 researchers, 2 project managers (who contribute to research), 1 administrator) and additionally supports 4 actively participating research associates, several regular collaborators, and all institute activities (conferences, outreach, active visitor programme).

The large majority of FHI funding comes from industry and academic funders (>80%) in the form of two large projects; one of these projects continues until August 2015, the other until April 2016. The FHI has no guaranteed grant income beyond this stage, but is expected to have success in funding applications in the meantime. The remainder comes from a combination of one large philanthropic donation from Alexander Tamas for Daniel Dewey's fellowship (until summer 2015), and a number of small donations from other philanthropic supporters.

How much does FHI have in reserves?

FHI has about £100,000 in reserves and "general purpose" funds – this unconstrained funding is of extremely high value to us.

What pending grant applications does FHI currently have?

FHI gets most of its funding from institutional sources. FHI has recently had success with a Leverhulme grant application on population ethics that will fund Toby Ord, a fraction of Hilary Greaves, and a new researcher to be hired. A small joint AHRC (Arts & Humanities Research Council) grant application is at the final round of assessment – if successful it will fund a fraction of Anders Sandberg's salary, and a postdoc at 50% for 1 year. Several further grant applications are planned.

Is FHI actively seeking funding from non-institutional donors?

FHI does not spend much time fundraising from non-institutional donors, but welcomes such donations.

What does a researcher-year cost FHI?

A portion of the cost of a researcher year must be paid to Oxford University for access to university services and affiliation. This includes an £8K university infrastructure charge, plus the cost of office rent and enough to cover the university's pension scheme and UK health insurance (national insurance). Please note that while Full Economic Costing (FEC), which makes a researcher year considerably more expensive, is required for research council-funded researchers, we can usually waive this for philanthropically-funded positions.

For salary, national insurance, pension, office rent, and bare essentials, a researcher year costs around £55,000. If you include travel funds, conference/workshop funds and a fraction of additional admin and supervisor time required to support research staff and publicise their research, it becomes around £77,000.

Other issues

What is typical output for a researcher-year?

About four single- or main-author published research papers, plus some amount of policy and media engagement, depending on which researcher it is, and what the specific research topic is.

To what extent are FHI researchers able to focus on what they believe is most important?

Daniel Dewey and Nick Bostrom can almost entirely focus on what they believe is most important. Other researchers at FHI can largely focus on what they believe is most important, but they are under fairly substantial constraints from institutional funders.

What would additional funding allow FHI to do?

With smaller amounts of funding (around £60,000 or \$100,000), FHI might:

1. **Fund thesis prize competitions and other research field-encouraging activities.** In the past, FHI has funded a couple of small prizes (a few thousand pounds each) for people who write dissertation abstracts on topics of interest to FHI. This has been a cheap and effective way to build relationships with young and promising researchers, and it is hard to get this money from institutional funders. Other examples of potentially valuable but ‘hard to fund’ projects FHI has discussed or undertaken include a summer school for very talented young people, distributing important books and papers free of charge, and media projects such as developing an educational documentary about existential risk research.
2. **Fly in good and interested journalists to learn about and report on FHI’s work.** FHI did this with Ross Anderson in the past and got a lot of great attention (including in the *Atlantic*) for FHI’s work.
3. **Build up its reserves.** Recently reserves were depleted to fund two researchers while they were between grants from institutional funders – these gaps happen and need to be covered quite frequently. FHI would like to have another £50,000 or so in reserves for this purpose.

With larger amounts of funding (around £600,000 or \$1M), FHI might:

1. **Build up its “core operations” staff, hiring a project manager, an administrative assistant, and/or research assistant.** Sean feels that FHI is currently constrained in this area. He believes additional capacity would substantially increase FHI’s overall output, including the output of Nick Bostrom. For one example, as academic manager one of Sean’s main goals is to free up Bostrom from many of the management, funder meeting and other directorial obligations that would normally fall on him, allowing Nick to focus on top-level FHI strategy and research prioritization, as well as his own research. FHI currently has 2 academic/project managers, 50% of an administrator and 50% of a PA (for Nick Bostrom) to manage, promote, fundraise, organize events, and assist in research for 16 researchers, research associates and regular collaborators.
2. **Fund Nick Bostrom and FHI’s core processes.** While FHI has had good success in securing researchers’ salaries recently, the project funding both Nick Bostrom’s salary and the FHI’s core processes (visitor and research associate programme, conferences, workshops,

websites/outreach) ends in August 2015 and a replacement has yet to be secured. This is therefore a high priority. [Note that employer cost of a professorial salary is much higher than that of a standard researcher.]

3. **Free up researchers from more restrictive funding sources.** Some strong FHI researchers could spend a significantly larger portion of their time on the projects that they believe are most important if they were funded from an unrestricted source.
4. **Extend the contracts of some researchers.** Some strong FHI researchers will be running out of institutional funding soon, and additional funding would allow FHI to keep them on.
5. **Hire researchers in new areas.** FHI would like to hire people with strengths in synthetic biology, generalists in the history of science and technology, and people with expertise in science and technology studies.

With much larger amounts of funding (around £3M or \$5M), FHI might:

1. Establish a joint research programme on existential risk with CSER (along the lines of a recently rejected joint grant proposal) and/or the Future of Life Institute.
2. Do extremely in-depth evaluations of specific risks, such as synthetic biology, nuclear war, and surveillance.
3. Bring in senior visiting researchers.
4. Fund additional workshops, events, and conferences.

Are there specific people FHI would have liked to hire, but couldn't hire because of lack of funds?

Yes. A year and a half ago ago, FHI advertised a position and had four highly qualified applicants that they would have liked to fund, but could only fund one of them. One of these applicants was hired a year later, and another was hired by another valuable project a couple of months later. Additional funding would have allowed FHI to get an additional few years of valuable research done.

Recently, FHI had two very good candidates for project management staff and was only able to hire one of them. An additional project manager would have been very valuable.

Sean believes that talent constraints would become a more significant factor if FHI had a large amounts of funding, such as £2M, at short notice. However, he believes this could easily be overcome by a more systematic search for talent, something that would be prioritized if a large increase in funding was on the horizon.

Monitoring and evaluation

What information does FHI currently share about its work?

All papers published are freely available on the FHI website. FHI keeps detailed records of its engagement with the media (which includes hundreds of press appearances).

FHI produces reports on its progress for its institutional funders, but these funders have previously requested that FHI not keep these reports online. They are available upon request for individuals.

FHI has substantial engagement with policymakers on core FHI issues, but much of this cannot be shared with the public due to the nature of the arrangement.

CSER questions

Current operations

What is CSER doing right now?

CSER's top priority is to acquire enough funding to commence research programmes. A previous effort to fund a large proposal from the European Research Council was nearly approved, but ultimately wasn't. Follow-on applications are in development.

CSER's second priority is to expand its research networks, meeting and finding support with other research groups and governmental representatives, and adding influential people in existential risk-relevant fields to its advisory boards. (It is important that CSER be a "powerful" organization academically and politically, able to translate existential risk mitigation research into real-world policy and impact, hence the high priority of this networking).

While CSER's research programmes haven't commenced, CSER's founders and advisors are already having a substantial impact by consulting policymakers, holding workshops and public lectures on different aspects of existential risk, and raising awareness of risks within relevant fields.

Who is actively working on CSER right now?

The co-founders are Huw Price, Jaan Tallinn, and Martin Rees. Huw Price is taking the lead, but all three have other full-time jobs and other demanding commitments.

Sean is currently transitioning from working 30% to 80% for CSER (20% at FHI). Once his FHI projects have been fully transferred, he may hire project assistance at CSER if funding can be secured, as there are more opportunities at present than can be taken advantage of. An administrator at CRAASH (the Centre for Research in the Arts, Social Sciences, and Humanities) is providing a small portion of time well.

Room for more funding

How much funding does CSER have right now?

CSER received £60,000 in initial seed funding from Jaan Tallinn, of which some has been spent. It is covering a portion of Sean's time, some funding for PIs to travel on CSER business, a small amount of administrative support at Cambridge, website support, and organizing workshops. Martin Rees is organizing two high-level workshops with senior scientists and policymakers to discuss existential risks from emerging technologies. This funding has been supplemented by an additional £24,000, which allows Sean's involvement to be increased, and for him to move to Cambridge to work on the project 4 days/week. In addition, we have recently secured a commitment of £50,000/year, expected to continue for three years – we aim to raise £20K/year in order to use these funds for a postdoc.

What pending grant applications does CSER have?

CSER has a £1.5M grant application in preparation to a major foundation, to be submitted in late May. Knowledge of outcome expected in November 2014. Several other academic and philanthropic options are being explored.

Questions about hiring additional research fellows

What would CSER do with additional funding?

With a smaller amount of funding (£60,000-£100,000 or \$100,000-\$160,000):

1. CSER would use £20K/year to bolster an additional £50k/year commitment in order to fund a postdoctoral researcher. This researcher would work within or in association with a larger existential risk project, funded from academic sources or a foundation.
2. CSER would use an additional £25-£35K to hire an assistant project manager for a year to work with Sean and Huw. Many opportunities are becoming available that time constraints don't allow us to take advantage of – additional project assistance would have an outsized impact.
3. Additional funds of £10-20K/year would allow more workshops, meetings, events, and travel by founders and staff.
4. All of these funds (plus additional smaller-scale funding) doubles up as funding that can be committed towards the cost of a larger foundation-funded project to increase its chances of success. For example, CSER has recently received an expression of interest from a foundation that typically funds 50% of a project – if seed funding combined with philanthropic donations can be used to meet the other 50% of the cost of the first year of the project, this may be the most direct route to the full launch of the centre.

With a larger amount of funding (£1.5-2M), CSER would hire a full-time director for CSER, 4 post-docs, and 50% of an administrator for 3 years. Finding the right person may be a significant hurdle for hiring a full-time director; however, Huw Price is likely to be able to take over as Director for a three year period for stage 1, giving enough time to get the Centre's work off to a good start, guide it to phase 2 (see below) and find a longer-term director. Currently, CSER is exploring academic/foundation grant options for this level of funding, but welcomes philanthropic interest. If funding for this "pilot phase" of CSER is achieved, there are indications that smaller grants and philanthropic donations should be available to complement the initial programme and develop this into a bigger and more lengthy project.

With a much larger amount of funding (around £5M or \$8M), CSER would do the work outlined in a large grant they recently applied for. This would involve hiring a full-time director, 8-9 postdocs, a project manager, and an administrator. Finding the right person may be a significant hurdle for the full-time director, and a moderate hurdle for the last 4-5 post-docs.

Sean feels that getting CSER to something close to the middle level, with a full-time coordinator and at least a couple of post-docs, is important for getting the project off the ground. Getting a single post-doc to do work for a few years would be less exciting as a test of the project or an attempt to get the project running at full steam.

Is getting CSER going important for ensuring that all the cool people who want to be associated with the project continue to want to stay associated with it?

Yes, but the people involved know that it will take some time to get everything set up. Some real output, even if just in the form of workshops, conferences, and lectures should help. Getting enough funding to hire a full-time director and a few post-docs would probably be sufficient.

Appendix: questions I sent to Sean prior to our meeting

FHI

Room for more funding

1. Pure finance issues
 - a. What is FHI's annual budget?
 - b. In the last year, how much did FHI get from non-institutional funding sources?
 - c. How much unrestricted funding does FHI have?
 - d. How much does FHI have in reserves?
 - e. In the last year, how much unrestricted funding did FHI get?
 - f. What pending grant applications does FHI have?
 - g. Is FHI actively seeking funds from non-institutional donors? How much would FHI like to have?
2. What would additional funding allow FHI to do?
 - a. What would we do with different amounts of money?
 - i. \$100,000
 - ii. \$1M
 - iii. \$5M
3. Questions about hiring additional research fellows
 - a. How much does it one research fellow-year cost?
 - b. What is typical output for a research fellow-year?
4. Funding constraints vs. talent constraints
 - a. Are there specific people FHI would have liked to hire, but couldn't hire because of lack of funds?
 - i. I think Owen Cotton-Barratt might be one
 - ii. I think I was one a year ago

Monitoring and evaluation

1. What does FHI produce showing its output?
2. What is it going to do in the future?

CSER questions

Current operations

1. What is CSER doing right now? Is it still basically just looking for funding in order to get going?

2. Apart from you, who is putting effort into CSER right now?

Room for more funding

1. How much funding does CSER have right now?
2. What pending grant applications does CSER have?
3. Questions about hiring additional research fellows
 - a. How much does it one research fellow-year cost?
 - b. What is typical output for a research fellow-year?
4. Funding constraints vs. talent constraints
 - a. Are there specific people CSER would have liked to hire, but couldn't hire because of lack of funds?
5. Is CSER actively seeking funds from non-institutional donors? How much would CSER like to have?
 - a. What would CSER do with the amount it would like to have?
 - b. What would we do with different amounts of money?
 - i. \$100,000
 - ii. \$1M
 - iii. \$5M

Other

1. Is getting CSER going important for ensuring that all the cool people who want to be associated with the project continue to want to stay associated with it?