Future of Humanity Institute

Bimonthly Progress Report

Issue 2 July 2006

Message from Nick

Our new whiteboards are being put to heavy use for thinking sessions in which new research ideas are being worked out. This Report outlines some of the research we are conducting – often in collaboration with researchers both inside and outside Oxford.

As the Report also mentions, we have been involved at the national and European level in giving advice to policymakers on issues relating to technology convergence (nano-bio-info-cogno) and human enhancement in sport. And we have been working to promote public awareness and discussion through a rather intense media coverage as well as through public lectures and by working with the Wellcome Trust on an exhibition on cognitive enhancers.

Dr Nick Bostrom, Director, FHI

Forthcoming Events

Wednesday 26 July

Nick Bostrom to deliver a keynote address at the Wellcome Trust's *Wellbeing in the 21st Century* series.

Saturday 5 August

Nick is presenting at the ENHANCE Workshop on Enhancement in Beijing , a satellite conference to the 8th World Congress of the International Association of Bioethicists.

Research is now progressing on several fronts in the FHI. Below are just a few samples of the ideas we are currently pursuing. Some of these will be described in more detail in future Reports.

On the ethics front, one topic we're working on is decision making under fundamental ethical uncertainty. The basic problem here is easy to state. There are various ethical theories that tell us how we ought to act. But what should we do if we are uncertain about which ethical theory is correct? In other words, how should a morally responsible agent act when uncertain not just about specific empirical matters but also about fundamental issues in ethics? Since there is no consensus about fundamental ethics, it seems important to address the problem of how to act in a responsible manner under such conditions. Nick is working on this together with Toby Ord. Nick and Toby's previous research collaboration, on status quo bias, resulted in a paper that is forthcoming in *Ethics*, and which is already being used as required course reading at several universities.

The prospect of human enhancement raises many well-known ethical and social issues. It also raises daunting practical and scientific challenges. How can one find feasible modifications of human capacities that will (a) work, and (b) not cause unacceptable medical side effects? The human organism is a marvel of complexity, of which medical science still has only a very limited understanding. Nick is working together with Dr. Anders Sandberg on developing a heuristic, based on evolutionary considerations, for guiding medical research in this area. The aim of the heuristic is to help identify promising enhancement opportunities and to indicate what side effects are likely to result from different kinds of interventions. Nick and Anders are also working on a state-of-the-art survey of current cognitive enhancements.

Ideas about "human nature" have played an important role in some arguments about the ethical acceptability of human enhancement. The concept of human nature is important but also problematic. Dr. Rebecca Roache is investigating how different views of human enhancement rest on assumptions about human nature and to what extent these assumptions hold up under critical scrutiny. She gives a preliminary synopsis of her work on page 5 in this Update.

Work is ongoing on an edited volume on global catastrophic risk, forthcoming with Oxford University Press. As part of this project, Nick is working on revising and updating his earlier writings on "existential risks". Existential risks are threats that endanger the survival of the human species or its future potential. In the last update, we reported that Nick had published a paper in *Nature* together with MIT physicist Max Tegmark, in which they derived an upper bound on probability of a particular subset of existential risks. Nick is now working on clarifying the conceptual framework for analyzing existential risks. The OUP volume is edited by Nick and Dr. Milan Cirkovic, an astrophysicist working from Belgrade in Serbia.

Contents

News

New Publications

Since our last bimonthly update, *Analysis and Metaphysics* has published Nick Bostrom's article, 'Desire, Time, and Ethical Weight'. He has written a chapter called 'Observation Selection Theory and Cosmological Fine-Tuning', which he was invited to contribute to *Universe or Multiverse?*, edited by Bernard Carr; and a chapter entitled 'Recent Developments in the Ethics, Science, and Politics of Life-Extension', which he was invited to contribute to *Death and Anti-Death, Volume 3: Fifty Years After Einstein, One Hundred Fifty Years After Kierkegaard*, edited by Charles Tandy.

Philosophy has published Rebecca Roache's article, 'A Defence of Quasi-Memory'. For more information about Rebecca's current research, see page 5 and for recent reprints and translations of Dr Bostrom's work please see below.

Reprints and Translations

April-July 2005

- Simulation Argument' reprinted in *Doing Philosophy: An Introduction through Thought Experiments*, 3rd edition by Theodor Shick and Lewis Vaughn;
- 'The Future of Human Evolution 'in *Futurology—Forecasts and Initiatives*, edited by P. Bala Bhaskaran;
- 'How long before Superintelligence?' (with a new postscript) and 'The Transhumanist FAQ, v. 2.1' in *Linguistic and Philosophical Investigations*;
- 'The Mysteries of Self-Locating Belief' in *Review of Contemporary Philosophy* (forthcoming in August);
- 'A History of Transhumanist Thought' and 'The Mysteries of Self-Locating Belief and Anthropic Reasoning' in *Analysis and Metaphysics* (forthcoming in October).
- 'The Singularity' has been translated into Romanian for *Net SF* as 'Bun venit in lumea schimbărilor exponenţiale',
- 'Letter from Utopia' for *Tendencias Cientificas* as 'Carta desde Utopía'.

Forthcoming 2006

- 'Are You Living in a Computer Simulation?' reprinted in *Linguistic and Philosophical Investigations*, 'Astronomical Waste'
- 'Human Genetic Enhancements: A Transhumanist Perspective' in *Review of Contemporary Philosophy*,
- 'Transhumanism: The World's Most Dangerous Idea?' and 'Observation Selection Effects, Measures, and Infinite Spacetimes' in *Analysis and Metaphysics*.

Lectures

"Human Enhancement and Sports Enhancement." Invited presentation for the *Science and Technology Select Committee, House of Commons* (UK parliament) (London, 21 June).

"Posthuman Dignity and the Rights of Artificial Minds" Invited closing keynote for the conference *Human Enhancement Technologies and Human Rights*, IEET and Stanford University Law School (San Francisco, 26-28 May).

"The Simulation Argument." Invited "annually hosted special lectures by speakers who have made distinguished contributions to the theory or applications of symbolic systems" at *Stanford University* (Stanford, 19 May).

"Existential Risks and Artificial Intelligence" Invited keynote at the *Singularity Summit* (Stanford, 13 May).

"Consequences of Cognitive Enhancement" *ENHANCE workshop* presentation (Oxford, 4 May)

Peter Taylor joins the FHI

Dr Peter Taylor

We are delighted to welcome a new Research Associate to the FHI. Dr Peter Taylor will be working with us part-time from October, and will be focusing on the area of catastrophic risk.

Peter has spent the last 25 years working in the Lloyd's insurance market, where he has managed IT and loss modelling departments and led and participated in many projects. He has been a director of insurance broking and underwriting companies and market organisations, and is currently the Project Director of the Lighthill Risk Network (www.lighthillrisknetwork.org), a non-profit organisation based at the Lighthill Institute of Mathematical Sciences, with a mission of bringing together the business and scientific communities for their mutual benefit. Peter has a long-standing interest in all aspects of risk, whether in insurance or in science generally, and has a particular background in the foundations of quantum theory for which he was awarded his D Phil at Oxford.

Nick's ideas were recently profiled in a regular *Guardian* column, 'The Ideas Interview', published on May 9, 2006, discussing his views on transhumanism. On genetic engineering, Nick said, " it will be increasingly possible to modify human capacities. The issue now is whether we should do it." And, if so, what are the ethical constraints?" You can read the full interview at http://www.guardian.co.uk/science/story/0,,1770695,00.html, or copied on pages 6 and 7.

Nick's Recent Media

Wellcome Trust Science Museum: background interview about cognitive enhancers.

Technocalypse (film documentary): interviewed about status quo bias, human rationality, and human enhancement.

The Sunday Times: interviewed about the impacts of growing up with digital technology on brain development and psychology.

Human Values in a Transhuman World: radio documentary about ethics, human enhancement, and new technologies.

Meme Therapy (blog): interviewed about transhumanism and related issues.

The Next Paradigm (TV documentary): interviewed about the singularity and the future of artificial intelligence.

TV documentary on transhumanism and related issues.
French Feature Film: interviewed about aging and life-extension.

Autopilots (TV documentary): interviewed about the future of robotics and artificial intelligence.

Bon Magazine: interviewed about memory enhancing and memory deleting drugs and their social and ethical ramifications.

Honours and Awards

Nick Bostrom has been named the Symbolic Systems Distinguished Speaker of 2006. Since 1991, the Symbolic Systems Program has annually hosted special lectures by speakers who have made distinguished contributions to the theory or applications of symbolic systems.

Previous Distinguished Speakers have included Daniel Kahneman, Michael Gazzaniga, Daniel Dennett, John Searle, and Steven Pinker.

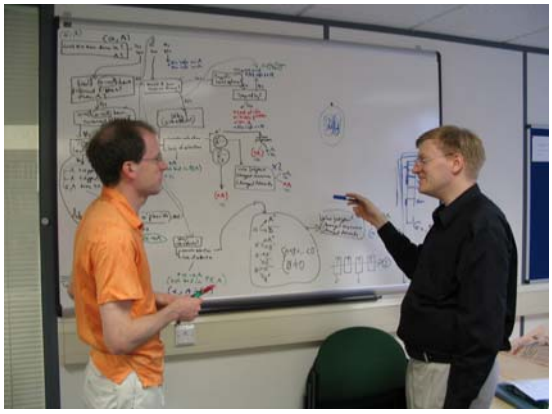
Nick Bostrom has also been named in *Marquis Who's Who in the World* (24th edition, 2007).

Policy

EU Building in Brussels

Nick Bostrom was recently called upon to give his views at the House of Commons Science and Technology Select Committee on human enhancement policy issues, focusing on sports enhancement. He has since been asked to become the official advisor to this committee.

Nick also travelled to Brussels for the European Parliament Scientific Technology Options Assessment, to discuss 'Converging Technologies in the 21st Century: Heaven, Hell or Down to Earth?'



Nick Bostrom and Anders Sandberg discuss a new idea

Launch

The Future of humanity Institute and its sister project the Program on the Ethics of the New Biosciences will host a workshop in October to initiate new collaborations and to celebrate their first few months working on the most important issues that we face.

Invited participants will work together brainstorming fruitful new areas of research.

FHI is recruiting for two postdoctoral positions. Please see page 7 for further details.

Please contact Miriam Wood (James Martin Projects Officer) in the first instance. Phone 01865 286279 or email miriam.wood@philosophy.ox.ac.uk

Visitors

FHI regularly invites visitors with relevant interests to collaborate in a variety of ways. Past visitors have presented at the James martin 21st Century School Advanced Research Seminar Series, or have given special public lectures, open to all to attend. Another important advantage of visiting scholars is collaboration and discussion with academics with specialist knowledge that can help forward the work of FHI. Robin Hanson, a professor of economics will arrive in July.

If you would like to visit FHI, please contact Miriam Wood on the contact details on p. 8, to discuss the possiblilites. FHI may be able to offer Bodleian reading rights, and office space, as well as the opportunity to collaborate on the FHI's areas of interest.

Recent and Forthcoming Visitors

Robin Hanson

Professor Robin Hanson

In July, we will welcome our first visiting scholar to the FHI. Professor Robin Hanson is Associate Professor of Economics at George Mason University. He is a pioneer of the concept of prediction markets and an expert on Bayesian epistemology. He has proved some interesting results showing, given a few seemingly weak assumptions, that it is impossible for rational Bayesian agents knowingly to disagree with one another, extending earlier work in this area by Robert Aumann. The topic of rational disagreement is important to the FHI for many reasons, but particularly in relation to the FHI's third major area of focus: the methodology of thinking about big picture questions for humanity (an area where disagreement is certainly rife!)

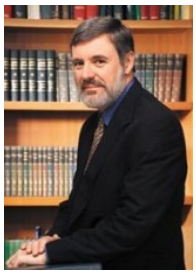
Professor Hanson has also made significant contributions to a wide range of other topics that are relevant to the FHI's work, including the future of economic growth, uploading, the nature of health care preferences and the origin of the placebo effect, the Fermi paradox, and the evolutionary dynamics of space colonizing replicators.

He has a PhD in economics, MA degrees in physics and in philosophy of science, and ten years of experience as an artificial intelligence researcher before switching field to economics. Combining a broad knowledge base with a highly analytic mind, Hanson is one of the most interesting thinkers in several of the areas that the FHI is concerned with, and we are looking forward to a period of fruitful collaboration and discussion. Find out more about his work at his website: http://hanson.gmu.edu/.

If you would be interested in meeting Professor Hanson whilst he is here, please consult the FHI website for information about when he is visiting and how to contact him.

Carl Cranor and Jonathan Wolff

Professor Carl Cranor, from the University of California Riverside, presented a paper earlier this month entitled 'Toward a Non-Consequentialist Approach to Acceptable Risks', as part of the series of 21st Century Advanced Research Seminars that the FHI runs in association with the Program on the Ethics of the New Biosciences.



Professor Carl Cranor

Professor Jonathan Wolff, from University College London, responded.

Jeff McMahan

Professor Jeff McMahan

Also presenting in our 21st Century Advanced Research Seminar series was Professor Jeff McMahan, from Rutgers University. His paper was entitled 'Killing Embryos for Stem Cell Research', and a webcast of his presentation is available for download from the FHI website.

Professor McMahan was visiting Oxford as the speaker for the 2006 Uehiro lectures, where he presented a series of three lectures on the topic 'Responsibility and Liability in War.'

For more information on Oxford's Uehiro Centre for Practical Ethics, see http://www.practicalethics.ox.ac.uk. This website offers downloadable abstracts for Professor McMahan's lectures, 'Unjust Warfare', 'Just Warfare', and 'Killing Civilians'.

The 21st Century Advanced Research Seminars will be continuing in Michaelmas Term. To be added to the mailing list to receive papers, please email miriam.wood@philosophy.ox.ac.uk

Enhancement, Politics, and Human Nature

A research update from Rebecca Roache

What is human nature, and what is the role of this concept in debates about enhancement? Fukuyama urges a cautious approach to enhancement technology because it could permanently change human nature, which he defines as 'the sum of the behaviour and characteristics that are typical of the human species, arising from genetic rather than environmental factors' (2002: 130). He thinks changing human nature would be bad because it 'shapes and constrains the possible kinds of political regimes, so [changing it] will have possibly malign consequences for liberal democracy and the nature of politics itself' (p. 7). Liberal democracy, he argues (1989, 1992, 2002), is the ideal form of government because its principles complement human nature; unlike, for example, communism, which demands that citizens care equally for all people, contrary to our desire to prioritise family and friends. Since liberal democracy is only ideal given current human nature, however, we can safeguard political stability only if we avoid making significant changes to our nature.

Fukuyama's argument is unlikely to convince advocates of enhancement. Their likely response to his claim that enhancement could undermine liberal democracy is, so what? If we cannot have both liberal democracy and enhancement, we should choose enhancement, since its benefits would outweigh the temporary inconvenience of having to develop a new political regime for our changed human nature. Focusing on the possibly disruptive social effects of enhancement, then, merely results in a stand-off between those who think enhancement is worth it and those who don't.

Let us instead consider the claim that liberal democracy is the ideal political regime because it best fits human nature. This claim is less plausible than it may initially seem, since the way we think about human nature may lead us to over-emphasise the fit between human nature and liberal democracy. This is because we define human nature in terms of dispositional properties. It would be less accurate, for example, to claim that the typical height of human beings falls within the range *n* to *n*+1 than to claim that our height typically falls within this range *given certain environmental conditions* (adequate nutrition, etc). In any given environment, we will see expressed in its inhabitants those features for which the environmental conditions are optimal, and less or no evidence of those features for which the environmental conditions are suboptimal. In the case of humans, environmental conditions should be taken to include social conditions. The expression of features like aggression, fear, and cooperativeness, for example, vary depending on the social situation in which people find themselves.

Since our political environment is a type of social environment, and since the environment allows the expression of only a subset of features of its inhabitants, the claim that liberal democracy is the ideal political regime given human nature becomes unconvincing, since it is based upon an incomplete understanding of human nature. History has allowed us to observe humans in a relatively small variety of political environments, and there may be features of human nature that have never been expressed because they have never found the right environment. Therefore, we should be sceptical of absolutist claims about what sort of political environment is ideal for humans.

Where does this leave the debate about enhancement? Well, Fukuyama is right that there may be some enhancements that we should avoid because they might have undesirable consequences. However, we have little reason to believe that changing human nature—if this is indeed where enhancement is likely to lead—is bad *per se*. It is possible that some changes may improve our ability to get along with each other, whatever the political environment. Deciding which enhancements will be beneficial in the long run is no easy feat, but the claim that the current state of human nature should guide such decisions is implausible.

References

Fukuyama, F. (1989) 'The End of History?' *The National Interest* 16: 3-16.

_____ (1992) *The End of History and the Last Man* (New York: Free Press).

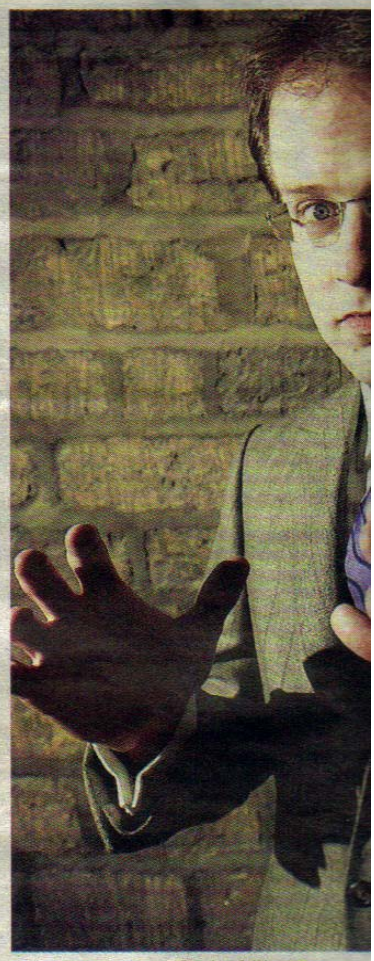
_____ (2002) *Our Posthuman Future* (New York: Farrar, Straus and Giroux).

Rebecca Roache is a Junior Research Fellow at the FHI. Please see p 8 for Rebecca's biography, including contact details.

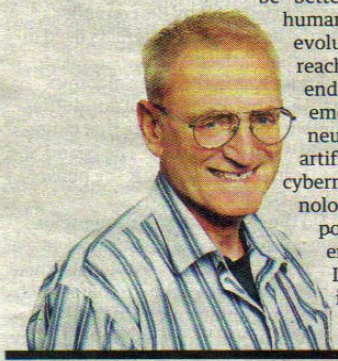
Ideas

The ideas interview
Nick Bostrom

John Sutherland meets a transhumanist who wrestles with the ethics of technologically enhanced human beings



The World Transhumanist Association was founded in 1998 by the philosophers Nick Bostrom and David Pearce. It describes itself as "an international nonprofit membership organisation which advocates the ethical use of technology to expand human capacities." Its proclaimed goal is that people should be "better than well", and that human development, in evolutionary terms, has not reached anything like an endpoint: all kinds of emerging technologies – neuropharmacology, artificial intelligence and cybernetics, and nanotechnologies – have the potential, it says, to enhance human abilities. In effect, it is interested in self-improvement and human perfectibility through the ethical application of science.



In a world suffused with gloom, is the WTA project not wildly utopian, I ask Dr Bostrom, who is the association's principal spokesperson and teaches at Oxford University.

"That might be true for some transhumanists," he replies. "I personally don't think of myself as either an optimist or a pessimist. I believe that if you look at the best-case scenarios, the upside is enormous. But there are clearly major risks that humanity will have to confront in this century. I can see a downside scenario as well, reaching down as far as the level of total human extinction. The possibilities range from the wonderful to the horrible. If I had to pull a number out of a hat, I'd say a 20% probability of extinction. Non-trivial."

How is transhumanism different from discredited notions of "creative evolution" – the idea that mankind, as a species, was evolving ever higher up the ladder, passing on its acquired traits to succeeding generations?

"Creative evolution, as propounded by Lamarck, was discredited by Darwin. Traits acquired during one's lifetime – muscles built up in the gym, for example – cannot be passed on to the next generation. Now with technology, as it happens, we might indeed be able to transfer some of our acquired traits on to our selected offspring by genetic engineering."

Transhumanism, as I understand it, is moving its focus on to ethics, regarding many of the technological enhancements as being in place. Is that the case?

"When I first got interested in this area a few years ago, the discussions would typically revolve around the question, 'Is this science fiction? Or are we dealing in realistic future possibilities?' Now the discussions tend to start from the position that, yes, it will be increasingly possible to modify human capacities. The issue now is whether we should do it. And, if so, what are the ethical constraints?"

When you say "modify human capacities", are you thinking of prenatal, postnatal, or midlife interventions? Prosthetic devices,for example?

"Prosthetic devices don't come into it except for people who happen to have some specific disability. For healthy adult people, the really big thing we can foresee are ways of intervening in the ageing process, either by slowing or reversing it."

How will technology achieve this?

"In the case of ageing, what you would need to do is either slow the rate at which this damage accumulates, or, even better, go in after the damage has accumulated and remove it. Stem cells, for instance, can be used to regrow cells that we have lost. And we might develop new enzymes which could break down those substances that the body, unaided, cannot deal with."

Transhumanist discourse often uses the term "post-human". What precisely is that?

"'Post-human' is a vague concept and people have used the term to mean entirely different things. It tends, in my opinion, to introduce more confusion than clarity. But one central meaning of the word would merely be an optimally enhanced human being."

Would this enhanced human being be what Nietzscheans call "the superman"?

"Nietzsche had a different view. He envisaged a moral and cultural transcendence: a very few people endowed with strong willpower and great refinement would throw off the shackles of traditional morality and

convention, and so rise above the rest of humanity. That's a very different mission from transhumanism where, ideally, everybody should have access to enhancement technologies."

Everyone their own superhuman?

"Well, it would be good if everyone had the option of, say, sharper memory and better health and longer life."

What are the ethical dilemmas that need to be solved?

"It's one thing if we are talking about adult, competent citizens deciding what to do with their own bodies. If, on the other hand, we are thinking of modifying children, or selecting embryos, then there is another set of ethical questions that arise. There is a further set of ethical questions relating to access. If some of the technologies, as they well might, turn out to be very expensive, then what mechanisms should be in place to ensure fairness?"

Surely the mechanisms are already in place? The rich will be able to afford them; the rest of us won't.

"One must ask, when these enhancement technologies are available and have been proven to work, whether they should be included in the package of treatments routinely offered to all by the NHS" ●

Bostrom . . . 'It will be increasingly possible to modify human capacities. The issue now is whether we should'

Illustration: Joe Magee

Nick Bostrom is the director of the Oxford Future of Humanity Institute. A list of his publications is available at www.nickbostrom.com

2 Postdoctoral Research Positions

FHI will shortly be advertising for two Postdoctoral Fellowship positions in Global Risk Modelling and theoretical risk. Further particulars will be available on the FHI website at http://www.fhi.ox.ac.uk/vacancies.html.

If you would like to receive the further particulars by email when they become available, please email miriam.wood@philosophy.ox.ac.uk.

Staff and Contact Details

Director
Dr Nick Bostrom

nick.bostrom@philosophy.ox.ac.uk

Nick Bostrom's research covers issues in the foundations of probability theory, global catastrophic risk, ethics of human enhancement, and consequences of potential future technologies such as artificial intelligence and nanotechnology, and related areas.
Bostrom has published more than 100 articles, including papers in journals such as Nature, Journal of Philosophy, Ethics, Bioethics, Mind, Journal of Medical Ethics, and Astrophysics & Space Science. He is the author of one monograph, Anthropic Bias (Routledge), and co-editor of two forthcoming volumes (OUP). His writings have been translated into more than 15 languages.

Bostrom has a background in physics and computational neuroscience as well as philosophy. Before moving to Oxford, he taught philosophy at Yale University. He is also a former British Academy Postdoctoral Fellow. He worked briefly as an expert consultant for the European Commission in Brussels and for the Central Intelligence Agency in Washington DC.

Bostrom is a frequently sought-after commentator in the media, having done nearly 200 interviews for television, radio, and print media.

Research Associate
Dr Peter Taylor

peter.taylor@faraday.com

Peter Taylor has spent the last 25 years working in the Lloyd's insurance market where he has managed IT and loss modelling departments and led and participated in many projects. He has been a director of insurance broking and underwriting companies and market organisations, and is currently the Project Director of the Lighthill Risk Network (www.lighthillrisknetwork.org), a non-profit organisation based at the Lighthill Institute of Mathematical Sciences with a mission of bringing together the business and scientific communities for their mutual benefit.

Peter intends to spend most of his time working for the Institute from October 2006.

Peter has a long-standing interest in all aspects of risk, whether in insurance or in science generally, and has a particular background in the foundations of quantum theory for which he was awarded his D Phil at Oxford.

His interests include chemistry, physical geography, mathematics, physics, climate change, literature, art, cricket, and philosophy.

Research Associate
Dr Anders Sandberg

anders.sandberg@philosophy.ox.ac.uk

Anders Sandberg is a Swedish neuroscientist, science debater, futurist, transhumanist, and author. He holds a Ph.D. in computational neuroscience from Stockholm University and has studied computer models of human memory at the Royal Institute of Technology, Stockholm, Sweden. He has also been scientific producer for the neuroscience exhibition "Se Hjärnan!" ("Behold the Brain!"), organized by Swedish Travelling Exhibitions, the Swedish Research Council and the Knowledge Foundation that is touring Sweden 2005-2006.

He is co-founder of and writer for the think tank Eudoxa. Between 1996 and 2000 he was Chairman of the Swedish Transhumanist Association.

His other interests include physics, astronomy, biomedicine, psychology, complexity theory, art, sciencefiction, roleplaying, computer graphics, artificial intelligence, cognitive science, information visualization, intelligence amplification technologies, and the philosophy and politics of human enhancement.

Anders is the Postdoctoral Research Assistant for the Oxford ENHANCE Project. the Enhance Project is hosted by the Uehiro Centre for Practical Ethics.

Junior Research Fellow
Dr Rebecca Roache

rebecca.roache@philosophy.ox.ac.uk

Rebecca Roache studied philosophy at the universities of Leeds and Cambridge, receiving a Ph.D. from the latter in 2002. She then spent three and a half years working in IT, and a short spell teaching philosophy at the University of London, before joining the Future of Humanity Institute in 2006. Her research interests centre around issues relating to personhood. In particular, she is interested in the extent to which our thought, self-conception, and patterns of self-concern are products of the sort of beings we are, biologically; and the effects that enhancing human capabilities might have on our beliefs about personhood.

Rebecca is a Research Associate at Balliol College.

We want to hear from you!

To contact the FHI, or to be added to our newsletter mailing list, please contact Miriam Wood.

The Future of Humanity Institute
Suite 7, Littlegate House
16-17 St Ebbe's Street
Oxford OX1 1PT

Telephone: +44 1865 286279
Email: miriam.wood@philosophy.ox.ac.uk