

FUTURE OF HUMANITY INSTITUTE REPORT

Wednesday 18 July, 2007

CONTENTS

- Section 1: Achievements Report**
- Section 2: Annexe 1 - Staff List**
- Section 3: Annexe 2 - Deliverables Report (November 2005 – July 2007)**
- Section 4: Annexe 3 - Workshops and Forums**
- Section 5: Annexe 4 - Budget**
- Section 6: Annexe 5 - Publication specimen: “The Reversal Test” (Bostrom & Ord)**
- Section 7: Annexe 6 - Publication specimen: “How Unlikely is a Doomsday Catastrophe?” (Tegmark & Bostrom)**
- Section 8: Annexe 7 - Publication specimen: “The Wisdom of Nature” (Bostrom & Sandberg)**
- Section 9: Annexe 8 - Publication specimen: “Dignity and Enhancement” (Bostrom)**

Achievements Report

Oxford Future of Humanity Institute

November 2005 – July 2007

FHI's mission is to pursue big picture questions for humanity. We study how anticipated technological developments may affect the human condition in fundamental ways, and how we can better understand, evaluate, and respond to radical change. We do this from a multidisciplinary perspective. The Institute currently runs four interrelated research programs:

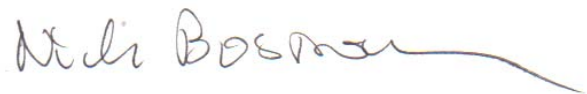
- human enhancement
- global catastrophic risks
- methodology and rationality
- impacts of future technologies

While the Institute was inaugurated on 29 November 2005, our first Research Fellow was not in post until 1 March 2006, and our other two James Martin Research Fellows were not appointed until 1 December 2006. We have thus been in operation at full capacity for less than eight months.

The following report (~2000 words) outlines our performance on the four criteria we have been asked to address: academic excellence, impact, additionality, and sustainability. We enclose the following annexes containing more detailed information and a few publication specimens:

1. Staff
2. Deliverables
3. Budget
4. Workshops and Forums
5. Publication specimen: "The Reversal Test" Bostrom & Ord
6. Publication specimen: "How Unlikely is a Doomsday Catastrophe?" Tegmark & Bostrom
7. Publication specimen: "The Wisdom of Nature" Bostrom & Sandberg
8. Publication specimen: "Dignity and Enhancement" Bostrom

We would like to express our thanks to Dr. James Martin and our other supporters, and to our colleagues in the James Martin 21st Century School, for providing us with the unique opportunity to address this range of pivotal issues within a single, mission-led Institute.

A handwritten signature in blue ink that reads "Nick Bostrom".

Dr. Nick Bostrom, Director

Oxford Future of Humanity Institute

Achievements Report

November 2005 – July 2007

1. Excellence

The FHI is uncompromisingly committed to academic excellence. The Institute has established itself as the top research centre of its kind in the world. Its academic excellence is revealed by the following indicators:

1.1 Publication record

Our research has been published in the world's top journals: *Nature*; *Ethics*; *Annals of the New York Academy of Sciences*; *Minds and Machines*; *Bioethics*; *Philosophical Quarterly*; *Journal of Medical Ethics*; *Journal of Law, Economics and Policy*; *Journal of Economic Methodology*; *Public Choice*; *Foreign Policy*; *Physica D*; *Physical Review E*; *Journal of Applied Philosophy*; *Philosophy*; *Analysis*; *Philosophy of Science*; and others. We have two books forthcoming with Oxford University Press.

The quantity as well as quality of research production is noteworthy:

- Academic journal papers: 40 (with a further 32 under review or in preparation)
- Articles and contributed book chapters: 43
- Books: 5
- Reprints and translations: 20

1.2 Prestigious lectures

FHI staff have delivered many prestigious lectures around the world. We have also had to turn down a large number of invitations because of the great demand. Three examples:

- FHI director is featured speaker at the *Second Annual Global Creative Leadership Summit* – “a unique platform for the best minds of our generation”, organized by LTB Foundation with support from the UN Fund for International Partnerships (UNFIP) (New York City, 2007)
- Delivering *The 11th Annual JUS Lecture* at the University of Toronto, 2007
- The *Symbolic Systems Distinguished Speaker of 2006*, Stanford University

Total number of lectures given: 95 (with a further 6 scheduled)

1.3 Prestigious invited contributions

We have been invited to contribute to numerous edited books. Three examples:

- Commissioned to write a contribution for the President’s Council of Bioethics in the United States, the world’s most influential public ethics body
- Invited to contribute *three* separate chapters to a series of books of philosophy-related work, published by Palgrave MacMillan, intended as “a showcase for original work from the best of the new generation of philosophers”
- Invited to contribute to *The New Stars of Science* (Vintage Books), a volume intended to include articles by “top young people in their respective fields ... i.e. who, among the young generation of scientists is most likely to turn out to be the next James Watson/Stephen Jay Gould/Martin Rees?”

1.4 Invitations to advise public bodies

FHI staff have been invited to give advice to a number of prestigious institutions (see 2.3 for details).

1.5 Translations and reprints

Some of our published works have been seen as sufficiently significant to warrant translation and reprinting in various collections.

- Writings by FHI researchers have been translated into 16 different languages
- There have been a total of 20 reprints

1.6 Academic ranking

While work in applied ethics constitutes only a fraction of the research conducted by the FHI, it is significant that The Philosophical Gourmet Report, the most important ranking of Graduate Programs in Philosophy in the English speaking world, rated Applied Ethics at Oxford University in the highest group, Group 1. We are also proud to be part of the Oxford Faculty of Philosophy, which is consistently rated among the top philosophy departments in the world (#2 in the latest ranking).

1.7 Academic influence

Work by FHI researchers has had impressive academic influence, as demonstrated by the fact that it has been very widely discussed and built-upon by other researchers around the world. Our work on existential risks, transhumanism, observation selection effects, and the simulation argument are among those topics to have made the greatest impact.

1.8 Miscellaneous

One staff member has been nominated for a Philip Leverhulme Prize (outcome pending).

2. Impact

FHI was created not only to produce world-class research, but also to have an impact outside academia. We take this mandate seriously. We have been highly successful at disseminating our findings to diverse constituencies, and at raising public awareness and

stimulating informed discussion in our areas of activity. There has been intense public interest in our work. Our advice has been sought by governments and influential organizations.

2.1 Media

FHI staff are frequently called upon by journalists to provide expert commentary. Our work has been featured widely on television, radio, and print media from around the world in outlets ranging from Chemistry World to GQ Magazine. Several television documentaries have been made focusing specifically on the work by done by our researchers.

FHI's work has been covered by BBC, CNN, Financial Times, The Independent, the Telegraph, The Guardian, ABC News, Times Higher Educational Supplement, Le Devoir, Toronto and Ottawa Suns, L'Hebdo, PC Plus magazine, Utildnings Radion, PBS, Boston Globe, San Diego Union-Tribune, The Sunday Herald, Discovery Channel, The Times, Nature, Italian National Television, New Scientist, The Observer, CBC, Science & Avenir, Volkskrant, Discover Magazine, New York Times, NBC, Forbes, and many others.

Total number of media appearances: 110

2.2 Forums

Another way in which we reach beyond the walls of academia is by creating forums where academics and other interested parties can interact, discuss, and develop collaborations. Examples include:

Overcoming Bias Blog, a highly successful new web forum, begun November 20, 2006, "for those serious about trying to overcome their own biases in beliefs and actions." This FHI blog was The Economist's 'BLOG OF THE WEEK' at Christmas 2006. The blog has attracted more than 241,000 unique visits. It currently attracts an average of 1,305 unique visitors per day (2,494 page views/day). It has published several hundred posts, and many thousand comments. www.overcomingbias.com

Lighthill Risk Network, created by FHI research associate Peter Taylor in conjunction with our program on global catastrophic risk. This is a newly established not-for-profit initiative with the aim of bringing world-wide scientific expertise in various aspects of risk to the financial services sector and (re)insurance industry. The Network provides business with a gateway to the latest knowledge and understanding of risk, while acting as a focal point for the research community to engage with industry. The expert panels will include Climate Change Implications, together with the Met Office (UK), Catastrophe Loss Modelling with the International Society of Catastrophe Managers (a US-based insurance organisation), Space Weather (partner: TBA), Quantitative Techniques in Insurance (partner: CASS Business School), and also Emerging Risks where the FHI will act as a partner.

2.3 Policy advice

FHI staff have been invited to give advice to the UK Parliament; the British Medical Association; the Royal Institution; the European Parliament; the European Commission; the UK's Parliamentary Office of Science and Technology; House of Commons Science and Technology Select Committee; UK's National Endowment for the Sciences, Technology, and Arts; the Academy of Medical Sciences; the organizers of the World Economic Forum in Davos; and others. We have also been commissioned to contribute a substantial piece of original research for the President's Council of Bioethics in the USA.

2.4 Web

Websites maintained by the FHI and its staff play a significant role in disseminating information to the public. We estimate that, in combination, these sites attract some 4,000 unique visitors per day.

2.5 Public lectures

FHI has delivered a total of 95 public lectures.

2.6 Events

The FHI has organized some half dozen academic conferences and workshops, and participated in the organization of several others. These have attracted world-leading scholars. One particularly noteworthy event was the Whole Brain Emulation Workshop, hosted by the FHI in Oxford in May 2007. This workshop assembled carefully selected experts from computational neuroscience, microscopy, and computer science, and will result in a technological roadmap for an important potential future technology. The roadmap is expected to become a defining document when it is released. Scientific research collaborations have already been established as a direct result of the workshop.

2.7 Teaching

The FHI has organized the James Martin Advanced Research Seminar (co-sponsored with the Program on Ethics and New Biosciences), which has run weekly since the launch of the Institute. The Seminar is attended by post-docs and Oxford graduate students from philosophy and other disciplines and has attracted speakers from around the world. FHI staff have also served as tutors for Oxford undergraduate students.

3. Additionality

The FHI was created from scratch and was not an extension of any pre-existing institute. The salary of the director, the three James Martin Research Fellows, and the administrative staff are all funded by the School. We have leveraged this by obtaining private philanthropic donations (details below), by recruiting five research associates to help with our research and other activities in an unsalaried capacity, and by forming research collaborations with academics in the James Martin 21st Century School and

elsewhere. None of the FHI's past or current activity could have taken place without the funding received from the School.

The FHI is, as far as we are aware, unique in the world. No other academic research institute is specifically devoted to the rigorous study of big-picture issues for the future of humanity. Characteristics of our approach include the emphasis on probabilistic methodology and biases, future technologies, global catastrophic risk, human enhancement, and on normative as well as positive questions.

The interdisciplinarity of our approach is reflected in the wide range of leading philosophical and scientific journals in which our research is published, and from the diverse academic backgrounds of our researchers.

Our researchers have made extensive use of the intellectual resources of the 21st Century School. We have established new lines of research following contact with other institutes, which includes attendance at research seminar series and discussions with researchers from the James Martin Institute for Science and Civilization, the Environmental Change Institute, the Programme on the Ethics of the New Biosciences, the Institute for Emergent Infections of Humans, the Institute for the Future of the Mind, and the Institute of Ageing. Our inter-institute connections have begun to bear fruit: five research papers have been or are being written by FHI researchers in collaboration with researchers from other institutes. Our researchers have a developmental role in four inter-institute research forums (including workshops, seminar series, and conferences) currently in preparation.

4. Sustainability

The FHI is a unique enterprise created to enable research on big picture questions for humanity's future. The research areas which we are pioneering are widely neglected, and traditional funding sources are lacking. We therefore face special difficulties in obtaining sustainability through research councils and similar funding agencies. Also, alone among the School's institutes, the FHI director's post is funded by the School as part of FHI's budget.

We have been encouraged to discover strong grassroots support, which has led to contributions from several private individuals (in addition to the original benefaction of Dr. James Martin) who strongly believe in our work. To date, the following donations have been received:

Philanthropist #1:	£101,714
Philanthropist #2:	\$25,000
Philanthropist #3:	\$11,000

We are preparing grant applications to the John Templeton Foundation for a Program on Wisdom in the 21st Century; to the Greek Ministry of National Education for the creation of a virtual centre called the Oxford Epictetus Center for the Promotion of Mental Health and the Study of Human Potential, which would include funding for some research within

the FHI; to the AHRC for research into normative judgement in the light of cognitive psychology and neuroscience; and to the Wellcome Trust for a program focusing on different concepts of risk. Other opportunities are also being actively explored.

All indicators unambiguously suggest that there is a vast demand for expertise and research into the topics we address. We would like to expand our research in several areas in order to build on our early successes. In particular, our current development priorities include the following:

- Develop a program on wisdom and rationality in relation to big picture questions for humanity
- Expand our program on global catastrophic risk, focusing in particular on (a) the very biggest risks, existential risks; (b) unduly neglected risks; (c) subtle risks; (d) risk assessment methodology
- Develop a program to analyze and evaluate global priorities in various areas and to identify strategic leverage points
- Expand our capacity to analyze radical future technologies such as nanotechnology and artificial intelligence

To realize these development goals, we will continue to leverage the support from the 21st Century School. We will continue to seek support from private philanthropists. We will also pursue grant funding from traditional funding agencies. It should be noted, however, that the latter approach – if carried too far – would incur the risk of distracting us from our original mission because of the necessity to tailor grant proposals to the allocation criteria used by conventional funding bodies.

In conclusion, in the brief time the FHI has been in operation – only eight months at full capacity – the Institute has achieved an astounding record of academic success, and it has achieved dissemination, public engagement, and policy impact out of all proportion to its size. The work is unique and would simply not be happening were it not for our place within the 21st Century School. We have leveraged the support from the School by obtaining external funding and by forming networks and collaborations. While we expect further successes in leveraging our core funding through financing by private philanthropists and by project funding from grants agencies, continued support within the 21st Century School is a necessity in order for the Institute to be able to continue to serve its role as an academic pioneer.

Annexe 1

Staff List

Research Staff	2
Dr. Nick Bostrom	2
Dr Rebecca Roache	2
Dr Rafaela Hillerbrand	3
Dr Nicholas Shackel	3
Dr Peter Taylor	4
Dr Anders Sandberg	5
Dr Guy Kahane	5
Professor Robin Hanson	5
Toby Ord	6
 Administrative staff	 6
Jo Armitage.....	6
Rachel Woodcock.....	6
 Fundraising Adviser	 6
Allison Taguchi	6

Research Staff



Director Dr. Nick Bostrom

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 130 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 16 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 more than 200 interviews for television, radio, and print media.



James Martin Research Fellow Dr Rebecca Roache

Rebecca Roache's research at the FHI centers around ethical issues surrounding human enhancement and new technology. Topics of particular interest include human nature and the relationship between humans and other species; the extent to which human values are products of the sort of beings we are, biologically, and the extent to which our values might change if we became different sorts of beings; rationality (and the lack of it) in popular thought about risk; and the role of intuition in philosophical reasoning. A philosopher by training, Rebecca's research also draws on material from social psychology, psychology, economics, politics, and biology.

In addition to her research, Rebecca enjoys participating in public consultations, educational initiatives, popular discussions, and media interviews relating to her work at

the FHI.

Rebecca 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 FHI in 2006.



James Martin Research Fellow in Global Risk Modelling
Dr Rafaela Hillerbrand

Rafaela Hillerbrand's research at FHI is about global catastrophic risk. Her research interests traverse epistemological problems related to the interpretations of probabilities, quantitative modelling, and foundational questions of statistical mechanics as well as ethical questions specific for decisions under risk or under uncertainty. The unifying question behind her research is the improvement of current risk assessments, with a particular focus on the unique problems of catastrophic risks.

Rafaela studied physics (with a minor in fluid mechanics) and philosophy (minor in political sciences) at the Universities of Erlangen-Nürnberg (Germany) and Liverpool. She received a Ph.D. in philosophy from the former in 2003 for a work on the ethics of technology. The work covered aspects of applied ethics just as well as genuine theoretical normative ethics, and was awarded the Lilli-Bechmann-Rahn-Preis in 2005.

Rafaela has done a Ph.D (2007) in Theoretical Physics at the University of Münster and the Observatoire de la Côte d'Azur in Nice (France). The thesis was on hydrodynamic turbulence.



James Martin Research Fellow in Theoretical Ethics
Dr Nicholas Shackel

Nicholas Shackel's research at FHI is focused on developing our rationality and methodology program. He has been doing research, including collaborations within and without the institute, on the ethics of expertise, epistemic ethics, public knowledge and complex systems, instrumental rationality and the relation of policy and science and on the scientific study of moral judgement. His prior philosophical research has been mainly on rationality. He has conducted research into the kinds of obligations there are to be rational in belief and in action, the relations

between practical and theoretical reason, paradoxes of rational decision, philosophy of probability, intentionality, and deontic logic. More recently, he has extended his research into the areas of neuroethics and neuroepistemology. His publications include papers in *Philosophy of Science*, *Mind*, *Erkenntnis* and the *British Journal for the Philosophy of Science*. Prior to joining FHI he was at the Oxford Centre for the Science of Mind, and before coming to Oxford he lectured in philosophy at the University of Aberdeen.



Research Associate
Dr Peter Taylor

Peter Taylor is a Research Associate concentrating on the area of risk with the FHI and James Martin School. Peter spent the 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 Deputy Chairman 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 is still an active consultant in the City of London, but spends as much time as he can working in the Institute. Peter has a long-standing interest in all aspects of risk, whether in insurance or in science generally, particularly the practical application of the theory of risk, and the analysis of emerging risks and has a particular background in the foundations of quantum theory for which he was awarded his D Phil at Oxford, and in July 2007 organised the Everett@50 Conference at the Philosophy Centre in Oxford. Peter makes regular conference speeches to the insurance industry on the subject of risk. His interests include chemistry, physical geography, mathematics, physics, climate change, literature, art, cricket, and philosophy.



Research Associate
Dr Anders Sandberg

Anders Sandberg is a Swedish computational neuroscientist and futurist. He holds a Ph.D. in computational neuroscience from Stockholm University. He is co-founder of and research director for the think tank Eudoxa. Between 1996 and 2000 he was Chairman of the Swedish Transhumanist Association.

Anders is the Postdoctoral Research Assistant for the Oxford part of the EU ENHANCE Project. Anders main research area is enhancement of human cognition and the philosophy and politics of human enhancement. In particular he links empirical sociological, economic and psychological research to issues in neuroethics. He is also interested in science popularization and the public engagement with emerging technologies.



Research Associate
Dr Guy Kahane

Dr Guy Kahane is the deputy director of the Uehiro Centre for Practical Ethics. He has BPhil and DPhil degrees in Philosophy from Oxford University, and has held postdoctoral positions at the Uehiro Centre for Practical Ethics and the Oxford Centre for the Science of Mind. He is currently heading a two-year project on neuroethics, funded by the John Fell OUP Research Fund. Starting September 2007, he will also be a Fulford Junior Research Fellow at Somerville College. Dr Kahane's research interests include neuroethics, practical ethics and value theory. He has written on pain, rationality, well-being and disability. Dr Kahane has also collaborated with Oxford neuroscientists in brain imaging studies of pain processing and moral judgement, and is a co-editor of the Blackwell volume 'Wittgenstein and His Interpreters'.



Research Associate
Professor Robin Hanson

Robin Hanson is an associate professor of economics at George Mason University, and a research associate at the Future of Humanity Institute of Oxford University. After receiving his Ph.D. in social science from the California Institute of Technology in 1997, Robin was a Robert Wood Johnson Foundation health policy scholar at the University of California at Berkeley. In 1984, Robin received a masters in physics and a masters in the philosophy of science from the University of Chicago, and afterward spent nine years researching artificial intelligence, Bayesian statistics, and

hypertext publishing at Lockheed, NASA, and independently. Robin has over 60 publications and since 1988 he has pioneered the new field of prediction markets. Robin also studies the social impact of future technologies.



**Research Associate
Toby Ord**

Toby Ord's research interests encompass both theoretical and practical ethics. He is currently focusing on a number of questions concerning the nature of consequentialism, its connection to practical decision making, and its relationship to other normative theories. In addition, he is looking at the moral status of the human embryo and at techniques to identify and overcome biases in ethical decision making.

Administrative staff



**James Martin Projects Co-ordinator
Jo Armitage**

Jo has worked for the Centre for Criminology, St Hilda's College and OUP during her time in Oxford. She has a BA from the University of Manchester, and a postgraduate qualification in personnel management from the Metropolitan University of Manchester.



**James Martin Projects Officer
Rachel Woodcock**

Rachel joined the team in February 2007. Prior to that, she worked as a Course Administrator in the Oxford Learning Institute. She has also worked at Oriel College and at OUP.

Fundraising Adviser



Allison Taguchi

Allison Taguchi serves as Funding Advisor to FHI. Allison has over 12 years of fund development experience at research institutes, universities, government agencies and nonprofit organizations. Some of the businesses she worked with in the past include: Rushford Nanotech Laboratory, Department of Defense, Oakland Military Institute, and University of Hawaii Biotech Research Center.

Annexe 2

Deliverables Report (November 2005 to July 2007)

Publications.....	2
Invited Lectures; Keynote Speeches; Conference Presentations.....	15
Workshops and Forums	23
Public Outreach	26
Media	27
Government, policy and advice	34
Teaching and seminars	36
Collaboration	37
Honours and Awards	40
Funding Procurement.....	42

Publications

Nick Bostrom

Books

Bostrom, N., (2007) *How Can Human Nature be Ethically Improved?* ed. with Julian Savulescu. (Oxford University Press: Oxford), forthcoming

Bostrom, N., (2007) *Global Catastrophic Risks*, ed. with Milan Cirkovic. (Oxford University Press: Oxford), forthcoming

Journal articles

Bostrom, N., Hillerbrand, R., and Meyer, A., (2007) "How reliable are forecasts on the technological progress", in preparation

Bostrom, N., (2007) "Sleeping Beauty and Self-Location: A Hybrid Model", *Synthese*, Volume 157, No. 1, forthcoming

Bostrom, N., and Sandberg, A., (2007) "Cognitive Enhancement: Methods, Ethics, Regulatory Challenges." *Science and Engineering Ethics*, forthcoming

Bostrom, N., (2007) "What is a Singleton", *Linguistic and Philosophical Investigations*, forthcoming

Bostrom, N., (2006) "Do We Live in a Computer Simulation" *New Scientist*, Volume 192, No. 2579, 19 November, pp 38-39

Bostrom, N., and Sandberg, A., (2006) "Converging Cognitive Enhancements" *Annals of the New York Academy of Sciences*, Volume 1093, pp 201-207

Bostrom, N., (2006) "Quantity of Experience: Brain-Duplication and Degrees of Consciousness", *Minds and Machines*, Volume 16, No. 2, pp 185-200

Bostrom, N., and Ord, T., (2006) "The Reversal Test: Eliminating Status Quo Bias in Bioethics", *Ethics*, Volume 116, No.4, pp 656-680

Bostrom, N., (2006) "Desire, Time, and Ethical Weight", *Analysis and Metaphysics*, Volume 4, No. 2, pp 59-83

Bostrom, N., (2006) "Ethical Principles in the Creation of Artificial Minds", *Analysis and Metaphysics*, October 2006

Bostrom, N., (2006) "Ethical Issues in Advanced Artificial Intelligence", *Review of Contemporary Philosophy*, August 2006

Bostrom, N., and Tegmark, M., (2005) "How Unlikely is a Doomsday Catastrophe?" *Nature*, Volume 438, No. 7069, p. 754 + supplementary materials

- Bostrom, N., (2005) “The Simulation Argument: Reply to Brian Weatherson,” *Philosophical Quarterly*, Volume 55, No. 218, pp 90-97
- Bostrom, N., (2005) “A History of Transhumanist Thought,” *Journal of Evolution and Technology*, Volume 14
- Bostrom, N., (2005) “Recent Arguments about Life-Extension,” *Aging Horizons*, No. 3, Autumn/Winter issue, pp 28-34
- Bostrom, N., (2005) “In Defence of Posthuman Dignity,” *Bioethics*, Volume 19, No. 3, pp 202-214
- Bostrom, N., (2005) “Understanding Quine’s Thesis of Indeterminacy,” *Linguistic and Philosophical Investigations*, Volume 4, No. 1, pp 60-96
- Bostrom, N., (2005) “Transhumanist Values,” *Review of Contemporary Philosophy*, Volume 4, No. 1-2, pp 87-101
- Bostrom, N., (2005) “The Fable of the Dragon-Tyrant,” *Journal of Medical Ethics*, Volume 31, No. 5, pp 273-277
- Bostrom, N., (2005) “Re: The Benevolent Dragon” *Journal of Medical Ethics*, 24 June 2005, 31/5/273#332, e-letter section
- Bostrom, N., (2005) “Scientist find death can damage your health” *Journal of Medical Ethics*, 31/5/273#308, e-letter section

Contributed book chapters, conference proceedings, and articles

- Bostrom, N., (2007) “Technological Revolutions: Ethics and Policy in the Dark” with a new Foreword, in J. Savulescu (ed) *Ethics of East and West: How they Contribute to the Quest for Wisdom*, (Oxford: Oxford Uehiro Center for Practical Ethics)
- Bostrom, N., and Roache, R., (2007) Human Enhancement.” Invited anthology chapter in J. Ryberg (ed) *New Waves in Applied Ethics*, (Palgrave Macmillan), forthcoming
- Bostrom, N., (2007) “Ethical and Political Challenges to the Prospect of Life Extension.” Invited article for *World Demographics Association Proceedings 2006*
- Bostrom, N., (2007) “Enhancement and Dignity.” Invited chapter for forthcoming book on human dignity by *The President’s Council of Bioethics*
- Bostrom, N., and Sandberg, A., (2007) “Brain-boosters” in P. Healey (ed) *Tomorrow’s People: The Challenges of Technologies for Life-Extension and Enhancement*, forthcoming
- Bostrom, N., (2007) “Technological Revolutions and the Problem of Prediction” in P. Lin, J. Moor and J. Weckert (eds) *Nanoethics*, (Wiley), forthcoming
- Bostrom, N., (2007) “Why I Want to be a Posthuman When I Grow Up.” in B. Gordijn and R. Chadwick (eds) *Medical Enhancement and Posthumanity*, (Springer)

Bostrom, N., (2007) “Technological Revolutions: Ethics and Policy in the Dark” in Nigel M. de S. Cameron and M. Ellen Mitchell (eds) *Nanotechnology and Society*, (John Wiley), forthcoming

Bostrom, N., (2007) “The Ethics of Artificial Intelligence.”, in W. Ramsey and K. Frankish (eds) *The Cambridge Handbook of Artificial Intelligence*, (Cambridge University Press), forthcoming

Bostrom, N., (2007) “The Future of Humanity.” Invited article in Jan-Kyrre Berg Olsen, Stig Andur Pedersen, and Vincent F. Hendricks (eds) *Companion to Philosophy of Technology*, (Blackwell), forthcoming

Bostrom, N., (2006) “Nanoethics and Technological Revolutions: A Précis.” *Nanotechnology Perceptions: A Review of Ultraprecision Engineering and Nanotechnology*, Volume 2 (1b), May Issue

Bostrom, N., (2006) “Observation Selection Theory and Cosmological Fine-tuning” Invited chapter in B. Carr (ed) *Universe or Multiverse?*, (Cambridge University Press: Cambridge)

Bostrom, N., (2006) “The Singularity.” Invited chapter in P. Miller and J. Wilsdon (eds) *Better Humans? The ethics and politics of human enhancement*, (DEMOS, January 26 2006)

Bostrom, N., (2006) “Growing Up: Human Nature and Enhancement Technologies” Invited chapter in E. Mitchell (ed) *Tomorrow’s People: The Challenge to Human Nature*

Bostrom, N., (2006) “Dinosaurs, Dodos, Humans?” Invited article for *Global Agenda*, the annual publication of the World Economic Forum, January 2006, pp 230-231

Bostrom, N., (2006) “Recent Developments in the Ethics, Science, and Politics of Life-Extension” Invited chapter in C. Tandy (ed) *Death And Anti-Death, Volume 3: Fifty Years After Einstein, One Hundred Fifty Years After Kierkegaard*, (Ria University Press)

Bostrom, N., (2005) “A Short History of Transhumanist Thought” contributed chapter to *The Prospect of Immortality*, by R. Ettinger, with Comments by Others, in C. Tandy (ed), in the Cultural Classics Series (Ria University Press, Palo Alto, California, 2005)

Bostrom, N., (2005) “The Future of Humankind: Heaven, Hell, with Stops Along the Way.” Review of “Radical Evolution: The Promise and Peril of Enhancing Our Minds, Our Bodies – and What it Means to be Human” by Joel Garreau. *Scientific American*, July, pp 86-87

Bostrom, N., (2005) “A Proactive Response to the Tsunami Disaster.” *BetterHumans*, 19 January

Bostrom, N., (2005) “Why Make a Matrix? And Why You Might Be In One.” contributed chapter in W. Irwin (ed) *More Matrix and Philosophy*, (New York: Open Court)

Reprints

- Bostrom, N., (2007) “Are You Living in a Computer Simulation?”, Reprinted in *Linguistic and Philosophical Investigations*, forthcoming, March 2007
- Bostrom, N., (2007) “Astronomical Waste”, Reprinted in *Review of Contemporary Philosophy*, forthcoming, August 2007
- Bostrom, N., (2007) “Human Genetic Enhancements: A Transhumanist Perspective”, Reprinted in *Review of Contemporary Philosophy*, forthcoming, August 2007
- Bostrom, N., (2007) “Transhumanism: The World’s Most Dangerous Idea?” Reprinted in *Analysis and Metaphysics*, forthcoming, October 2007 (this is an expanded version of earlier note in *Foreign Policy*)
- Bostrom, N., (2007) “Observation Selection Effects, Measures, and Infinite Spacetimes”, Reprinted in *Analysis and Metaphysics*, forthcoming, October 2007
- Bostrom, N., (2006) “Bun venit în lumea schimbărilor exponențiale” (translation into Romanian of “The Singularity”), *Net SF*, 30 March
- Bostrom, N., (2006) “The Simulation Argument”, Reprinted in *Doing Philosophy: An Introduction through Thought Experiments*, 3rd edition by Theodor Shick and Lewis Vaughn
- Bostrom, N., (2006) “The Future of Human Evolution.” Reprinted in *Futurology-Forecasts and Initiatives*, ed. P. Bala Bhaskaran (ICFAI University Press, Hyderabad, 2006)
- Bostrom, N., (2006) “How long before Superintelligence?” Reprinted in *Linguistic and Philosophical Investigations*, Volume 5, No. 1, pp 11-30, with a new postscript
- Bostrom, N., (2006) “Carta desde Utopía” (translation into Spanish of Letter from Utopia), *Tendencias Cientificas*, 28 January 2006
- Bostrom, N., (2006) “The Transhumanist FAQ, v. 2.1” Reprinted in *Linguistic and Philosophical Investigations*, Volume 5, No. 2, April 2006
- Bostrom, N., (2006) “The Mysteries of Self-Locating Belief”, Reprinted in *Review of Contemporary Philosophy*, August 2006
- Bostrom, N., (2006) “A History of Transhumanist Thought”, Reprinted in *Analysis and Metaphysics*, forthcoming, October 2006
- Bostrom, N., (2006) “The Mysteries of Self-Locating Belief and Anthropic Reasoning”, Reprinted in *Analysis and Metaphysics*, October 2006
- Bostrom, N., (2005) In Defence of Posthuman Dignity” *Linguistic and Philosophical Investigations*, Volume 4, No. 2. (Reprinted from *Bioethics*)
- Bostrom, N., (2005) “The Fable of the Dragon-Tyrant” *Linguistic and Philosophical Investigations*, Volume 4, No. 2. (Reprinted from *Journal of Medical Ethics*)

Robin Hanson

Journal articles

Hanson, R., (2007) “Insider Trading and Prediction Markets”, *Journal of Law, Economics, and Policy*, forthcoming

Hanson, R., and Cowen, T., (2007) “Are Disagreements Honest?”, *Journal of Economic Methodology*, forthcoming

Hanson, R., (2007) “Logarithmic Market Scoring Rules for Modular Combinatorial Information Aggregation”, *Journal of Prediction Markets* 1(1) pp 3-15

Hanson, R., (2006) “Designing Real Terrorism Futures”, *Public Choice* 128 (1-2) pp 257-274, July 2006. Also to appear in C. Rowley (ed) *The Political Economy of Terrorism* (2007), and W. Hancock (ed) *Business Continuity and Homeland Security* (2007) (Edward Elgar)

Book Chapters

Hanson, R., (2007) “Catastrophe, Social Collapse, and Human Extinction”, in N. Bostrom and M. Cirkovic (eds) *Global Catastrophic Risks*, (Oxford: Oxford University Press) forthcoming

Hanson, R., (2007) “Enhancing Our Truth Orientation”, in J. Savulescu (ed) *How Can Human Nature be Ethically Improved?*, (Oxford: Oxford University Press) forthcoming

Other Publications

Hanson, R., “Birds of a Feather; Letter On Why Hawks Win”, *Foreign Policy*, pp 10-12, March/April 2007

Rafaela Hillerbrand

Journal articles

Hillerbrand, R., (2007) Book review of Brian Leiter (ed), *The Future for Philosophy*, to appear in: *International Studies in the Philosophy of Science*

Hillerbrand, R., Bec, J. and Cencini, M., (2007) “Large Stokes number particles in incompressible flows,” *Physica D*, Volume 226, No. 11

Hillerbrand, R., Bec, J. and Cencini, (2007) “Inertial particles suspended in suspended in turbulent flows,” *Physical Review E*, Volume 75

Journal articles in preparation/submitted

Hillerbrand, R., Bec, J., Cencini, M., and Turitsyn, K., (2007) “Heavy particles in stochastic flows” submitted to *Physica D*

Hillerbrand, R., and Shackel, N., (2007) “Epistemic Ethics and Complex Systems”, in preparation

Hillerbrand, R., (2007) “Empirical arguments against current cost-benefit analysis of the aftermaths of a manmade greenhouse effect. How the latest IPCC report undermines economic and philosophical assessments”, in preparation

Hillerbrand, R., (2007) “The moral threat of a manmade climate change” to be submitted to *Physica D*

Meyer, A., Hillerbrand, R., and Bostrom, N., (2007) “Predicting technological progress. The predictions of the 1960s revisited”, in preparation

Hillerbrand, R., and Sandberg, A., (2007) “Quantum field theoretic risks”, in preparation

Hillerbrand, R., (2007) “The limits of the central limit theorem,” in preparation

Books

Hillerbrand, R., and Karlsson, R., (2007) *Environmental Justice and Global Citizenship*, EBook, forthcoming

Hillerbrand, R., (2007) *Distribution of massless and massive particles in turbulent flows. Differences and commons between Lagrangian tracers and inertial particles* (Mensch und Buch), forthcoming

Contributed book chapters, conference proceedings, and articles

Hillerbrand, R., (2007) “On the rationality and stability of a Minimal Consensus” in J. Kühnelt (ed), *Political legitimization without morality*, (Springer: Berlin, Heidelberg, New York), in press

Hillerbrand, R., (2007) “Dianoetic Virtues in Addressing a Morally Correct Treatment of GM” contributed paper at *Environmental Justice and Global Citizenship*, Mansfield College, Oxford

Hillerbrand, R., (2006) “Uncertainty as a challenge for ethics” contributed paper in Gasser, Georg/Kanzian, Christian/Runggaldier, Edmund: *Kulturen: Streit-Analyse-Dialog*

Guy Kahane

Journal articles under review

Savulescu J. and Kahane, G., (2007) “Procreative Beneficence and Disability: Is There a Moral Obligation to Create Children with the Best Chance of the Best Life?” (revised and resubmitted to *Ethics*)

Kahane, G., (2007) “Pain, Dislike, and Experience” (under review, *Utilitas*)

Kahane, G., (2007) “Non-Identity, Self-Defeat, and Attitudes to Future People” (under review, *Philosophical Studies*)

Kahane, G. and Shackel, N., (2007) “Utilitarian Bias or Defective Moral Dilemmas?” (communications arising, under review, *Nature*)

Books

Kahane, G., Kanterian, E., Kuusela, O., (2007) *Wittgenstein and His Interpreters*, (Oxford: Blackwell Publishers)

Contributed book chapters, conference proceedings, and articles

Kahane, G. Shackel, N. and Farias, M. (forthcoming) “Conceptual Problems in the Scientific Study of the Influence of Religious Belief on Pain” in C. Jäger ed., *Brain—Religion—Experience: Multidiscipline Encounters*, (New York: Springer)

Kahane, G., Kanterian, E., Kuusela, O. (2007) ”Interpreting Wittgenstein: An Introduction” in *Wittgenstein and His Interpreters*, pp 1-36, (Oxford: Blackwell Publishers)

Kahane, G. and Savulescu J. (forthcoming) “The Welfarist Account of Disability” in K. Brownlee and A. Cureton *Disability and Disadvantage*, (Oxford: Oxford University Press)

Kahane, G. (2006) “Pain, Ethical Significance of” in D. Borchert, ed., *The Encyclopaedia of Philosophy*, 2nd Edition, (Macmillan)

Rebecca Roache

Journal articles

Roache, R., and Clarke, S., (2008) “Bioconservatism, Bioliberalism, and the Wisdom of Reflecting on Repugnance” *Journal of Applied Philosophy*, forthcoming

Roache, R., (2006) “A Defence of Quasi-Memory”, *Philosophy* 81, pp 323-355

Roache, R., (2007) Review of Jonathan Glover’s *Choosing Children*, in *Philosophical Books*, forthcoming

Roache, R., (2007) “Self-Esteem, Mood Enhancement, and Human Flourishing”, to appear in a special issue of a journal (details currently in negotiation) dedicated to the proceedings of a University of Ghent workshop entitled ‘Neuroenhancement of Mood: Social and Ethical Issues at the Forefront of the Debate’, forthcoming

Journal articles in preparation/submitted

Roache, R., (2007) “Fission and Survival”, submitted to *Erkenntnis* on 2 May 2007.

Viewable online at:

[http://www.fhi.ox.ac.uk/Papers/Fission%20and%20Survival%20\(FHI\).pdf](http://www.fhi.ox.ac.uk/Papers/Fission%20and%20Survival%20(FHI).pdf).

Roache, R., (2007) “Human Enhancement and the Risk of Social Disruption”, in preparation

Roache, R., (2007) “What is Human Nature?”, in preparation

Roache, R., (2007) “Human Nature, Self-Interest, and Enhancement”, in preparation

Contributed book chapters, conference proceedings, and articles

Roache, R., and Bostrom, N., (2007) “Ethical Issues in Human Enhancement” in C. Wolf, J. Ryberg and T. Petersen (eds.) *New Waves in Applied Ethics* (Palgrave Macmillan), forthcoming. Viewable online at <http://www.nickbostrom.com/ethics/human-enhancement.pdf>

Reprints/translations

Roache, R., (2008) “Bioconservatism, Bioliberalism, and the Wisdom of Reflecting on Repugnance” (see above entry under ‘Journal articles’) is to appear in German translation in a Volume on the Ethics of Enhancement by N. Knoepffler (ed) and published by Verlag Karl Alber

Anders Sandberg

Journal articles

Sandberg, A., Chakraborty, S., and Greenfield, S., (2007) “Differential Dynamics of Transient Neuronal Assemblies in Visual Compared to Auditory Cortex” *Experimental Brain Research* (accepted)

Sandberg, A., and Bostrom, N., (2006) “Converging Cognitive Enhancements” in W. Bainbridge and M. C. Roco (eds) *Ann. N.Y. Acad. Sci (Special Issue: Progress in Convergence: Technologies for Human Wellbeing)* 1093: 201–227

Sandberg, A., and Bostrom, N., (2006) “Cognitive Enhancement: Methods, Ethics, Regulatory Challenges”. In the proceedings of *Forbidding Science? Balancing Freedom, Security, Innovation and Precaution Conference*, (Arizona State University's College Law Center for the Study of Law, Science & Technology, January 12 and 13)

Journal articles in preparation

Liao, S.L., & Sandberg, A., (2007) “The Normativity of Memory Modification” (submitted to *Philosophical Psychology*)

Sandberg, A., and Savulescu, J., (2007) “Mozart, Folic Acid, Choline and Genetic Modification: Prenatal Cognitive Enhancement” (to be submitted)

Sandberg, A., and Savulescu, J., (2007) “Intelligence and Happiness” (to be submitted)

Sandberg A., and Savulescu, J., (2007) “Performance enhancing drugs and marriage: the chemicals between us” (to be submitted)

Ravelingien, A., and Sandberg, A., (2007) “Sleep better than medicine? Ethical and philosophical issues related to ‘wake enhancement’” (to be submitted)

Sandberg, A., (2007) *Definitions of Enhancement, Review of Cognitive Enhancement Technologies, Review of Ethical Topics in Cognitive Enhancement*, reports from ENHANCE Project, forthcoming

<http://www.enhanceproject.org/Internal/Cognitive%20Enhancement%20Review.pdf>

Contributed book chapters, conference proceedings, and articles

Bostrom N., and Sandberg, A., (2007) “The Wisdom of Nature: An Evolutionary Heuristic for Human Enhancement” in N. Bostrom and J. Savulescu (eds) *Enhancing Humans*, (Oxford: Oxford University Press) <http://www.nickbostrom.com/evolution.pdf>

Sandberg, A., (2007) “Biotechnology and the promise of tailor-made medicine” in F. Fici (ed) *Unlocking Ideas: Essays from the Amigo Society*, (London, Stockholm Network 2007), pp 61-67

Sandberg, A., (2006) “Den forstaerkede hjerne” in G. Balling & K. Lippert-Rasmussen (eds), *Det Menneskelige Eksperiment*, (Museum Tusculanum Forlag: Denmark) pp 75-114

Reprints/translations

Sandberg, A., (2007) “Cognitive Enhancement: Can we afford to ban it?” , forthcoming in proceedings of the *Enhancement and Genetics Conference* (Friedrich-Schiller-University Jena, Alber Verlag, June, 22-24) (work being translated)

Reports

Sandberg, A., and Bostrom, N., (2006) “Cognitive Enhancement: A Review of Technology”. *Report for the ENHANCE Project (first draft)*

Journalism

Sandberg, A., (2007) “Cognitive Enhancement: Can we afford to ban it?” , forthcoming in proceedings of the *Enhancement and Genetics Conference* (Friedrich-Schiller-University Jena, Alber Verlag, June, 22-24) (work being translated)

Sandberg, A., (2007) “Cognition enhancement: Upgrading the Brain”, forthcoming in ENHANCE project anthology

Sandberg, A., and Savulescu, J., (2007) “Cognition Enhancement and Happiness”, forthcoming in ENHANCE project anthology

Sandberg, A., Savulescu, J., and Bostrom, N., (2007) “The Economic and Social Impact of Enhancement”, forthcoming in ENHANCE project anthology

Sandberg, A., (2007) “The Blue or the Pink Pill: Are Enhancements Gendered?”
Forthcoming in ENHANCE project anthology

Nicholas Shackel

Journal articles

Shackel, N., (2007) “Parting smoothly?”, *Analysis*, forthcoming

Shackel, N., (2007) “Bertrand's Paradox and the Principle of Indifference”, *Philosophy of Science*, forthcoming

Shackel, N., (2007) “Pragmatism and sophism”, *Logique et Analyse*, forthcoming

Shackel, N., (2006) “Shutting Dretske's door”, *Erkenntnis*, (Accompanied by a reply from Dretske), Volume 64, October 2006, pp 393-401

Shackel, N., and Clark, M., (2006) “The Dr. Psycho Paradox and Newcomb's Problem”, *Erkenntnis*, Volume 64, pp. 85-100

Journal articles in preparation/under consideration

Shackel, N., (2007) “Two kinds of Normativity”, *Ethics*

Shackel, N., (2007) “Pluralism for the normativity of rationality”, *Mind*

Shackel, N., and Kahane, G., (2007) “Utilitarian bias or defective moral dilemmas?”, *Nature*

Shackel, N., (2007) “Epistemic ethics”

Shackel, N., (2007) “Epistemic ethics and public knowledge”

Shackel, N., (2007) “The instrumental rationality model for the relation of policy and science”

Shackel, N., (2007) “Epistemic blame and epistemic duty”

Shackel, N., and Hillerbrand, R., (2007) “Epistemic ethics and complex systems”

Shackel, N., and Ravetz, J., (2007) “Ethics of expertise”

Shackel, N., and Ravetz, J., (2007) “What is a citizen to believe?”

Shackel, N., and Liao, M., (2007) “Disagreement and Philosophical Bayesianism”

Contributed book chapters, conference proceedings, and articles

Shackel, N., (2007) “Paradoxes of Probability” in T. Rudas (ed) *Handbook of Probability Theory with Applications*, (Sage)

Shackel, N., Faria, M., and Kahane, G., (2007) “Conceptual problems in the scientific study of belief”, in C. Jaeger (ed) *Brain -- Religion -- Experience. Multidiscipline Encounters* (Dordrecht, New York: Springer)

Peter Taylor

Journal articles

Taylor, P., (2006) “The Lighthill Risk Network”, *Mathematics Today*, December 2006

Taylor, P., (2006) “The Lighthill Risk Network”, *Catastrophe Risk Management*, April edition

Journal articles in preparation

Taylor, P., (2007) “Assessing catastrophic risk,” in preparation

Taylor, P., (2007) “Unprecedented risks,” in preparation

Taylor, P., (2007) “Uncertainty in Climate Change Prediction”, in preparation

Taylor, P., (2007) “Model choice”, in preparation

Contributed book chapters, conference proceedings, and articles

Taylor, P., (2007) "Catastrophes and Insurance" contributed chapter in N. Bostrom and M. Cirkovic (eds) *Global Catastrophic Risks*, (Oxford: Oxford University Press), forthcoming

Invited Lectures; Keynote Speeches; Conference Presentations

Nick Bostrom

Bostrom, N., (2007) “The Values that Should Guide us in Managing the Fast-Expanding Frontier of Science and Technology” Invited plenary speaker for Women's Forum 2007 (Deuville, France, 11-13 October)

Bostrom, N., (2007) “The Future of Humanity” Invited speaker at the *TransVision 2007 Conference* (Chicago, 24-26)

Bostrom, N., (2007) “Ethical Objections to Life Extension” Invited speaker at *Securing the Longevity Dividend Symposium* (Chicago, 23 July)

Bostrom, N., (2007) “The Future of Humanity” Keynote speaker at *Sedbergh Festival of Ideas* (Sedbergh, 20-22 July)

Bostrom, N., (2007) Invited to deliver the *11th Annual JUS Lecture* (Joint Centre for Bioethics, University of Toronto, October). Lectures are delivered by an internationally recognized major contributor to the advancement of genetics, neuroscience, psychiatry and its ethical implications". Previous speakers have included James D. Watson, Jean-Pierre Changeux, Anne Young, and Floyd Bloom

Bostrom, N., (2007) “My Challenges for the next 15 years” Featured speaker at the *Second Annual Global Creative Leadership Summit* – “a unique platform for the best minds of our generation”, organized by LTB Foundation with support from the UN Fund for International Partnerships (UNFIP) (New York City, 23-25 September)

Bostrom, N., (2007) “Enhancements: A Practical Approach” Invited speaker at *Enhancement and Genetics* (Jena, Germany, 22-24 June)

Bostrom, N., (2007) “Cognitive Enhancement: Methods, Ethics, and Challenges for Policy” Keynote presentation for *Oxford Forum for the Medical Humanities: Neuroethics Symposium* (Oxford, 11 May)

Bostrom, N., (2007) “Policy Issues” Invited session chair for *The Human Enhancement Colloquium* at the British Ambassador's Residence in the Hague (the Hague, 10 May)

Bostrom, N., (2007) “Dignity and Enhancement” Presentation for *Cognitive Enhancement Conference* organized by the ENHANCE project (Stockholm, 27-28 March)

Bostrom, N., (2006) “Ethical and Social Implications of Cognitive Enhancement” Invited presentation for *Cognitive Enhancement Workshop organized by the British Medical Association* (London, 24 November)

Bostrom, N., (2006) Panellist in debate on the topic “Will our Grandchildren be Robotic?” at the *BBC Festival of Ideas*, broadcast on Radio 3 (Liverpool, 5 November)

- Bostrom, N., (2006) “Dignity and Enhancement” presentation for the *James Martin Advanced Research Seminar* (Oxford, 20 October)
- Bostrom, N., (2006) “Human Enhancement and Sports Enhancement.” Invited presentation for the *Science and Technology Select Committee, House of Commons* (UK parliament) (London, 21 June)
- Bostrom, N., (2006) “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)
- Bostrom, N., (2006) “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)
- Bostrom, N., (2006) “Existential Risks and Artificial Intelligence” Invited keynote at the *Singularity Summit* (Stanford, 13 May)
- Bostrom, N., (2006) “Consequences of Cognitive Enhancement” *ENHANCE workshop* presentation (Oxford, 4 May)
- Bostrom, N., (2006). “The Big Picture for Humanity” Special invited forum speaker, *Rutherford Appleton Laboratory* (Abingdon, Oxfordshire, 6 October)
- Bostrom, N., (2006) “Political and Ethical Challenges for Society from the Prospect of Life-Extension.” Invited keynote address for the *2nd World Aging & Generations Congress* (St. Gallen, Switzerland, 27-29 September)
- Bostrom, N., (2006) “Wiser and Smarter.” Invited lecture for *Annual Investors Forum 2006*, organized by Oxford Capital Partners, Said Business School (Oxford, 20 September)
- Bostrom, N., (2006) “What is Enhancement?” Invited closing address for *TransVision 2006* (Helsinki, Finland, 17-19 August)
- Bostrom, N., (2006) “An Evolutionary Heuristic for Identifying Promising Human Enhancements” Invited opening plenary for *TransVision 2006* (Helsinki, Finland, 17-19 August)
- Bostrom, N., (2006) “A Practical Approach to Human Enhancement.” Satellite meeting to the *8th World Congress in Bioethics* (Beijing, 5 August)
- Bostrom, N., (2006) “The Future of Aging.” Invited lecture for the *Wellcome Trust* (London, 26 July)
- Bostrom, N., (2006). “Human Capital” Invited lecture for *The Royal Society for the Encouragement of Arts, Manufactures, and Commerce* (London, 22 March)
- Bostrom, N., (2006) “Cognitive Enhancement” Invited plenary presentation at the *World Forum for Science and Civilization* (Oxford, 14-17 March)
- Bostrom, N., (2006) “Existential Risks: what’s the probability that humanity will go extinct in the 21st century? What can we do to reduce the probability?” Invited presentation for the *World Forum for Science and Civilization* (Oxford, 14-17 March)

Bostrom. N., (2006) “Human Enhancement, Transhumanism, and Genetics.” Keynote address for *Great Expectations: On our genetic future* (Amsterdam, 21 February)

Bostrom. N., (2006) “Transhumanist Values.” Invited speaker at the *Institute for Science, Innovation & Society* (Nijmegen, 21 February)

Bostrom. N., (2006) “Cognitive Enhancement” Invited speaker at the *Forbidding Science: Balancing Freedom, Security, Innovation & Precaution* conference (Tempe, Arizona, January 10-11)

Bostrom. N., (2006) “The Transhumanist Vision.” Invited closing keynote presentation for *The Future of Human Nature: Science, Ethics, and Democracy* (University of Utah)

Bostrom, N., (2005) “Status Quo Bias in Bioethics.” Invited speaker for the DeCamp Seminar Series at the Princeton Center for Human Values (Princeton, 30 November)

Rafaela Hillerbrand

Hillerbrand, R., (2007) "Scale Separation as a Condition for Quantitative Modelling. Why Mathematics Works for some Problems and Fails for Others", presentation at *Models and Simulations 2* (Tillburg, Netherlands, October 2007)

Hillerbrand, R., (2007) "Uncertainty as a challenge for moral philosophy", presentation at *Societa Ethica Conference* (Leysin, Switzerland, August 2007)

Hillerbrand, R., (2007) "The communication of epistemic uncertainties", presentation at the *International Congress for Logic, Methodology, and Philosophy of Science* (Beijing, China, August 2007)

Hillerbrand, R., (2007) "Dianoetic Virtues in addressing a Morally Correct Treatment of GM", presentation at the conference *Environmental Justice and Global Citizenship* (Oxford, July 2007)

Hillerbrand, R., (2007) "The moral threat of a manmade climate change," invited talk at the conference *EE250: The Euler Equations: 150 years on* (Aussoi, France, June 2007)

Hillerbrand, R., (2007) "Nanotechnology – why worry", *James Martin Advanced Research Seminar* (Oxford, February 2007)

Hillerbrand, R., (2007) "Decision making under uncertainty. How to handle an anthropogenic greenhouse effect", *James Martin Advanced Research Seminar* (Oxford, February 2007)

Hillerbrand, R., (2006) "Uncertainty as a challenge for ethical reasoning," Rafaela Hillerbrand, contributed talk at the *Wittgenstein Symposium*, (Kirchberg am Wechsel, August 2006)

Guy Kahane

Kahane, G., (2007) “Pain, Experience and Frontal Lobotomy”, *Neuroethics Workshop*, Oxford University

Kahane, G. (2007) “Genetic Selection and Disability”, *The Oxford Medical Humanities Forum*, February

Kahane, G., (2007) “If Nothing Matters”, *The Moral Philosophy Seminar*, Oxford University, 26 February

Kahane, G., (2007) “Cognitive Enhancement: A Perspective From Value Theory”, *EU ENHANCE Project Workshop on Cognitive Enhancement*, Stockholm University, 27-8 March

Kahane, G., (2007) “A Welfarist Account of Disability”, *Disability and Disadvantage*, Manchester University (with Julian Savulescu), 12-13 May

Kahane, G., (2007) “Non-identity, Self-defeat, and Attitudes to Future People”, *Disability and Disadvantage*, Manchester University, 12-13 May

Kahane, G., (2006) “Psychology, Neuroethics and Society”, *Wellcome Trust Workshop*, Cambridge

Kahane, G., (2006) “Brain Reading and the Privacy of the Inner”, *Oxford Neuroethics Workshop*, Oxford University,

Kahane, G., (2006) “Procreative Beneficence and Disability”, *Pacific American Philosophical Association Meeting*, Portland, Oregon, 24 March

Kahane, G., (2006) “Cognitive Enhancement: A Perspective from Value Theory”, *Enhance Workshop*, St Cross College, Oxford

Rebecca Roache

Roache, R., (2007) “Self-Esteem, Mood Enhancement, and Human Flourishing”, at *Neuroenhancement of Mood: Social and Ethical Issues at the Forefront of the Debate* (University of Ghent, Belgium, 29 November 2007)

Roache, R., (2007) “Ethical Issues in Mood Enhancement”, *Moral Philosophy Seminar*, (University of Oxford, 15 October 2007)

Roache, R., (2007) “Bioconservatism, Bioliberalism, and the Wisdom of Reflecting on Repugnance”, at *Enhancement and Genetics Conference* (University of Jena, 22 June 2007) (unable to attend, presentation given by co-author, Steve Clarke)

Roache, R., (2007) Response to “Personal Identity and Life Span Extension” by G. Barazzetti and M. Reichlin, *ENHANCE workshop on Life Span Extension* (Università Vita-Salute, San Raffaele, Italy, 17-18 May 2007)

Roache, R., (2007) “Cognitive Bias and Human Enhancement”, *James Martin Advanced Research Seminar* (Oxford)

Roache, R., (2007) “Enhancement and the Risk of Social Disruption”, *James Martin Advanced Research Seminar* (Oxford)

Roache, R., (2006) “Human Nature and Enhancement”, *James Martin Advanced Research Seminar* (Oxford)

Anders Sandberg

Sandberg, A., (2007) Participated in *Mind and Body 2025* dialogue event (Dana Centre, London Science Museum, 28 June)

Sandberg, A., (2007) “The Blue or the Pink Pill: are Enhancements Gendered?”, talk held at the *ENHANCE Workshop on Cognitive Enhancement*, (Stockholm 27-28 April)

Sandberg, A., (2007) “Scenario planning for life extension”, talk held at *Extrobritannia*, (Conway Hall, London, 13 May)

Sandberg, A., (2007) “Are Enhancements Gendered?”, talk held at the *ENHANCE Workshop on Mood Enhancement*, (Maastricht April 26-28)

Sandberg, A., (2007) “The Chemicals Between Us”, lecture at the *Neuroethics Symposium, Oxford Forum for Medical Humanities*, (St. John College, Oxford, 11 May)

Sandberg, A., (2007) “Unfair to Allow or Unfair to Ban?”, lecture on the economic impact of cognitive enhancement *Friedrich-Schiller-University* (Jena, Germany, 23 July)

Sandberg, A., (2006) “The Social Impact of Cognitive Enhancement”, talk held at the *ENHANCE Conference* (Beijing, August 5)

Sandberg, A., (2006) “Mozart, Folic Acid, Choline and Genetic Modification: Prenatal Cognitive Enhancement”, talk held at *The 8th World Congress of Bioethics*, (Beijing, August 6)

Sandberg, A., (2006) “Memory Enhancement: what, how and why”, talk held at *Extrobritannia*, (Conway Hall, London, 25 February)

Sandberg, A., (2006) “Biotechnology and the promise of tailor-made medicine” at the *Amigo Society*, (Brussels, 21 February) as part of ‘Segundas Jornadas sobre Convergencia Ciencia-Tecnología,’ sponsored by the Vodafone Foundation

Sandberg, A., (2006) “The Transhumanist Vision” (Universidad de Alcalá de Henares, 9 March) as part of ‘Segundas Jornadas sobre Convergencia Ciencia-Tecnología,’ sponsored by the Vodafone Foundation

Sandberg, A., (2006) “Memory Modification and Authenticity” at the *Human Enhancement Technologies and Human Rights* conference (Stanford University Law School, San Francisco, 26-28 May)

Sandberg, A., (2006) “Cognitive Divide or a Mind-Meld?: Scenarios of Cognitive Enhancement”, at the *Transvision 2006 conference*, (Helsinki University, 17-19 August)

Sandberg, A., (2006) “Genius as a commodity: cognitive enhancement technology and scenarios of its social effects”, *2006 Annual Meeting of the Society for the Social Studies of Science*, (Vancouver, 1-5 November)

Nicholas Shackel

Shackel, N., (2007) “Epistemic blame, public knowledge and epistemic duty”, *James Martin Advanced Research Seminar* (Oxford)

Shackel, N., (2006) “Society, artificial persons, rights and responsibility”, *James Martin Advanced Research Seminar*, (Oxford)

Shackel, N., (2006) “Reasons, rationality and externalism”, Joint Session of the *Mind and Aristotelian Societies*, (Southampton)

Shackel, N., (2006) “Rhetorical manoeuvres, rationalism and sophism”, to the *Logic and Rhetoric Conference*, (Cambridge)

Shackel, N., (2006) Pluralism for the normativity of rationality”, *Rationality and Normativity Seminar*, (Oxford)

Shackel, N., (2006) On the obligation to be rational”, *Oxford Moral Philosophy Seminar*, (Oxford)

Shackel, N., (2006) “Relations of belief and consciousness”, to *Oxford Centre for the Science of Mind*, (Oxford)

Peter Taylor

Taylor, P., (2007) “Unprecedented Excessive Risks,” contributed talk, *James Martin Advanced Research Seminar* (Oxford, January 2007)

Taylor, P., and Hillerbrand, R., (2007) “Global catastrophic Risks. Nanotechnology,” contributed talk *James Martin Advanced Research Seminar* (Oxford, February 2007)

Taylor, P., (2007) "Ten Key Issues in Catastrophe Modelling" talk at the *RAA/IUA Conference* (London, 20 June)

Taylor, P., (2007) "Emerging Risks" talk at the *Risk Modelling for P&C Insurers Conference* (London, 12 April)

Taylor, P., (2007) "Ten Challenges in Catastrophe Modelling" talk at the *RAA/IUA Conference* in (London, 21 June)

Workshops and Forums¹

FHI International Methodology Workshop

Oxford (13 March 2006)

On 13 March 2006, in advance of the James Martin Institute inaugural 2006 World Forum, the FHI held a ‘Big Issues for Humanity’ advanced methodology workshop. Speakers included Joel Garreau (Washington Post), Julian Savulescu (University of Oxford), James Hughes (Trinity College, Connecticut), William Bainbridge (National Science Foundation), and Nick Bostrom (FHI)

ENHANCE Workshops – Cognition Enhancement

Oxford (4 May 2006)

Anders Sandberg helped organise the ENHANCE Workshops on cognition enhancement in Oxford (4 May 2006) and Stockholm (27-28 March 2007)

ENHANCE Workshops – Cognition Enhancement

Stockholm (27-28 March 2007)

Anders Sandberg helped organise the ENHANCE Workshops on cognition enhancement in Oxford (4 May 2006) and Stockholm (27-28 March 2007)

Whole Brain Emulation Workshop

St Hilda’s College, Oxford (26-27 May 2007)

Rebecca Roache, Nick Bostrom and Anders Sandberg organised an FHI workshop ‘Whole Brain Emulation’, dedicated to estimating when and how an emulation of a whole human brain might be possible. Attended by an international panel of neuroscientists and relevant researchers, and followed by the ongoing development of a ‘roadmap’ document. (University of Oxford, 26-27 May 2007)

¹ See Annexe 3 for further details of FHI Workshops and Forums

Bayesian Approaches to Agreement Conference

Pembroke College, Oxford (4 June 2007)

Nicholas Shackel organized an international conference for those interested in Bayesian approaches to agreement and disagreement. Richard Bradley of L.S.E presented his paper comparing deliberation and aggregation as ways of dealing with disagreement and producing collective opinion. Robin Hanson of George Mason University presented his recent research on disagreement, in which he has been extending Aumann's theorem to conclude that rational disagreement requires origin disputes. Christian List and Franz Dietrich of L.S.E. presented the paper they have written with Richard Bradley "Aggregating Causal Judgments". The occasion was a very interesting exposure of the Bayesian approach to the issue and provoked considerable discussion among the participants. We are looking at publishing the papers as special edition of a journal: possibly the Knowledge, Rationality and Action section of Synthese.

Conference 'EE250: The Euler Equations, 250 years on'

Aussois, France (June 2007)

Rafaela Hillerbrand assisted in organizing a conference held on the tercentenary of the birth of Leonhard Euler and the 250th anniversary of his seminal publications on fluid mechanics. The conference will cover the latest research within fluid mechanics related to the modelling and predictability of nonlinear systems. The conference was held in Aussois, France in June 2007, and was organized by Uriel Frisch (Observatoire de la Côte d'Azur, Nice) and funded by CNRS and under the patronage of the French Academy of Science.

Everett@50 Conference

Oxford (19 – 21 July 2007)

Peter Taylor was Conference Manager for the Everett@50 Conference in Oxford held between 19th through 21st July 2007 to bring together the leading philosophers and physicists interested in the interpretation of quantum mechanics to discuss Everett's theory on the 50th Anniversary of publication of the "relative state formulation of quantum mechanics", sometimes called the many-worlds or multiverse interpretation. Oxford philosophers have led the revival of the Everett interpretation in the past ten years, and this Conference will see if Everett's explanation of quantum mechanics has at last come of age. Aspects of the conference from an administrative point of view include a live webcast, will have an on-line Blog, and is being included in a BBC4 documentary on Everett to be broadcast in November 2007. Website: <http://users.ox.ac.uk/~everett/>

Existential Risk Workshop

Oxford (autumn 2007)

Nick Bostrom and Rafaela Hillerbrand are organizing workshop on Existential Risks to be held in Oxford in autumn 2007. Leading experts on different existential risks will be invited.

Cross-Disciplinary Seminar on Risks

Oxford (MT07 and HT08)

Peter Taylor and Rafaela Hillerbrand are organizing a programme of cross-disciplinary seminars for the James Martin 21st Century School on risk, including emerging risks, in the Academic Year 2007/2008 Michaelmas and Hilary terms. This will cover such areas as the effect of new technologies as well as the socio-economic and political response to disruptive change.

Human Nature

Hong Kong (December 2007)

Rebecca Roache is helping Matthew Liao from our sister JM institute, BEP, to organise ‘Human Nature and Bioethics’ conference at City University, Hong Kong, China, in December 2007. This included making a successful application for funding to the British Academy. To date, confirmed speakers at the conference are Jonathan Glover (King’s College, London), Jeff McMahan (Rutgers), John Harris (Manchester), Dan Brock (Harvard), Dan Wikler (Harvard), Ingmar Persson (Göteborg), Jo Wolff (University College, London). Conference on ‘Global catastrophic risks’. With the launch of the Oxford University monograph on global catastrophic risk edited by Nick Bostrom and Milan Cirkovic, there will be a conference hosted by the FHI with all the contributing authors as well as invited speakers.

Forthcoming

Rebecca Roache is currently working with the James Martin Institute of Ageing to identify possibilities for collaboration. One possibility is a jointly-organised workshop to predict the social impact of life-extension technology.

Public Outreach

Nick Bostrom presented a version of his paper *Are We Living in a Computer Simulation?* to a group of schoolchildren, as part of *Vice Magazine's* project to introduce children to ideas about the future. Nick asked the children what kind of people they would create, and how they envisaged the future, and discussed the possibility of living in a computer simulation, and what it would be like. This was part of the magazine's project to introduce children to academic ideas about the future.

“NEURObotics: the future of thinking?”, exhibition was launched on October 10. Nick Bostrom and Anders Sandberg had been involved in setting up this exhibition with the producers, and Anders Sandberg was science advisor for the project, which was part of the Wellcome trust wing of the London Science museum.

F.H.I. is sponsoring a new web forum (i.e., blog), which began November 20, 2006. Called Overcoming Bias, it is "A forum for those serious about trying to overcome their own biases in beliefs and actions." As of July 17, 2007 it has 40 contributors who had made 215 posts, 240,136 visits, and 500,563 page views. The FHI blog was The Economist's 'BLOG OF THE WEEK' at Christmas 2006. www.overcomingbias.com

EE250 discussion group (<http://groups.google.com/group/ee250>). The website was set up by Rafaela Hillerbrand in connection with the above mentioned EE250 conference. It give participants of the conference as well as others interested in the topic the opportunity to discuss questions related to the modeling of hydrodynamical nonlinear systems. This touches for example on questions related to climate.

Rebecca Roache gave a presentation and led a discussion on ethical aspects of cognitive enhancement with members of the public for the Academy of Medical Sciences 'Drugs Futures' project, which explored public views on the sort of drug culture we want for the future. (30 March 2007)

Anders Sandberg has designed and maintains www.enhanceproject.org, the ENHANCE project website. This includes online forum for document sharing among project members, a public blog and an emerging wiki database about enhancement.

Media

Nick Bostrom

Bostrom, N., (2007) *TV documentary*. Interviewed about life extension and biogerontology.

Bostrom, N., (2007) *The Sunday Times*. Interviewed about cognitive enhancing drugs, particularly Modafinil, and its use by some academics

Bostrom, N., (2007) *BBC Radio 3*, ‘The Essay’ – lecture on extraterrestrial intelligence (19 July)

Bostrom, N., (2007) *Ci’num* (Digital Civilizations Forum). Interviewed about existential risks, human enhancement, public engagement with emerging technologies, and utopian visions

Bostrom, N., (2007) *BBC Radio 2*. Interviewed about the future of the Internet and its impacts, for the series *Why didn’t I think of that?*

Bostrom, N., (2007) *GQ Magazine*. Long feature article on me and my work and the FHI

Bostrom, N., (2007) *Times Higher Educational Supplement*. Interviewed about cognitive enhancement medicine, including practical and ethical issues, and implications for policy

Bostrom, N., (2007) *Le Devoir*. Interviewed about my reactions to the Body Worlds exhibition

Bostrom, N., (2007) *BBC Television*. Interviewed for two documentaries about biotechnology and nanotechnology, about expected developments in these fields and what they might mean for the future of humanity

Bostrom, N., (2007) *Seed Magazine*. Interviewed about robot ethics

Bostrom, N., (2007) *Toronto and Ottawa Suns* (Canadian newspapers). Asked about the future of surveillance technology

Bostrom, N., (2007) *The Independent* (UK newspaper). Asked about efforts by industry to develop ethics codes for robots

Bostrom, N., (2007) *The Telegraph*, Interviewed about use of isotopes to slow the ageing process.

<http://www.telegraph.co.uk/news/main.jhtml?xml=/news/2007/03/26/norganic126.xml>

Bostrom, N., (2007) *L’Hebdo* (Swiss newsmagazine). Asked about cyborg technologies such as implants versus other ways of enhancing human performance

Bostrom, N., (2007) *Associated Press*. Interviewed about transhumanism, future technologies, and Fukuyama’s critique

Bostrom, N., (2007) *Financial Times*. Interviewed about “where philosophy is going” today

Bostrom, N., (2007) *BBC World Service* (radio). Commenting on three topics: the “doomsday vault” (seed bank) in Svalbard, Microsoft’s “immortal computing” project, and plans to send out messages aimed for extraterrestrial civilizations

Bostrom, N., (2007) *PC Plus magazine*. Interviewed about pervasive computing

Bostrom, N., (2007) *Chemistry World* (magazine). Interviewed about the ethics of life extension research in conjunction with a forthcoming paper in the journal *Rejuvenation Research*

Bostrom, N., (2007) *Utilidnings Radion* (Swedish educational radio). Interviewed about the future of human evolution, the impacts of transformative technologies, and transhumanist ethics

Bostrom, N., (2007) *CLOSER TO TRUTH* (public television / PBS series). Four-hour interview covering topics for many aspects of my work, with material to be included in at least five programs: the simulation argument, anthropic reasoning, the future of intelligent life, multiverse theories, the Doomsday argument, etc

Bostrom, N., (2007) *Drivetime With Dave Fanning on RTE Radio One* (Irish radio). Interviewed about human enhancement and nanotechnology

Bostrom, N., (2007) *Urbania* (Montreal-based magazine). Short interview about what the priorities should be on the environmental agenda

Bostrom, N., (2007) *Boston Globe*. Interviewed about existential risks, and about the invocation of the term “existential threat” in relation the war on terror

Bostrom, N., (2007) *San Diego Union-Tribune*. Interviewed about human evolution and its possible future directions

Bostrom, N., (2007) *The Sunday Herald* (Scottish newspaper). Interviewed about life-extension and transhumanism

Bostrom, N., (2007) *Drivetime With Dave Fanning on RTE Radio One* (Irish radio). Interviewed about the simulation argument

Bostrom, N., (2007) *Odd at Large* (Swedish Magazine). Interviewed about the activities of the Oxford Future of Humanity Institute, global catastrophic risks, and the simulation argument

Bostrom, N., (2007) *BBC Focus Magazine*. Interviewed about the simulation argument

Bostrom, N., (2007) *Fast Thinking* (Australian magazine). Interviewed about DNA technology and about the possibility of eradicating aging and disease

Bostrom, N., (2007) *The Today Programme* (BBC Radio 4). Interviewed about the role of instincts and moral intuitions in bioethics and discussions about human enhancement

Bostrom, N., (2006) *McCalmont’s web forum*. 2006. Interviewed about my background and miscellaneous topics

Bostrom, N., (2006) *The Today Programme* (BBC Radio 4). Interviewed about gene doping and performance enhancement in sport

- Bostrom, N., (2006) *Discovery Channel* (United States). Interviewed about cyborg technology, uploading, and the future of humanity
- Bostrom, N., (2006) *PIMM* (book blog). Interviewed about life extension
- Bostrom, N., (2006) *BBC Focus Magazine*. Interviewed about cosmological problems and the simulation argument
- Bostrom, N., (2006) *Personal Computer World*. Interviewed about transformative future technologies
- Bostrom, N., (2006) *Mongrel Magazine*. Interviewed about transhumanism
- Bostrom, N., (2006) *Cryonics Magazine*. Interviewed about ethical issues related to the practise of cryonic suspension
- Bostrom, N., (2006) *BBC Radio 3*. Hour-long debate about the future of human enhancement and robotics, part of the BBC Festival of Ideas in Liverpool
- Bostrom, N., (2006) *The Times*. Interviewed about human enhancement and associated ethical issues
- Bostrom, N., (2006) *Independent documentary film*. Interviewed about future energy sources
- Bostrom, N., (2006) *Nature*. Interviewed about the conference TransVision 2006 and my presentations therein
- Bostrom, N., (2006) *Italian National Television*. Interviewed about the future and about philosophical questions related to artificial intelligence and human enhancement
- Bostrom, N., (2006) *New Scientist*. Interviewed about my work on the evolution heuristic
- Bostrom, N., (2006) *BBC Radio Five*. Interviewed about performance enhancement in sport
- Bostrom, N., (2006) *Monthly Vision* (Italian magazine). Interviewed about artificial intelligence and where it will go in the future
- Bostrom, N., (2006) *Maxim magazine*. Interviewed about future technologies and body modification
- Bostrom, N., (2006) *CNN Future Summit*. Interviewed about the present state of artificial intelligence and its future prospects
- Bostrom, N., (2006) *National Journal*. Interviewed about the future of intelligent machines, the possibility of a technological singularity, and the implications for governance and public policy
- Bostrom, N., (2006) *The Times* (UK newspaper). Interviewed about life-extension research and the desirability of longer life
- Bostrom, N., (2006) *The Times* (UK newspaper). Interviewed about memory and how new technology may change the demands on human memory

- Bostrom, N., (2006) *Wellcome Trust science museum*. Background interview about cognitive enhancers
- Bostrom, N., (2006) *Technocalypse* (film documentary). Interviewed about status quo bias, human rationality, and human enhancement
- Bostrom, N., (2006) *The Sunday Times*. Interviewed about the impacts of growing up with digital technology on brain development and psychology
- Bostrom, N., (2006) *Human Values in a Transhuman World* (radio documentary). About ethics, human enhancement, and new technologies
- Bostrom, N., (2006) *Meme Therapy* (blog). Interviewed about transhumanism and related issues
- Bostrom, N., (2006) *The Next Paradigm* (TV documentary). Interviewed about the singularity and the future of artificial intelligence
- Bostrom, N., (2006) *TV documentary*. On transhumanism and related issues
- Bostrom, N., (2006) *French Feature Film*. Interviewed about aging and life-extension
- Bostrom, N., (2006) *Autopilots* (TV documentary). Interviewed about the future of robotics and artificial intelligence
- Bostrom, N., (2006) *Bon Magazine*. Interviewed about memory enhancing and memory deleting drugs and their social and ethical ramifications
- Bostrom, N., (2006) *Isis Magazine*. Interviewed about what kinds of technological change students alive today can expect to experience in their lifetime
- Bostrom, N., (2006) *Eureka* (French magazine). Interviewed about human nature
- Bostrom, N., (2006) *Muy interesante* (popular science magazine for Central America). Main feature (15-20 pages) on my work on existential risks
- Bostrom, N., (2006) *The Observer* (UK newspaper). Interviewed about my RSA lecture and the prospects of human enhancement technologies
- Bostrom, N., (2006) *Galileu* (Brazilian science magazine). Interviewed about the science of life extension
- Bostrom, N., (2006) *Guardian* (UK newspaper). Feature interview about the work of the Oxford Future of Humanity Institute, and about transhumanism
- Bostrom, N., (2006) *CNN Future Summit*. Interviewed about a variety of future-related topics
- Bostrom, N., (2006) *The Meaning of the 21st Century* (TV documentary). Reading from my “Letter from Utopia”
- Bostrom, N., (2006) *Philosophy Now*. Interviewed about David Pearce’s work
- Bostrom, N., (2006) *CBC* (Canadian Broadcasting Corporation television). Interviewed about the future of aging

Bostrom, N., (2006) *Science & Avenir* (French science magazine). Interviewed about the Future of Humanity Institute and transhumanism

Bostrom, N., (2006) *London Update*. Interviewed about transhumanism

Bostrom, N., (2006) *ABC News documentary*. Interviewed about global catastrophic risks and threats to human survival

Bostrom, N., (2006) *The Future* (Dutch National Television). Followed for one day and interviewed about my work and about transhumanism

Bostrom, N., (2006) *BBC World Service*. Interviewed about aging and life-extension

Bostrom, N., (2006) *The Today Program (Channel 4, UK Radio)*. Interviewed about human enhancement

Bostrom, N., (2006) *Good Morning Scotland (BBC Radio)*. Interviewed about life-extension

Bostrom, N., (2006) *Radio Oxford*. Interviewed about the ethics and science of life-extension

Bostrom, N., (2006) *Catastrophe* (UK magazine). Interviewed about the Future of Humanity Institute and its interdisciplinary work

Bostrom, N., (2006) *Volume kskraut* (Dutch newspaper). Interviewed at length about my work, and about human enhancement technologies

Bostrom, N., (2006) *Discover Magazine*. Interviewed in relation to Nature publication on the cosmic disaster frequency

Robin Hanson

Hanson, R., (2007) *The Corporate Board*, “Spoken and Written” ,p 30 (January/February 2007)

Hanson, R., (2007) *Forbes.com* “Catch The Carbon, Win a \$25M Prize”, (9 February 2007)

Hanson, R., (2007) *New York Times* B1, “You Want Innovation? Offer a Prize” (31 January 2007)

Rafaela Hillerbrand

Hillerbrand, R., (2007) Invited talk on “Global Catastrophic Risks” for Chinese Journalists who have been assigned a journalist prize by Elsevier, (April 2007)

Guy Kahane

Kahane, G., (2007) *Times Higher Education Supplement*, “Meditations on the Flourishing and the Fallen”, (11 May 2007)

Rebecca Roache

Roache, R., (2007) *BBC Radio 4*, “The Defeat of Sleep”. Interviewed about cognitive enhancement (16 April)

Roache, R., (2007) *American GQ magazine*. Interviewed about the future. (June)

Anders Sandberg

Sandberg, A., (2007) Interviewed by Linus Brohult, Editor-in-Chief of *Mobil Magazine* on cognition enhancement, privacy and human-machine symbiosis for report on “The Mobile Human 2.0” (23 February)

Sandberg, A., (2007) Appearance on NBC program “Dawn of the robot age?” (21 February)

<http://www.msnbc.msn.com/id/17244922/>

Sandberg, A., (2007) Interviewed by Jorun Modén, health editor, on pharmacological enhancement of memory and love relations for the e-health.se newsletter (16 February)

Sandberg, A., (2007) Participated as expert in a public consultation on enhancement drugs in Glasgow, part of the Academy of Medical Sciences’ project *Drugsfutures* (3 March)

Sandberg, A., (2007) Interview for *Gate Report Magazine* on the future of brain research (8 March)

Sandberg, A., (2007) Interview Maria Cheng, AP on transhumanism, enhancement and future studies (12 March)

Sandberg, A., (2006) Participated in the Delphi study *Education and Learning: Possibilities by 2030 organized by the UN Millennium Project*.

<http://www.realtimedelphi.com/STUDIES/education/kedu.php>

Sandberg, A., (2007) Interview by Tomas Lindblad in “Allt om Vetenskap” (Swedish popular science magazine) issue 6/7-2007 p. 95 about the survival of humanity

Sandberg, A., (2006) Interviewed for BBC channel 4, *Leading Edge* on “The Future of Thinking?” and BBC News, “Science has designs on your brain” (10 October)

<http://news.bbc.co.uk/1/hi/health/5410092.stm>

<http://www.bbc.co.uk/radio4/science/leadingedge.shtml>

Sandberg, A., (2006) Film crew from 3sat filmed for the science program *Nano’s series on visionaries*. The program aired in November 2006 (8-9 September)

Sandberg, A., (2006) Interview on Monitor, Swedish Radio on artificial intelligence and AI ethics (10 October).

<http://www.sr.se/cgi-bin/p2/program/index.asp?programID=2098>

Sandberg, A., (2006) Article in perfil.com, “A un paso de conseguir superhumanos” about the ENHANCE project by Martín De Ambrosio. (4 October)

Sandberg, A., (2006) Short radio interview on neuroethics and enhancement, Swedish radio, TV 5 (31 August)

Sandberg, A., (2006) "The honorable tag", seminar at Hitachi Sweden on RFID tag policy and identity technology. This also led to a radio interview aired on Swedish radio (23 May)

Sandberg, A., (2006) Interview, Swedish youth radio program "Stjärnstopp" about life extension, cryonics and identity

Sandberg, A., (2006) Anders and Nick Bostrom were interviewed by Belgian filmmaker Frank Theys for an upcoming documentary

Sandberg, A., (2006) Lectured on the social impact and ethics of life extension at a seminar organized by Eudoxa at Uvvy Island in the virtual reality world Second Life (18 December).

<http://www.eudoxa.se/content/archives/Keepraging.pdf>

Sandberg, A., (2006) Participated in debate program “The Philosophical Room” on Swedish national radio, discussing human enhancement with professor Lars Bergström (philosophy), professor Maria Strömme (nanotechnology) and bishop Antje Jackelén (31 December)

Sandberg, A., (2006) Participated in “Our Sci-Fi Future”, a dialogue event at the Dana centre at the London Science Museum (10 January).

<http://www.danacentre.org.uk/events/2007/01/10/215>

Sandberg, A., (2006) Lectured on “New Media, New Brains” at Thames Valley University (12 February)

Sandberg, A., (2006) “Doubting Ageing”, a popular article in the magazine *Persuader* June 2006 about the ethics and consequences of life extension

Government, policy and advice

Nick Bostrom

Bostrom, N., (2007) Advising the UK's *National Endowment for the Sciences, Technology, and Arts* on developing a new 5 year £25m Talent Fund

Bostrom, N., (2007) Advising the *Parliamentary Office of Science and Technology* (POST) on the subject of cognitive enhancers and associated ethical issues

Bostrom, N., (2007) 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

Bostrom, N., (2007) Attended a London brainstorming meeting with organizers to advise on and select a science and technology topic for the next *World Economic Forum in Davos*

Bostrom, N., (2007) 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?”

Bostrom, N., (2007) Advising the UK's *National Endowment for the Sciences, Technology, and Arts* on developing a new Talent Fund

Bostrom, N., (2007) Advising the *Parliamentary Office of Science and Technology* (POST) on the subject of cognitive enhancers and associated ethical issues

Bostrom, N., (2006) Advisory roundtable on cognitive enhancement for the *British Medical Association and the Royal Institution* (London, November, 2006)

Bostrom, N., (2006) Advising the *European Commission* on the implementation of the information and communication technologies (ICT) theme in the *Community 7th Framework Programme for Research and Technological Development (2007-2013)*, (Brussels)

Bostrom, N., (2006) London brainstorming meeting with organizers to select science and technology topic for the next *World Economic Forum in Davos*

Bostrom, N., (2006) Expert advisor for the *STOA-panel of the European Parliament* (Brussels) on NBIC convergence and human enhancement

Bostrom, N., (2006) Expert advisor for the *Science and Technology Select Committee, House of Commons* (UK Parliament) on human enhancement policy issues, particularly sports enhancement

Bostrom, N., (2006) Invited essay for the *President's Council on Bioethics* on the concept of human dignity and its application in current bioethics controversies

Rebecca Roache

Roache, R., (2007) *Academy of Medical Sciences*, ‘Drugsfutures’. Public consultation about cognitive enhancement. <http://www.drugsfutures.org.uk/>

Anders Sandberg

Sandberg, A., (2007) *Academy of Medical Sciences*, ‘Drugsfutures’. Public consultation about cognitive enhancement. <http://www.drugsfutures.org.uk/>

“I had the chance to participate in a public consultation on cognition enhancers in Glasgow, part of the project Drugsfutures, organized by the Academy of Medical Sciences. Members of the public were called upon to tell the organizers their views on enhancement, react to future scenarios and formulate what they thought were the best policy approaches. My role was to be an expert, providing information when needed.”

Teaching and seminars

2.00 – 4.00, Tuesdays Weeks 1-8 Trinity Term 2007, Faculty of Philosophy James Martin Seminar Series, academics, undergraduate and graduate students

2.00 pm-4.00 pm, Fridays Weeks 1-8 Hilary Term 2007, Faculty of Philosophy James Martin Seminar Series, academics, undergraduate and graduate students

Special lecture, Tuesday May 1 2007

Presenter: Michael Boylan (Marymount University, Virginia)

Topic: Worldview and the Value-Duty Link to Environmental Ethics

Special lecture, Friday May 11 2007-03-26

Presenter: Roland Benedikter (University of Vienna and University of Innsbruck, Austria)

Topic: Global Systemic Shift

Special lecture, Friday May 25 2007

Presenter: Ralph Merkle (Georgia Tech. College of Computing, Atlanta, Georgia)

Topic: Nanotechnology: the coming revolution in manufacturing

Rafaela Hillerbrand

Taught a second-year undergraduate for a one-to-one tutorial on the ethics of technology within the Stanford-in-Oxford program in the academic year 2007

Supervised Andrew Meyer, a second-year student from Stanford University, worked as a research assistant at the FHI this summer

Rebecca Roache

Taught a third-year undergraduate for a course of one-to-one tutorials on bioethics

Read extensive sections of theses for two D.Phil. students and provided detailed feedback

Presented papers at three James Martin Advanced Research Seminars

Collaboration

The Future of Humanity Institute and its sister project The Program on the Ethics of the New Biosciences hosted a workshop in October 2006 to initiate new collaborations and to celebrate their first few months working on the most important issues that we face. Invited participants worked together brainstorming fruitful new areas of research.

Nick Bostrom

Nick has collaborated with MIT physicist Max Tegmark to develop a new way to apply observation selection theory to derive an upper bound on the probability of a certain category of existential disasters. This work resulted in a co-authored paper published in *Nature*.

Nick is collaborating with Belgrade astrophysicist Dr. Milan Cirkovic on a co-edited volume on Global Catastrophic Risks, which will be published by Oxford University Press.

Contributor to Overcoming Bias Blog: with Professor Robin Hanson, George Mason University, U.S.A.

Contributor to Ethics Etc Blog: with Dr Matthew Liao

Guy Kahane

fMRI experiment on moral judgment: Dr Nicholas Shackel, FHI; Dr Katja Wiech, Dept. of Physiology Anatomy and Genetics; Dr Miguel Farias, Ian Ramsey Centre, [Oxford University](#), and the Psychology and Religion Research Group, Cambridge University.

fMRI experiment on belief and pain: Professor John Brooke, Andreas Idreos Professor of Science and Religion; Dr Nicholas Shackel, FHI; Dr Katja Wiech, Dept. of Physiology Anatomy and Genetics; Dr Miguel Farias, Ian Ramsey Centre, [Oxford University](#), and the Psychology and Religion Research Group, Cambridge University.

Kahane, G., Wiech, K., Farias, M., Shackel, N., and Tracey I., "Neuroimaging of Religious Analgesia" (under review, *Science*). This interdisciplinary study is the first of its kind to demonstrate the phenomenon of religious analgesia in a controlled laboratory setting and the first to use fMRI brain imaging to identify the neural pathways that underlie it. This study is the result of collaboration between Drs Kahane and Shackel and researchers from the Pain Imaging Neuroscience Group (Department of Physiology, Anatomy & Genetics), the Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (Department of Clinical Neurology), the Ian Ramsey Centre (Theology Faculty), and the Oxford Centre for the Science of Mind.

Contributor to Overcoming Bias Blog: with Professor Robin Hanson, George Mason University, U.S.A..

Contributor to Ethics Etc Blog: with Dr Matthew Liao

Rafaela Hillerbrand

Planning interdisciplinary seminar series of risk and risk assessment with Peter Taylor

Writing a research paper with Nicholas Shackel

Rebecca Roache

Assisting Matthew Liao, from the Program on Ethics of the New Biosciences, to organise an international conference on 'Human Nature and Bioethics', at City University, Hong Kong, in December 2007. This included making an application for funding to the British Academy, which was successful.

Co-authored a paper, 'Bioliberalism, Bioconservatism, and the Wisdom of Reflecting on Repugnance' (forthcoming in *Journal of Applied Philosophy* and, in German translation, in a volume on the ethics of enhancement edited by Nikolaus Knoepffler and published by Verlag Karl Alber) with Steve Clarke, from the Program on Ethics of the New Biosciences.

Contributor to Ethics Etc blog

Nicholas Shackel

Research papers written or being written with:

- Rafaela Hillerbrand, Future of Humanity Institute
- Dr Matthew Liao, Programme on the Ethics of the New Biosciences
- Dr Jerome Ravetz, James Martin Institute for Science and Civilization
- Professor Michael Clark, University of Nottingham

Kahane, G., Wiech, K., Farias, M., Shackel, N., and Tracey I., "Neuroimaging of Religious Analgesia" (under review, *Science*) This interdisciplinary study is the first of its kind to demonstrate the phenomenon of religious analgesia in a controlled laboratory setting and the first to use fMRI brain imaging to identify the neural pathways that underlie it. This study is the result of collaboration between Drs Kahane and Shackel and researchers from the Pain Imaging Neuroscience Group (Department of Physiology, Anatomy & Genetics), the Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (Department of Clinical Neurology), the Ian Ramsey Centre (Theology Faculty), and the Oxford Centre for the Science of Mind.

fMRI experiment on moral judgment: Dr Guy Kahane, Oxford Uehiro Centre for Practical Ethics; Dr Katja Wiech, Dept. of Physiology Anatomy and Genetics; Dr Miguel

Farias, Ian Ramsey Centre, [Oxford University](#), and the Psychology and Religion Research Group, Cambridge University.

fMRI experiment on belief and pain: Professor John Brooke, Andreas Idreos Professor of Science and Religion; Dr Guy Kahane, Oxford Uehiro Centre for Practical Ethics; Dr Katja Wiech, Dept. of Physiology Anatomy and Genetics; Dr Miguel Farias, Ian Ramsey Centre, [Oxford University](#), and the Psychology and Religion Research Group, Cambridge University.

Contributor to Overcoming Bias Blog: with Professor Robin Hanson, George Mason University, U.S.A.

Contributor to Ethics Etc Blog: with Dr Matthew Liao

Peter Taylor

Involved with the Environmental Change Institute in the Commodifying Carbon workshop

Planning interdisciplinary seminar series of risk and risk assessment with Rafaela Hillerbrand

Honours and Awards

Applied Ethics at Oxford ranked ‘in the highest group in the world’

The [Philosophical Gourmet Report](http://www.philosophicalgourmet.com/breakdown/breakdown12.asp), the most important ranking of Graduate Programs in Philosophy in the English speaking world, has just published their 2006 rankings. Applied Ethics at Oxford University appears in the highest group, Group 1, with median and mean scores of (4, 4). This is a tremendous achievement for Applied Ethics at Oxford. Our applied ethics program was only established in Oxford in 2003 on a modest budget. Philosophy overall at Oxford is ranked equal second in the world.

<http://www.philosophicalgourmet.com/breakdown/breakdown12.asp>

Nick Bostrom

Nick Bostrom is to deliver *The 11th Annual JUS Lecture* at the University of Toronto, 2007. (This lecture is delivered by "an internationally recognized major contributor to the advancement of genetics, neuroscience, psychiatry and its ethical implications". Previous speakers have included James D. Watson, Jean-Pierre Changeux, Anne Young, and Floyd Bloom).

“My Challenges for the next 15 years” Featured speaker at the *Second Annual Global Creative Leadership Summit* – “a unique platform for the best minds of our generation”, organized by LTB Foundation with support from the UN Fund for International Partnerships (UNFIP) (New York City, 2007, 23-25 September).

Nick Bostrom has been invited to contribute an essay for a new book project, "The New Stars Of Science" (working title), rights to which have been sold to Vintage Books for publication in 2008. Foreign rights (to date) have been acquired by S. Fisher Verlag (Germany), RBA Libros (Spain), and Het Spektrum (The Netherlands), edited by Max Brockman, who says "To come up with an invitation list appropriate to the goals of the book I asked a number of leading third culture scientists/authors to recommend the top young people in their respective fields (i.e. "who, among the young generation of scientists is most likely to turn out to be the next James Watson/Stephen Jay Gould/Martin Rees?".

Palgrave MacMillan Publishing is preparing a series of edited books of philosophy-related work, intended as “a showcase for original work from the best of the new generation of philosophers”, Dr. Bostrom has been invited to contribute to this series – not one, but three chapters.

Nick Bostrom has been invited as a featured presenter at the *Second Annual Global Creative Leadership Summit* – “a unique platform for the best minds of our generation” – organized by LTB Foundation with support from the UN Fund for International Partnerships (UNFIP) (New York City, 23-25 September).

Nominated for the *2007 Philip Leverhulme Prize* (awaiting outcome)

Marquis Who's Who in the World (24th Edition, 2007)

Dictionary of International Biography (34th Edition, 2007)

Fellow (University of St Gallen & World Demographic Association)

The Symbolic Systems Distinguished Speaker of 2006, Stanford University. ("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 been Daniel Kahneman, Michael Gazzaniga, Daniel Dennett, John Searle, Steven Pinker, and others).

Rafaela Hillerbrand

Rafaela defended her PhD in Theoretical Physics with distinction (summa cum laude) in Münster, Germany

Guy Kahane

Elected Fulford Junior Research Fellow, Somerville College, Oxford (2007-2009)

Elected Research Member, *Exeter College*, Oxford (2007)

Fundacao Bial Research Grant (awarded, with Drs N. Shackel and K. Wiech (Oxford), a major research grant for an fMRI study of moral judgement)

John Fell OUP Research Fund (awarded, with Professor Savulescu, a major research grant for a two-year project on neuroethics)

Merit Award, Oxford University, Humanities Division

Funding Procurement

Nick Bostrom

Philanthropist #1: £101,714

Philanthropist #2: \$25,000

Philanthropist #3: \$11,000

Developing proposal to the Greek Ministry of National Education for the creation of a virtual centre called the Oxford Epictetus Center for the Promotion of Mental Health and the Study of Human Potential

Developing proposal for Templeton Foundation: 'Program for Wisdom in the 21st Century'

Developing proposal to the Wellcome Trust for a program focusing on different concepts of risk

Guy Kahane

Co-investigator for *Fundacao Bial Research Grant* (PI, Dr N. Shackel, CI, Dr K. Wiech (Oxford)), a major research grant for an fMRI study of moral judgement)

Kahane, G., *John Fell OUP Research Fund* (awarded, with Professor Savulescu, a major research grant for a two-year project on neuroethics)

Wrote, with Nicholas Shackel, application for AHRC Research Grant, £700,000, *Normative Judgement in the Light of Cognitive Psychology and Neuroscience*. Rated A+ by the board, which means rated as of the highest quality and significance, but not funded due to pressure of funds. However, being graded A+ allows us to resubmit the very same application, which we will do.

Rafaela Hillerbrand

Developing proposal to the Wellcome Trust for a program focusing on different concepts of risk

Rebecca Roache

With Dr Matthew Liao, wrote an application to the British Academy for funding for 'Bioethics and Human Nature' conference. The application was successful.

Helped to develop funding proposal for Epictetus project (discussions ongoing)

Nicholas Shackel

Principal Investigator for the *Fundacao Bial Research Grant* (CIs, Drs G. Kahane and K. Wiech (Oxford)) for fMRI experiment on moral judgement: Value of award: Euros 43000

Wrote, with Guy Kahane, application for AHRC Research Grant, £700,000, *Normative Judgement in the Light of Cognitive Psychology and Neuroscience*. Rated A+ by the board, which means rated as of the highest quality and significance, but not funded due to pressure of funds. However, being graded A+ allows us to resubmit the very same application, which we will do.

Developing proposal for Templeton Foundation: ‘Program for Wisdom in the 21st Century’

Peter Taylor

Lighthill Risk Network

The Lighthill Risk Network (LRN) by Peter Taylor is a newly established not-for-profit initiative with the aim of bringing world-wide scientific expertise in various aspects of risk to the financial services sector and (re)insurance industry in particular.

The network provides business with a gateway to the latest in knowledge and understanding of risk, while acting as a focal point for the research community to engage with industry.

The expert panels will include Climate Change Implications, together with the Met Office (UK), Catastrophe Loss Modelling with the International Society of Catastrophe Managers (a US-based insurance organisation), Space Weather (partner TBA), Quantitative Techniques in Insurance (partner CASS Business School), and also Emerging Risks where the FHI will act as a partner.

Annexe 3

Workshops and Forums

FHI International Methodology Workshop	2
Public outreach project	5
ENHANCE Workshops	5
Whole Brain Emulation Workshop	5
Bayesian Approaches to Agreement Conference	6
Conference ‘EE250: The Euler Equations, 250 years on’	8
Everett@50 Conference	8
Existential Risk Workshop	17
Cross-Disciplinary Seminar on Risks	17
Human Nature	18
Conference on ‘Global catastrophic risks’	18
Forthcoming	18

FHI International Methodology Workshop

March 13th 2006

On 13th March 2006, in advance of the James Martin Institute inaugural 2006 World Forum, the FHI held a 'Big Issues for Humanity' advanced methodology workshop. Speakers included Joel Garreau (Washington Post), Julian Savulescu (University of Oxford), James Hughes (Trinity College, Connecticut), William Bainbridge (National Science Foundation), and Nick Bostrom (FHI).

Information Release for Big Issues for Humanity: Methodology Workshop

In conjunction to the World Forum, organized by the James Martin Institute, the Future of Humanity Institute will be holding an advanced Methodology Workshop. This will take place on Monday, 13 March 2006, i.e. the day before the start of the Forum, enabling you to kill two birds with one stone.

Speakers at the Forum, and a few other selected guests, will be invited to participate in the Workshop. We envisage this to be a very small, informal event, allowing for in-depth discussion in a group of distinguished minds.

The purpose of the workshop is to discuss and explore ideas, rather than to showcase completed work to a large audience. Papers will be circulated ahead of time to maximize time for thought and discussion.

Proposals for papers are welcome. Broadly, the Workshop will focus on methodological tools and difficulties and meta-level issues that arise when thinking about the kinds of topic that will be addressed at the Forum. The exact topics that will be covered will depend on the interests of the participants and presenters, but the following list illustrates some of the possibilities:

- Information markets as an institutional mechanism for aggregating information to yield probabilistic forecasts of future events or scientific hypothesis
- Observation selection theory – how to avoid anthropic bias when considering questions where observation selection effects filter our evidence, e.g. the Fermi paradox, the Doomsday argument, the Simulation argument etc.
- Disagreement and rationality. Can rational, truth-seeking Bayesians agree to disagree about factual questions? If not, what accounts for the pervasive disagreements we find among actual humans?
- Heuristics and biases. A rich literature has been developed in the last couple of decades on biases and heuristics that affect human cognition and decision-making. How do these findings relate to the prospects of human transformation and other issues arising from anticipated future technologies.
- Applied ethics and ELSI. Big government-sponsored techno-scientific projects like the human genome project and the National Nanotechnology Initiative now include

substantial funding for applied ethics and other “ELSI” research. Does such research produce useful results? What impact does it have over the way technology is developed and used?

- Scenario planning – is this a useful framework for thinking about future possibilities?
- Technological determinism – to what extent is technological determinism true, and how does this influence where people of “good will” should focus their efforts?

The Future of Humanity Institute World Forum Advanced Methodology Workshop (13 March 2006) Faculty of Philosophy, Oxford	
Programme	
10.30 am	Nick Bostrom Welcome and Introduction
10.40 am	Julian Savulescu Methods in Applied Ethics
11.40 am	Coffee break
12.00	Nick Bostrom Observation Selection Theory as a Methodology for Thinking about the Big Picture
1.00 pm	Lunch
2.30 pm	James Hughes Ensuring Universal Access to Safe Human Enhancement Technologies Response: James Tansey
3.30 pm	Coffee break
4.00 pm	William Bainbridge Advances in cognitive science and related fields are challenging traditional notions of human nature in profound ways that this paper will outline. As we understand ourselves better, painful questions arise: Are we less intelligent than we had imagined? Are we sufficiently noble to build a sustainable civilization? Is morality anything more than a rhetoric to bind a society together and justify its bloody conflict with other societies? Is the fertility collapse in advanced post-industrial nations a sign of impotence, implying that only barbarism is viable in the long run? Or, could awareness of the real characteristics of human

	nature be the first necessary step toward transforming humans into a new species, both philosophical and fertile, fulfilling the human creative potential?
5.00 pm	End

Participants	
<i>Name</i>	<i>University</i>
James Hughes	Trinity College, Connecticut
Kevin Warwick	University of Reading
William Bainbridge	US National Science Foundation
Joel Garreau	Washington Post
Nick Bostrom	FHI
Julian Savulescu	BEP
Bill Sharpe	
Guy Kahane	BEP
Connal Mannion	
Justin Holme	Cambridge Student
Rebecca Roache	FHI
Peter Ward	Stage research/ JMI/ World Forum
Angela Wilkinson	SBS
Nick Shackel	FHI
Peter Houghton	

Public outreach project

Glasgow (30 March 2007)

Rebecca Roache gave a presentation and led a discussion on ethical aspects of cognitive enhancement with members of the public for the Academy of Medical Sciences ‘Drugs Futures’ project (<http://www.drugsfutures.org.uk>), which explored public views on the sort of drug culture we want for the future. 30 March 2007.

ENHANCE Workshops

Oxford (4 May 2006) and Stockholm (27-28 March 2007)

Anders Sandberg helped organise the ENHANCE Workshops on cognition enhancement.

Whole Brain Emulation Workshop

St Hilda’s College, Oxford (26-27 May 2007)

Rebecca Roache, Nick Bostrom and Anders Sandberg organised an FHI workshop ‘Whole Brain Emulation’, dedicated to estimating when and how an emulation of a whole human brain might be possible. Attended by an international panel of neuroscientists and relevant researchers, and followed by the ongoing development of a ‘roadmap’ document.

Workshop Participants

<i>Name</i>	<i>Institution</i>
John Fiala	Research Assistant, Professor of Health Sciences, Boston University
Kenneth Hayworth	Research Fellow, University of Southern California
Todd Huffman	Research Assistant, Alcor
Randal Koene	Postdoctoral Research Fellow, Boston University
Eugen Leitl	Independent Researcher

Bruce McCormick	Professor Emeritus, Texas A&M University
Ralph Merkle	Professor, Georgia Tech College of Computing
Peter Passaro	Research Officer in Informatics, University of Sussex
Robin Hanson	Associate Professor of Economics, George Mason University
Rebecca Roache	Research Fellow, Future of Humanity Institute, University of Oxford
Nick Bostrom	Director, Future of Humanity Institute, University of Oxford
Anders Sandberg	Research Associate, Future of Humanity Institute, University of Oxford
Toby Ord	Research Associate, Future of Humanity Institute, University of Oxford
Non-attending participant: Robert Freitas (Senior Research Fellow, Institute for Molecular Manufacturing)	

Bayesian Approaches to Agreement Conference

Pembroke College, Oxford (4 June 2007)

<http://www.fhi.ox.ac.uk/bayesian.htm>

Nick Shackel organized an international conference for those interested in Bayesian approaches to agreement and disagreement. Richard Bradley of L.S.E presented his paper comparing deliberation and aggregation as ways of dealing with disagreement and producing collective opinion. Robin Hanson of George Mason University presented his recent research on disagreement, in which he has been extending Aumann's theorem to conclude that rational disagreement requires origin disputes. Christian List and Franz Dietrich of L.S.E. presented the paper they have written with Richard Bradley "Aggregating Causal Judgments". The occasion was a very interesting exposure of the Bayesian approach to the issue and provoked considerable discussion among the participants. We are looking at publishing the papers as special edition of a journal: possibly the Knowledge, Rationality and Action section of Synthese.

Workshop Programme

**Future of Humanity Institute
Bayesian Approaches to Agreement Workshop
Pembroke College, Oxford
4 June 2007**

Itinerary

1.00 – 2.00 pm	Sandwich lunch (the Gallery)
2.00 – 3.15 pm	Christian List and Franz Dietrich: Aggregating Casual Judgments (List, Franz and Bradley)
3.15 – 3.30 pm	Tea and coffee
3.30 – 4.45 pm	Robin Hanson: Recent research on common knowledge and disagreement.
4.45 – 6.00 pm	Richard Bradley: Deliberation and aggregation as ways of dealing with disagreement
6.00 pm	Close

Participants

<i>Name</i>	<i>Institution</i>
Dr Nick Bostrom	Director, Future of Humanity Institute, University of Oxford
Professor Robert Stalnaker	Massachusetts Institute of Technology
Professor Robin Hanson	Associate Professor of Economics, George Mason University
Dr Rebecca Roache	Research Fellow, Future of Humanity Institute, University of Oxford
Professor Richard Bradley	London School of Economics
Dr Franz Dietrich	London School of Economics
Dr Christian List	London School of Economics
Professor Wlodek Rabinowicz	Lund University

Dr Ralph Wedgwood	Merton College, University of Oxford
Mr Michael Blome-Tillmann	Stevenson Junior Research Fellow, University College
Professor Luciano Floridi	St Cross College, University of Oxford
Dr Nick Shackel	Research Fellow, Future of Humanity Institute, University of Oxford
Dr Peter Taylor	Research Associate, Future of Humanity Institute, University of Oxford
Dr Matthew Liao	Research Fellow, Ethics of the New Biosciences, University of Oxford
Mr Sebastian Sequoiah-Grayson	Balliol College, Oxford
Mr Matteo Turilli	Lady Margaret Hall, Oxford

Conference ‘EE250: The Euler Equations, 250 years on’

Rafaela Hillerbrand assisted in organizing a conference held on the tercentenary of the birth of Leonhard Euler and the 250th anniversary of his seminal publications on fluid mechanics. The conference will cover the latest research within fluid mechanics related to the modelling and predictability of nonlinear systems. The conference was held in Aussois, France in June 2007, and was organized by Uriel Frisch (Observatoire de la Côte d’Azur, Nice) and funded by CNRS and under the patronage of the French Academy of Science.

<http://www.obs-nice.fr/etc7/EE250/> and <http://groups.google.com/group/ee250>

Everett@50 Conference

Peter Taylor was Conference Manager for the Everett@50 Conference in Oxford held between 19th through 21st July 2007 to bring together the leading philosophers and physicists interested in the interpretation of quantum mechanics to discuss Everett's theory on the 50th Anniversary of publication of the "relative state formulation of quantum mechanics", sometimes called the many-worlds or multiverse interpretation. Oxford philosophers have led the revival of the Everett interpretation in the past ten years, and this Conference will see if Everett's explanation of quantum mechanics has at last come of age. Aspects of the conference from an administrative point of view include a live webcast, will have an on-line Blog, and is being included in a BBC4 documentary on Everett to be broadcast in November 2007.

Website: <http://users.ox.ac.uk/~everett/>

Conference Programme

Thursday 19th July

10.30 - 11.00 Coffee

11.00 - 1.00 The Everett interpretation: 50 years on. Simon Saunders

How to think about ontology. David Wallace

Commentator: Robert Geroch

1.00 - 2.00 Sandwich lunch, Ryle Room, 10 Merton St.

2.00 - 3.30 Can the world be only wave-function? Tim Maudlin

Commentator: Adrian Kent

3.30 - 4.00 Tea

4.00 - 5.30 Two Dogmas About Quantum Mechanics. Jeff Bub and Itamar Pitowsky

Commentator: Chris Timpson

6.30 - 10.00 Evening drinks and dinner, Cherwell boathouse

Friday 20th July

9.30 - 11.00 Everett and Evidence. Wayne Myrvold and Hilary Greaves

Commentator: Barry Loewer

11.00 - 11.30 Coffee

11.30 - 1.00 Probability in the Everett picture. David Z. Albert

Commentator: David Papineau

1.00 - 3.00 Lunch at the Head of the River

- 3.00 - 4.30 Apart from universes. David Deutsch
 The time symmetric QM and the MWI. Lev Vaidman
- 4.30 - 5.00 Tea
- 5.00 - 6.30 Quantum cosmology. James B. Hartle
 Probability without time. Andreas J. Albrecht
- 7.00 - 10.00 Conference dinner, Oriel College

Saturday 21st July

- 9.30 - 11.00 A metaphysician looks at the Everett interpretation. John Hawthorne
 Commentator: James Ladyman
- 11.00 - 11.30 Coffee
- 11.30 - 1.00 Explaining probability. Simon Saunders
 Commentator: Oliver Pooley
- 1.00 - 2.30 Lunch, local restaurants
- 2.30 - 4.00 Pilot-wave theory: Everett in denial? Antony Valentini
 Commentator: Harvey Brown
- 4.00 - 4.30 Tea
- 4.30 - 6.00 Round table discussion: David Wallace, Jeremy Butterfield,
 David Albert
- 7.00 Drinks, Old Cloisters, New College
- Farewell dinner, The Undercroft, New College

Speakers

David Z. Albert

David Z. Albert is the Frederick E. Woodbridge Professor of Philosophy at Columbia University and the Director of the M.A. Program in the Philosophical Foundations of Physics. His areas of specialisation are the philosophical problems of modern physics, philosophy of quantum mechanics, philosophy of time and space, and the philosophy of science. He is the author of *Quantum Mechanics and Experience*, and *Time and Chance*.

Andreas J. Albrecht

Andreas J. Albrecht is a Professor in the Department of Physics at the University of California, Davis, and a Fellow of the American Physical Society and the Institute of Physics (UK). His focus is on the field of Cosmology and issues surrounding the

formation and evolution of the Universe. His specific research problems currently include: fundamental issues with the theory of cosmic inflation, the formation of cosmic structure, and searching for the understanding of the "dark energy" that currently suggests is accelerating the Universe. This recent work includes serving on the "Dark Energy Task Force" to develop the US observational program to study the cosmic acceleration.

Jeffrey Bub

Jeffrey Bub is a Distinguished Professor in the Department of Philosophy at the University of Maryland and on the Committee for Philosophy and the Sciences. Rooted in the foundations of physics, his current interests are in the rapidly developing fields of quantum computation, quantum cryptography, and especially quantum information: 'how information is stored, how it can be moved around, what you can do about it, and what this tells us about the quantum world.' He is the author of *The Interpretation of Quantum Mechanics and Interpreting the Quantum World*, which won the Lakatos Award in 1998.

David Deutsch

David Deutsch is an associate of the Department of Atomic and Laser physics at the Centre for Quantum Computation, University of Oxford. He is one of the founders of the field of quantum computing and a long-standing proponent of the multiverse interpretation of quantum mechanics, as set out in his book *The Fabric of Reality*. He was the recipient of the Dirac Prize of the Institute of Physics in 1998 and the Edge of Computation Science Prize in 2005, and is currently working on a book entitled *The Beginning of Infinity*.

Hilary Greaves

Hilary Greaves is completing her Ph.D in the Department of Philosophy at Rutgers University and is Junior Research Fellow at Merton College, University of Oxford, in philosophy of physics. She is the author of a number of papers in confirmation theory and Bayesian epistemology, most of them focused explicitly on the Everett interpretation. Her most recent is "On the Everettian problem", published in *History and Philosophy in Modern Physics*.

James B. Hartle

James B. Hartle is a Research Professor of Physics at the University of California, Santa Barbara. His scientific work is concerned with the application of Einstein's theory of gravity to realistic astrophysical situations, especially cosmology. He has contributed usefully to the understanding of gravitational waves, relativistic stars, and black holes. His current interest is in understanding the quantum origin of the universe and the generalizations of quantum mechanics necessary for that. He is a member of the US National Academy of Sciences, a fellow of the American Academy of Arts and Sciences, and a founder and past director of the Institute for Theoretical Physics in Santa Barbara.

John Hawthorne

John Hawthorne is the Waynflete Professor of Metaphysical Philosophy at Magdalen College, University of Oxford. His research interests include metaphysics, epistemology, philosophy of language, philosophy of mind, and early modern philosophy. His most

recent book is *Metaphysical Essays*, 2006. Visit his website at <http://www.philosophy.ox.ac.uk/members/jhawthorne/index.htm>

Tim Maudlin

Tim Maudlin is Professor in the Department of Philosophy at Rutgers University. His areas of research include the philosophy of science, philosophy of physics, and metaphysics. He is the author of *Quantum Non-Locality and Relativity: Metaphysical Intimations of Modern Physics* and *Truth and Paradox: Solving the Riddles*.

Wayne C. Myrvold

Wayne C. Myrvold is Associate Professor in the Department of Philosophy at Talbot College, University of Western Ontario, and is Associate Member of the Perimeter Institute of Theoretical Physics, Waterloo. His work is chiefly concerned with the philosophy of physics and the interpretation of quantum mechanics, but he has also published in confirmation theory, Bayesian epistemology, and the philosophy of biology. His recent publications include “Modal Interpretations and Relativity” and “Relativistic Quantum Becoming.”

Itamar Pitowsky

Itamar Pitowsky is a Professor in the Philosophy Department and The Program for the History, Philosophy and Sociology of Science at The Hebrew University of Jerusalem. His research is in philosophy of physics. He has contributed extensively to the foundations of quantum mechanics. In his monograph, *Quantum Probability, Quantum Logic*, he recast the Bell inequalities as general theorems about classical probability. He is currently working in information-theoretic approaches to quantum mechanics. Visit his website at <http://edelstein.huji.ac.il/staff/pitowsky/>.

Simon Saunders

Simon Saunders is Reader in the Philosophy of Physics and Fellow of Linacre College at the University of Oxford. He has worked in the foundations of quantum field theory, quantum mechanics, symmetries, and thermodynamics and statistical mechanics. He was an early proponent of the view of branching in the Everett interpretation as an ‘effective’ process based on decoherence. His most recent work include ‘On the explanation of quantum statistics’ and (with D. Wallace) ‘Branching and uncertainty’. Visit his website at <http://users.ox.ac.uk/~lina0174/Saunders.html>.

Lev Vaidman

Lev Vaidman is a professor of physics at Tel-Aviv University. His scientific interests are Foundations of Quantum Mechanics and Quantum Information. His main achievements include Continuous-Variables Teleportation, Weak Measurements (with Yakir Aharonov), Interaction-free Measurements (with Avshalom Elitzur), and Cryptography with Orthogonal States (with Lior Goldenberg). For a long time he is one of the strongest proponents of the Everett Interpretation as can be seen from his SEP entry *The Many-Worlds Interpretation*.

Antony Valentini

Antony Valentini is a Visiting Professor at the Centre de Physique Théorique de Luminy, Université de la Méditerranée, Marseilles, and a member of the Foundational Questions Institute. His research focuses on the possible role of hidden variables in quantum theory and cosmology --- in particular, in the very early universe (including inflationary cosmology), in quantum information and computation, and in the physics of black holes. His research interests also include the history and philosophy of modern physics: he is co-author (with G. Bacciagaluppi) of *Quantum Theory at the Crossroads: Reconsidering the 1927 Solvay Conference* (CUP, forthcoming, and quant-ph/0609184). He is the principal exponent of the 'quantum nonequilibrium' hypothesis, according to which quantum theory is not fundamental but merely describes a statistical equilibrium state, which the universe happens to be in at the present time. Recent papers include 'Astrophysical and Cosmological Tests of Quantum Theory', *J. Phys. A: Math. Theor.* 40, 3285-3303 (2007) [hep-th/0610032]. He is completing a book (also to be published by CUP) that re-examines modern physics and cosmology from a pilot-wave and general hidden-variables viewpoint.

David Wallace

David Wallace is Fellow in Philosophy of Balliol College, University of Oxford. His research has concentrated on the interpretation of quantum mechanics, and in particular on the Everett interpretation of quantum mechanics, although he has also worked on the foundations of statistical mechanics and of quantum field theory. He is co-author (with D. Deutsch) of the decision-theory argument for quantum probability and of a number of influential papers on the Everett interpretation.

Commentators

Guido Bacciagaluppi

Guido Bacciagaluppi is a philosopher of physics at the Centre for Time, University of Sydney. He works on the foundations of quantum mechanics and is a principal contributor to modal interpretation of quantum mechanics. He has just completed (with A. Valentini) an English translation of and commentary on the *Proceedings of the Fifth Solvay Congress of 1927*.

Harvey Brown

Harvey Brown is Professor of Philosophy at the University of Oxford and Fellow of Wolfson College. He has published widely in the foundations of quantum mechanics, relativity theory, and thermal physics. He is the author of *Physical Relativity: Space-time structure from a dynamical perspective*, for which he was co-winner of the 2006 Lakatos prize in philosophy of science.

Jeremy Butterfield

Jeremy Butterfield is a Senior Research Fellow at Trinity College, University of Cambridge. He has published widely in the philosophy of space-time, and in the foundations of quantum mechanics and classical mechanics. His most recent book (co-edited with J. Earman) is *A Handbook of Philosophy of Physics*.

Robert Geroch

Robert Geroch is Professor of Physics at the University of Chicago. His research interests lie in relativity and quantum mechanics. He is the author of *General Relativity From A to B*.

Meir Hemmo

Meir Hemmo is Professor in the Philosophy Department at the University of Haifa. His main research areas are philosophy of modern physics, philosophy of science, probability and metaphysics.

Michel Janssen

Michel Janssen is a Professor at the Center for Philosophy of Science at the University of Minnesota. He is a regular visitor at the Max Planck Institute for the History of Science in Berlin. His research area is the history of modern physics, particularly the history of relativity theory.

Adrian Kent

Adrian Kent is Professor of Physics at the Centre for Quantum Computation, University of Cambridge. His research interests are in quantum information theory, quantum cryptography, and foundations of quantum theory. He is the author of a number of critical articles on the consistent-histories approach to quantum mechanics.

James Ladyman

James Ladyman is Reader in Philosophy at the University of Bristol. His research interests are primarily in philosophy of science, and especially in constructive empiricism and structural realism. He is the author of *Understanding Philosophy of Science*, which received a Choice Outstanding Academic Title Award, and has just completed (with D. Ross, D. Spurrett and J. Collier) *Everything Must Go: Metaphysics Naturalized*. He is the recipient of the Philip Leverhulme Prize in Philosophy and Ethics.

Christoph Lehner

Christoph Lehner is a Research Scholar at the Max Planck Institute for the History of Science in Berlin, and coordinator of the research project on history and foundations of quantum physics. He got his Ph.D. from Stanford with a dissertation about the Everett interpretation and has worked at the Einstein Papers project. Recently, he was one of the organizers of the 2005 exhibition "Albert Einstein, Engineer of the Universe." Right now, he is working on the history of wave mechanics and on a Cambridge Companion to Albert Einstein.

Peter Lewis

Peter Lewis is Associate Professor of Philosophy at the University of Miami. His research interests are in philosophy of science, especially philosophy of physics, scientific realism and scientific methodology. He has published articles on the foundations of quantum mechanics and on scientific realism.

Barry Loewer

Barry Loewer is a Professor and chair of the Department of Philosophy at Rutgers University and is Director of the Rutgers Center for Philosophy and the Sciences. His

areas of research include the philosophy of science, philosophy of physics, the philosophy of mind, metaphysics, and the philosophical logic. He has published articles on Bohemian mechanics, GRW, and on the foundations of Stat Mech (may with D. Albert) and is the author of *Meaning in Mind* (with George Rey Blackwell), and the creator (with D. Albert) of the ‘many-minds’ interpretation of quantum mechanics.

David Papineau

David Papineau is Professor of Philosophy of Science at King's College London. His areas of focus are epistemology, philosophy of science, and philosophy of mind. He is currently working on consciousness, practical reasoning, and the evolution of cognition more generally. His recent books include *Thinking about Consciousness* and *The Roots of Reason: Philosophical Essays on Rationality, Evolution and Probability*.

Oliver Pooley

Oliver Pooley is a Fellow in Philosophy of Oriel College, University of Oxford. His primary area of research is in the philosophy of physics, where he is especially interested in topics that overlap with metaphysics and the philosophy of language. He has written influential articles on Mach’s principle in special and general relativity and on relationist approaches to mirror-symmetry. He is currently completing a book on spacetime.

Alastair Rae

Alastair Rae is a Reader in Quantum Physics at the School of Physics and Astronomy at the University of Birmingham until he retired in 2003. He is the author of *Quantum Physics: Reality or Illusion?*, *Quantum Mechanics* (an undergraduate text) and *Quantum Physics: a Beginner’s Guide*.

Tony Sudbery

Tony Sudbery is a Professor in the Department of Mathematics, University of York. His areas of interest are quantum information theory, foundations of quantum mechanics, and exceptional Lie algebras. Recent articles include "Why Am I Me?" (quant-ph/00011084) and "Alice and Bob Get Away With It" (physics/0606186).

Paul Tappenden

Paul Tappenden obtained his Ph.D. from the Department of Philosophy, Kings College, London. His research interests lie in metaphysics, philosophy of mind, and the Everett interpretation of quantum mechanics. His most recent article is 'Saunders and Wallace on Everett and Lewis'.

Christopher Timpson

Christopher Timpson is currently a Lecturer in Philosophy at the University of Leeds. He will be joining Oxford faculty in September 2007. His research interests are in the philosophy of physics, especially quantum mechanics and quantum information theory, philosophy of science, and philosophy of mind and language. Recent publications include “The Grammar of Teleportation” and (with H. R. Brown) of “Why Special Relativity Should not be a Template for a Fundamental Reformulation of Quantum Mechanics”. His work on quantum information theory will shortly be forthcoming with an Oxford

University Press monograph Quantum Information Theory and the Foundations of Quantum Mechanics .

Tim Williamson

Tim Williamson is the Wykeham Professor of Logic at the University of Oxford, and Fellow of New College, Oxford. His main research interests are in philosophical logic, philosophy of language, epistemology and metaphysics. He is the author of *Vagueness, Knowledge and Its Limits* and the forthcoming *The Philosophy of Philosophy*. He was this year elected Foreign Honorary Member of the American Academy of Arts and Sciences.

Existential Risk Workshop

Nick Bostrom and Rafaela Hillerbrand are organizing workshop on Existential Risks to be held in Oxford in autumn 2007. Leading experts on different existential risks will be invited, such as:

Gaverick Matheny

John Leslie

Richard Posner

Elizier Yudkowsky

the person to be nominated with the Winton Professorship (Cambridge, UK)

Robin Hanson

Eric Drexler

Lou Sulkind

Jared Diamond

Cross-Disciplinary Seminar on Risks

Peter Taylor and Rafaela Hillerbrand are organizing a programme of cross-disciplinary seminars for the James Martin 21st Century School on risk, including emerging risks, in the Academic Year 2007/2008 Michaelmas and Hilary terms. This will cover such areas as the effect of new technologies as well as the socio-economic and political response to disruptive change.

Peter Taylor will be organising a Carbon Trading seminar in the City of London in the Autumn to follow-up the Commodifying Carbon Workshop, and it is intended to invite key people from the Environmental Change Institute.

Human Nature

Rebecca Roache is helping Matthew Liao from our sister JM institute, BEP, to organise 'Human Nature and Bioethics' conference at City University, Hong Kong, China, in December 2007. This included making a successful application for funding to the British Academy. To date, confirmed speakers at the conference are Jonathan Glover (King's College, London), Jeff McMahan (Rutgers), John Harris (Manchester), Dan Brock (Harvard), Dan Wikler (Harvard), Ingmar Persson (Göteborg), Jo Wolff (University College, London).

Conference on 'Global catastrophic risks'

With the launch of the Oxford University monograph on global catastrophic risk edited by Nick Bostrom and Milan Cirkovic, there will be a conference hosted by the FHI with all the contributing authors as well as invited speakers. Conference on 'Global catastrophic risks'. With the launch of the Oxford University monograph on global catastrophic risk edited by Nick Bostrom and Milan Cirkovic, there will be a conference hosted by the FHI with all the contributing authors as well as invited speakers.

Forthcoming

Rebecca Roache is currently working with the James Martin Institute of Ageing to identify possibilities for collaboration. One possibility is a jointly-organised workshop to predict the social impact of life-extension technology.

Annexe 5

Publication Specimen

“The Reversal Test” (Bostrom & Ord)

The Reversal Test: Eliminating Status Quo Bias in Applied Ethics*

Nick Bostrom and Toby Ord

I. INTRODUCTION

Suppose that we develop a medically safe and affordable means of enhancing human intelligence. For concreteness, we shall assume that the technology is genetic engineering (either somatic or germ line), although the argument we will present does not depend on the technological implementation. For simplicity, we shall speak of enhancing “intelligence” or “cognitive capacity,” but we do not presuppose that intelligence is best conceived of as a unitary attribute. Our considerations could be applied to specific cognitive abilities such as verbal fluency, memory, abstract reasoning, social intelligence, spatial cognition, numerical ability, or musical talent. It will emerge that the form of argument that we use can be applied much more generally to help assess other kinds of enhancement technologies as well as other kinds of reform. However, to give a detailed illustration of how the argument form works, we will focus on the prospect of cognitive enhancement.

Many ethical questions could be asked with regard to this prospect, but we shall address only one: do we have reason to believe that the long-term consequences of human cognitive enhancement would be, on balance, good? This may not be the only morally relevant question—

* For comments, we are grateful to Daniel Brock, David Calverley, Arthur Caplan, Jonathan Glover, Robin Hanson, Michael Sandel, Julian Savulescu, Peter Singer, Mark Walker, and to the participants of the “Methods in Applied Ethics” seminar at Oxford, the “How Can Human Nature Be Ethically Improved” conference in New York, the “Sport Medicine Ethics” conference in Stockholm, and the Oxford-Scandinavia Ethics Summit, where earlier versions of this article were presented. We are also grateful for the helpful comments from two anonymous referees and six anonymous members of the editorial board.

we leave open the possibility of deontological constraints—but it is certainly of great importance to any ethical decision making.¹

It is impossible to know what the long-term consequences of such an intervention would be. For simplicity, we may assume that the immediate biological effects are relatively well understood, so that the intervention can be regarded as medically safe. There would remain great uncertainty about the long-term direct and indirect consequences, including social, cultural, and political ramifications. Furthermore, even if (*per impossibile*) we knew what all the consequences would be, it might still be difficult to know whether they are on balance good. When assessing the consequences of cognitive enhancement, we thus face a double epistemic predicament: radical uncertainty about both prediction and evaluation.

This double predicament is not unique to cases involving cognitive enhancement or even human modification. It is part and parcel of the human condition. It arises in practically every important deliberation, in individual decision making as well as social policy. When we decide to marry or to back some major social reform, we are not—or at least we shouldn't be—under any illusion that there exists some scientifically rigorous method of determining the odds that the long-term consequences of our decision will be a net good. Human lives and social systems are simply too unpredictable for this to be possible. Nevertheless, some personal decisions and some social policies are wiser and better motivated than others. The simple point here is that our judgments about such matters are not based exclusively on hard evidence or rigorous statistical inference but rely also—crucially and unavoidably—on subjective, intuitive judgment.

The quality of such intuitive judgments depends partly on how well informed they are about the relevant facts. Yet other factors can also have a major influence. In particular, judgments can be impaired by various kinds of biases. Recognizing and removing a powerful bias will sometimes do more to improve our judgments than accumulating or analyzing a large body of particular facts. In this way, applied ethics could benefit from incorporating more empirical information from psychology and the social sciences about common human biases.

In this article we argue that one prevalent cognitive bias, status quo bias, may be responsible for much of the opposition to human en-

1. In parallel to affirming deontological side constraints, one might also hold that the value of a state of affairs depends on how that state was brought about. For instance, one might hold that the value of a state of affairs is reduced if it resulted from a decision that violated a deontological side constraint. When we discuss the consequentialist dimension of ethical or prudential decision making in this article, we mainly set aside this possibility.

hancement in general and to genetic cognitive enhancement in particular. Our strategy is as follows: first, we briefly review some of the psychological evidence for the pervasiveness of status quo bias in human decision making. This evidence provides some reason for suspecting that this bias may also be present in analyses of human enhancement ethics. We then propose two versions of a heuristic for reducing status quo bias. Applying this heuristic to consequentialist objections to genetic cognitive enhancements, we show that these objections are affected by status quo bias. When the bias is removed, the objections are revealed as extremely implausible. We conclude that the case for developing and using genetic cognitive enhancements is much stronger than commonly realized.

II. PSYCHOLOGICAL EVIDENCE OF STATUS QUO BIAS

That human thinking is susceptible to the influence of various biases has been known to reflective persons throughout the ages, but the scientific study of cognitive biases has made especially great strides in the past few decades.² We will focus on the family of phenomena referred to as status quo bias, which we define as an inappropriate (irrational) preference for an option because it preserves the status quo.

While we must refer the reader to the scientific literature for a comprehensive review of the evidence for the pervasiveness of status quo bias, a few examples will serve to illustrate the sorts of studies that have been taken to reveal this bias.³ These examples will also help delimit the particular kind of status quo bias that we are concerned with here.

The Mug Experiment.—Two groups of students were asked to fill out a short questionnaire. Immediately after completing the task, the students in one group were given decorated mugs as compensation, and the students in the other group were given large Swiss chocolate bars. All participants were then offered the choice to exchange the gift they had received for the other, by raising a card with the word “Trade” written on it. Approximately 90 percent of the participants retained the original reward.⁴

Since the two kinds of reward were assigned randomly, one would have expected that half the students would have got a different reward from the one they would have preferred *ex ante*. The fact that 90 percent of the participants preferred to retain the award they had been given

2. See, e.g., Thomas Gilovich, Dale W. Griffin, and Daniel Kahneman, *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge: Cambridge University Press, 2002).

3. For a good introduction to the literature on status quo bias and related phenomena, see Daniel Kahneman and Amos Tversky, *Choices, Values, and Frames* (Cambridge: Cambridge University Press, 2000).

4. Gilovich et al., *Heuristics and Biases*.

illustrates the “endowment effect,” which causes an item to be viewed as more desirable immediately upon its becoming part of one’s endowment.

The endowment effect may suggest a status quo bias. However, we have defined status quo bias as an inappropriate favoring of the status quo. One may speculate that the favoring of the status quo in the Mug Experiment results from the subjects forming an emotional attachment to their mug (or chocolate bar). An endowment effect of this kind may be a brute fact about human emotions and as such may be neither inappropriate nor in any sense irrational. The subjects may have responded rationally to an a-rational fact about their likings. There is thus an alternative explanation of the Mug Experiment which does not involve status quo bias.

In this article, we want to focus on genuine status quo bias that can be characterized as a cognitive error, where one option is incorrectly judged to be better than another because it represents the status quo. Moreover, since our concern is with ethics rather than prudence, our focus is on (consequentialist) ethical judgments. In this context, instances of status quo bias cannot be dismissed as merely apparent on grounds that the evaluator is psychologically predisposed to like the status quo, for the task of the evaluator is to make a sound ethical judgment, not simply to register his or her subjective likings. Of course, people’s emotional reactions to a choice may form part of the consequences of the choice and have to be taken into account in the ethical evaluation. Yet status quo bias remains a real threat. It is perfectly possible for a decision maker to be biased in judging the strength of people’s emotional reactions to a change in the status quo.⁵ Explanations in terms of emotional bonding seem less likely to account for the findings in the following two studies.

Hypothetical Choice Tasks.—Some subjects were given a hypothetical choice task in the following “neutral” version, in which no status quo was defined: “You are a serious reader of the financial pages but until recently you have had few funds to invest. That is when you inherited a large sum of money from your great-uncle. You are considering different portfolios. Your choices are to invest in: a moderate-risk company, a high-risk company, treasury bills, municipal bonds.” Other subjects were presented with the same problem but with one of the options designated as the status quo. In this case, the opening passage continued: “A significant portion of this portfolio is invested in a moderate risk company . . . (The tax and

5. Independent of the issue of status quo bias, there is evidence of a durability bias in affective forecasting, which leads people to systematically overestimate the duration of emotional reactions to future events; see, e.g., Gilovich et al., *Heuristics and Biases*, 292ff.

broker commission consequences of any changes are insignificant.)” The result was that an alternative became much more popular when it was designated as the status quo.⁶

Electric Power Consumers.—California electric power consumers were asked about their preferences regarding trade-offs between service reliability and rates. The respondents fell into two groups, one with much more reliable service than the other. Each group was asked to state a preference among six combinations of reliability and rates, with one of the combinations designated as the status quo. A strong bias to the status quo was observed. Of those in the high-reliability group, 60.2 percent chose the status quo, whereas a mere 5.7 percent chose the low-reliability option that the other group had been experiencing, despite its lower rates. Similarly, of those in the low-reliability group, 58.3 chose their low-reliability status quo, and only 5.8 chose the high-reliability option.⁷

It is hard to prove irrationality or bias, but taken as a whole, the evidence that has accumulated in many careful studies over the past several decades is certainly suggestive of widespread status quo bias. In considering the examples given here, it is important to bear in mind that they are extracted from a much larger body of evidence. It is easy to think of alternative explanations for the findings of these particular studies, but many of the potential confounding factors (such as transaction costs, thinking costs, and strategic behavior) have been ruled out by further experiments. Status quo bias plays a central role in prospect theory, an important recent development in descriptive economics (which earned one of its originators, Daniel Kahneman, a Nobel Prize in 2002).⁸ Psychologists and experimental economists have found extensive evidence for the prevalence of status quo bias in human decision making.⁹

6. William Samuelson and Richard Zeckhauser, “Status Quo Bias in Decision Making,” *Journal of Risk and Uncertainty* 1 (1988): 7–59.

7. Raymond S. Hartman, Michael J. Doane, and Chi-Keung Woo, “Consumer Rationality and the Status Quo,” *Quarterly Journal of Economics* 106 (1991): 141–62.

8. The work of Kahneman and Amos Tversky and their collaborators has convinced many economists that the standard economic paradigm, which postulates rational expected-utility maximizing agents, is, despite its simplicity and convenient formal features, not descriptively adequate for many situations of human decision making.

9. The exact nature and the psychological factors contributing to status quo bias are not yet fully understood. Loss aversion—the tendency to place a greater weight on aspects of outcomes when they are represented as “losses” (rather than, e.g., forfeited gains)—seems to be a significant part of the picture (James N. Druckman, “Evaluating Framing Effects,” *Journal of Economic Psychology* 22 [2001]: 91–101). It has also been suggested that omission bias may account for some of the findings previously ascribed to status quo bias. Omission bias is diagnosed when a decision maker prefers a harmful outcome that results from an omission to a less harmful outcome that results from an action (even in cases

Let us consider one more illustration of the empirical literature on status quo bias. One source of status quo bias is loss aversion, which can seduce people into judging the same set of alternatives differently depending on whether they are phrased in terms of potential losses or potential gains.

The Asian Disease Problem.—The same cover story was presented to all the subjects: “Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows.” One group of subjects was presented with the following pair alternatives (the percentage of respondents choosing a given program is given in parentheses):

If Program A is adopted, 200 people will be saved (72 percent).
If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved (28 percent).

Another group of subjects were instead offered the following alternatives:

If Program C is adopted, 400 people will die (22 percent).
If Program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die (78 percent).¹⁰

It is easy to verify that the options A and B are indistinguishable in real terms from options C and D, respectively. The difference is solely one of framing. In the first formulation, the outcomes are represented as gains (people are saved), while in the second formulation, outcomes are represented as losses (people die). The second formulation, however, assumes a reference state where nobody dies of the disease, and Program D is the only way to possibly avoid a loss. In the first formulation, by contrast, the assumed reference state is that nobody lives, and ordinary risk aversion explains why people prefer Program A (the safe bet).

The bias to avoid outcomes that are framed as “losses” is both pervasive and robust.¹¹ This has long been recognized by marketing

where presumably no moral deontological constraints are involved; Ilana Ritov and Jonathan Baron, “Status-Quo and Omission Biases,” *Journal of Risk and Uncertainty* 5 [1992]: 49–61).

10. Amos Tversky and Daniel Kahneman, “The Framing of Decisions and the Psychology of Choice,” *Science* 211, no. 4481 (1981): 453–58.

11. For a review of more recent confirmations of this framing effect, see, e.g., Druckman, “Evaluating Framing Effects.”

professionals. Credit card companies, for instance, lobbied vigorously to have the difference between a product's cash price and credit card price labeled "cash discount" (implying that the credit price is the reference point) rather than "credit card surcharge," presumably because consumers would be less willing to accept the "loss" of paying a surcharge than to forgo the "gain" of a discount.¹² The bias has been demonstrated among sophisticated respondents as well as among naive ones. For example, one study found that preferences of physicians and patients for surgery or radiation therapy for lung cancer varied markedly when their probable outcomes were described in terms of mortality or survival.¹³

Changes from the status quo will typically involve both gains and losses, with the change having good overall consequences if the gains outweigh these losses. A tendency to overemphasize the avoidance of losses will thus favor retaining the status quo, resulting in a status quo bias. Even though choosing the status quo may entail forfeiting certain positive consequences, when these are represented as forfeited "gains" they are psychologically given less weight than the "losses" that would be incurred if the status quo were changed.

Having noted that a body of data from psychology and experimental economics provides at least *prima facie* grounds for suspecting that a status quo bias may be endemic in human cognition, let us now turn to the case of human cognitive enhancement. Does status quo bias affect our judgments about such enhancements? If so, how can the bias be diagnosed and removed?

III. A HEURISTIC FOR REDUCING STATUS QUO BIAS

Many people judge that the consequences of increasing intelligence would be bad, even assuming that the method used would be medically safe. While enhancing intelligence would clearly have many potential benefits, both for individuals and for society, some feel that the outcome would be worse on balance than the status quo because increased intelligence might lead people to become bored more quickly, to become more competitive, or to be better at inventing destructive weapons; because social inequality would be aggravated if only some people had access to the enhancements; because parents might become less accepting of their children; because we might come to lose our "openness to the unbidden"; because the enhanced might oppress the rest; or

12. R. Thaler, "Toward a Positive Theory of Consumer Choice," *Journal of Human Behavior and Organization* 1 (1980): 39–60.

13. B. J. McNeil, S. G. Pauker, H. G. Sox, and A. Tversky, "On the Elicitation of Preferences for Alternative Therapies," *New England Journal of Medicine* 306 (1982): 1259–62.

because we might come to suffer from “existential dread.”¹⁴ These worries are often combined with skepticism about the potential upside of enhancement of cognitive and other human capacities:

Whether a general ‘improvement’ in height, strength, or intelligence would be a benefit at all is even more questionable. To the individual such improvements will benefit his or her social status, but only as long as the same improvements are not so widespread in society that most people share them, thereby again levelling the playing field. . . . What would be the status of Eton, Oxford and Cambridge if all could go there? . . . In general there seems to be no connection between intelligence and happiness, or intelligence and preference satisfaction. . . . Greater intelligence could, of course, also be a benefit if it led to a better world through more prudent decisions and useful inventions. For this suggestion there is little empirical evidence.¹⁵

In a recent article, another author opines: “Crucially, though, despite the fact that parents may want their children to be ‘intelligent’, where all parents want this any beneficial effect is nullified. On the one hand, intelligence could be raised to the same amount for all or, alternatively, intelligence could be raised by the same amount for all. In either case no one actually benefits over anyone else. . . . [The] aggregate effect, if all parents acted the same, would be that all their children would effectively be the same, in terms of outcome, as without selection.”¹⁶

Others have argued that the benefits of cognitive enhancement (for rationality, invention, or quality of life) could be very large and

14. See, e.g., Søren Holm, “Genetic Engineering and the North-South Divide,” in *Ethics and Biotechnology*, ed. A. Dyson and J. Harris (New York: Routledge, 1994), 47–63; Gregory S. Kavka, “Upside Risks: Social Consequences of Beneficial Biotechnology,” in *Are Genes Us? The Social Consequences of the New Genetics*, ed. C. Cranor (New Brunswick, NJ: Rutgers University Press, 1994), 155–79; George J. Annas, Lori B. Andrews, and Rosario M. Isasi, “Protecting the Endangered Human: Toward an International Treaty Prohibiting Cloning and Inheritable Alterations,” *American Journal of Law and Medicine* 28 (2002): 151–78; Francis Fukuyama, *Our Posthuman Future: Consequences of the Biotechnology Revolution* (New York: Farrar, Straus & Giroux, 2002); Leon Kass, *Life, Liberty, and the Defense of Dignity: The Challenge for Bioethics* (San Francisco: Encounter Books, 2002); Michael Sandel, “The Case against Perfection,” *Atlantic Monthly* 293 (2004): 51–62. For some of these, such as Leon Kass, it is sometimes difficult to discern to what extent the objection refers to the (narrow) effects of the intervention or to the mere fact that intervention and control is exercised. We partially address objections regarding the degree of control in Sec. V.

15. Holm, “Genetic Engineering and the North-South Divide,” 60 and n. 9.

16. Kean Birch, “Beneficence, Determinism and Justice: An Engagement with the Argument for the Genetic Selection of Intelligence,” *Bioethics* 16 (2005): 12–28, 24.

that many of the risks have been overstated.¹⁷ To proponents of this view, opinions like those expressed in the above quotation tend to seem puzzling: why think that greater mental faculties would be of no value if everybody shared in the improvement? Why be so suspicious of the consequences of the biological enhancement of intelligence when more familiar efforts to improve thinking ability (such as education) are met with near-universal approbation? To proponents, the idea that these negative judgments might derive partially from a bias against the new might seem plausible even without further argument. Opponents, of course, could return fire by charging proponents with a contrary bias in favor of the new. We need some way of adjudicating between the differing intuitions.

How can we determine whether the judgments opposing cognitive enhancement result from a status quo bias? One way to proceed is by reversing our perspective and asking a somewhat counterintuitive question: "Would using some method of safely lowering intelligence have net good consequences?"

The great majority of those who judge increases to intelligence to be worse than the status quo would likely also judge decreases to be worse than the status quo. But this puts them in the rather odd position of maintaining that the net value for society provided by our current level of intelligence is at a local optimum, with small changes in either direction producing something worse. We can then ask for an explanation of why this should be thought to be so. If no sufficient reason is provided, our suspicion that the original judgment was influenced by status quo bias is corroborated.

In its general form, the heuristic looks like this:

Reversal Test: When a proposal to change a certain parameter is thought to have bad overall consequences, consider a change to the same parameter in the opposite direction. If this is also thought to have bad overall consequences, then the onus is on those who reach these conclusions to explain why our position cannot be improved through changes to this parameter. If they are unable to

17. See, e.g., Ainsley Newson and Robert Williamson, "Should We Undertake Genetic Research on Intelligence?" *Bioethics* 13 (1999): 327–42; James Hudson, "What Kinds of People Should We Create?" *Journal of Applied Philosophy* 17 (2000): 131–43; Mark Walker, "Prolegomena to Any Future Philosophy," *Journal of Evolution and Technology* 10 (2002), <http://jetpress.org/contents.htm>; Nick Bostrom, "Human Genetic Enhancements: A Transhumanist Perspective," *Journal of Value Inquiry* 37 (2003): 493–506; see also Jonathan Glover, *What Sort of People Should There Be?* (Harmondsworth: Pelican, 1984); Anjan Chatterjee, "Cosmetic Neurology: The Controversy over Enhancing Movement, Mentation, and Mood," *Neurology* 63 (2004): 968–74; M. J. Farah, J. Illes, R. Cook-Deegan, H. Gardner, E. Kandel, P. King, E. Parens, B. Sahakian, and P. R. Wolpe, "Neurocognitive Enhancement: What Can We Do and What Should We Do?" *Nature Reviews Neuroscience* 5 (2004): 421.

do so, then we have reason to suspect that they suffer from status quo bias.

The rationale of the Reversal Test is simple: if a continuous parameter admits of a wide range of possible values, only a tiny subset of which can be local optima, then it is *prima facie* implausible that the actual value of that parameter should just happen to be at one of these rare local optima (fig. 1). This is why we claim that the burden of proof shifts to those who maintain that some actual parameter is at such a local optimum: they need to provide some good reason for supposing that it is so.

Obviously, the Reversal Test does not show that preferring the status quo is always unjustified. In many cases, it is possible to meet the challenge posed by the Reversal Test and thus to defeat the suspicion of status quo bias. Let us examine some of the possible ways in which one could try to do this in the case of medically safe, financially affordable, cognitive enhancement.

The Argument from Evolutionary Adaptation

For some biological parameters, one may argue on evolutionary grounds that it is likely that the current value is a local optimum. The idea is that we have adapted to live in a certain kind of environment, and that if a larger or a smaller value of the parameter had been a better adaptation, then evolution would have ensured that the parameter would have had this optimal value. For example, one could argue that the average ratio between heart size and body size is at a local optimum, because a suboptimal ratio would have been selected against. This argument would shift the burden of proof back on somebody who maintains that a particular person's heart—or the average human heart-to-body-size ratio—is too large or too small.

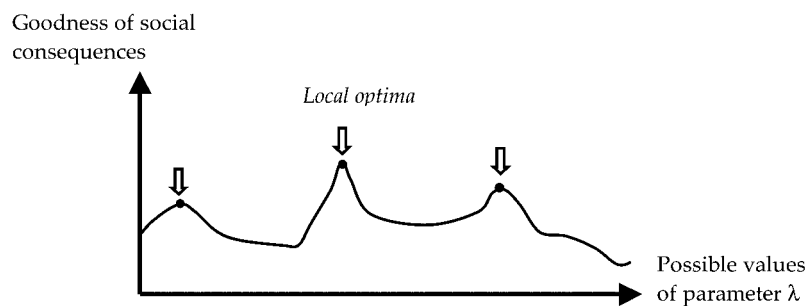


FIG. 1.—Only the points indicated with arrows are local optima. Typically, only a few points will be local optima, and most points will be such that a small shift in λ in the appropriate direction will increase the goodness of the social consequences.

The applicability of this evolutionary argument, however, is limited for several reasons. First, our current environment is in many respects very different from that of our evolutionary ancestors. A sweet tooth might have been adaptive in the Pleistocene, where high-calorie foods were scarce and the risk of starvation outweighed the health risks of a sugary diet. In wealthy modern societies, where a Mars bar is never far away, the risks of obesity and diabetes outweigh the risk of undernutrition, and a sweet tooth is now maladaptive. Our modern environment also places very different demands on cognitive functioning than did an illiterate life on the savanna: numeracy, literacy, logical reasoning, and the ability to concentrate on abstract material for prolonged periods of time have become important skills that facilitate successful participation in contemporary society.

Second, even if, say, a greater capacity for abstract reasoning had in itself been evolutionarily adaptive in the period of human evolutionary adaptation, there may have been trade-offs that made an increase in this parameter on balance maladaptive. For example, a larger brain might be correlated with greater cognitive capacity, yet a larger brain incurs substantial metabolic costs.¹⁸ These metabolic costs are no longer significant, thanks to the easy availability of food, suggesting that we may not be optimally adapted to the current environment. Similarly, the size of the birth canal used to place severe limitations on the head size of newborns, but this constraint is ameliorated by modern obstetrics and the possibility of cesarean section. An extended period of maturation was also vastly riskier ten thousand years ago than it is today.

Third, even if some trait would have been adaptive for our Pleistocene predecessors, there is no guarantee that evolutionary trial and error would have discovered it. This is especially likely for polygenic traits that are only adaptive once fully developed but that incur a fitness penalty in their intermediary stages of evolution. In some cases, the evolution of such traits may require an improbable coincidence of several simultaneous mutations that may simply not have occurred among our finite number of ancestors. An advanced genetic engineer, by contrast, may be able to solve some of the problems that proved intractable to blind evolution. She can think backward, starting with a goal in mind and working out what genetic modifications are necessary to attain it.

Fourth, there is no general reason for thinking that what evolution selects for—inclusive fitness—coincides with what makes our lives go well individually, much less collectively. The traits that would maximize our individual or collective well-being are not always the ones that maximize our tendency to propagate our genetic material. Evolution doesn't

18. Richard J. Haier, Rex E. Jung, Ronald A. Yeo, Kevin Head, and Michael T. Alkire, "Structural Brain Variation and General Intelligence," *Neuroimage* 23 (2004): 425.

care about human happiness. A capacity for rape, plunder, cheating, and cruelty might well have been evolutionarily adaptive, yet they have disastrous consequences for human welfare. Regarding intellectual faculties, we place a value on understanding, knowledge, and cognitive sensitivity that goes beyond the contribution these traits may make to our ability to survive and reproduce.

If we have reason for thinking that, for some human parameter, its role in contemporaneous society is identical to its role in Pleistocene human society, and that the trade-offs that this parameter involves have not changed, and that evolution would have had enough time to chance upon the optimum value, and that this parameter bears the same relation to human well-being as it did to reproductive success in the Pleistocene, then the argument from evolutionary adaptation would successfully meet the challenge posed by the Reversal Test. While these conditions may well hold for the heart-to-body-size ratio, they do not hold for all human parameters. In particular, they do not hold for human cognitive ability.

The Argument from Transition Costs

Consider the reluctance of the United States to move to the metric system of measurement units. While few would doubt the superiority of the metric system, it is nevertheless unclear whether the United States should adopt it. In cases like this, the transition costs are potentially so high as to overwhelm the benefits to be gained from the new situation. Those who oppose both increasing and decreasing some parameter can potentially appeal to such a rationale to explain why we should retain the status quo without having to insist that the status quo is (locally) optimal.

In the case of cognitive enhancements, one can anticipate many transition costs. Maybe school curricula would have to be redesigned to match the improved learning capacity of enhanced children. Tax codes and other regulations are often designed to strike a trade-off between how well they serve their intended function and how complex they are. (Complex regulations are harder to learn and enforce.) If people could learn complex rules more quickly, it may be appropriate to reevaluate these trade-offs and perhaps to adopt a more nuanced and complex set of social norms and regulations. Some games and recreational activities may likewise have to be modified to provide interesting levels of challenge to smarter participants. In the case of germline interventions, cognitively enhanced children might be raised by parents of normal cognitive ability, which could conceivably create some friction in such families and necessitate more preschool educational opportunities.

It is easy to overstate such transitional burdens. The cost would be

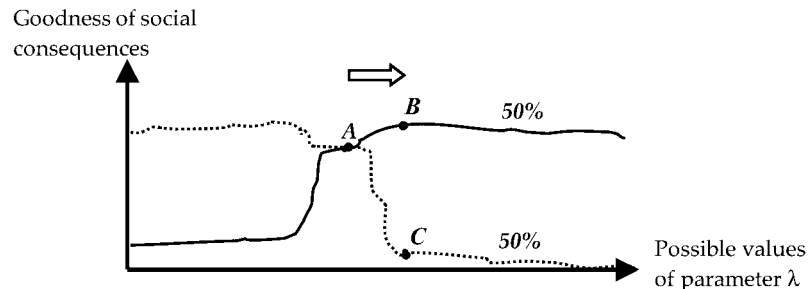


FIG. 2.—There is uncertainty whether the goodness of social consequences of a given value of a parameter λ is represented by the solid or the dotted line. A society that is currently in state A is not at a local optimum but may nevertheless resist a small shift in the parameter λ because of the risk that it would bring about state C rather than state B.

one-off while the benefits of enhancement would be permanent. School curricula are frequently rewritten for all sorts of trivial reasons. Modifying tax codes and regulations to fit a population with increased average intelligence would not be strictly necessary; it would simply be an opportunity to reap additional benefits of enhancement. Games and recreational activities are easy to invent, and we already have many games and cultural treasures that would presumably remain rewarding to people with substantially enhanced cognitive capacities. Even today, smart children are often raised by less smart parents, and while this might create problems in a few cases it certainly does not justify the conclusion that it would have been better, all things considered, if these children had been less talented than they are.

It would, however, be very difficult to exhaustively evaluate each possible transition cost against the permanent gains. Judgments about the balance between transition costs and long-term benefits will inevitably involve appeals to subjective intuitions. Such intuitions can easily be influenced by status quo bias. In Section IV, we will therefore present an extended version of our heuristic that takes account of transition costs.

The Argument from Risk

Even if it is agreed that we are probably not at a local optimum with respect to some parameter under consideration, one could still mount an argument from the risk against varying the parameter. If it is suspected that the potential gains from varying the parameter are quite low and the potential losses very high, it may be prudent to leave things as they are (fig. 2).

Uncertainty about the goodness of the consequences also means that results may be much better than anybody expected. It is not clear

that such uncertainty by itself provides any consequentialist ground whatsoever for resisting a proposed intervention. Only if the expectation value of the hypothetical negative results is larger than the expectation value of the hypothetical positive results does the uncertainty favor the preservation of the status quo.

The potential for unexpected gains should not be dismissed as a far-fetched theoretical possibility. In the case of cognitive enhancement, unanticipated consequences of enormous positive value seem not at all implausible. The fact that it may be easier to vividly imagine the possible downsides than the possible upsides of cognitive enhancement might psychologically join with other sources of human loss aversion to form a particularly strong status quo bias.

Imagine a tribe of *Australopithecus* debating whether they should enhance their intelligence to the level of modern humans. Is there any reason to suppose that they would have been able to foresee all the wonderful benefits we are enjoying thanks to our improved intellect? Only in retrospect did the myriad technological and social gains become apparent. And it would have been even less feasible for an *Australopithecus* to foresee the qualitative changes in our ways of experiencing, thinking, doing, and relating that our greater cognitive capacity have enabled, including literature, art, music, humor, poetry, and the rest of Mill's "higher pleasures." All these would have been impossible without our enhanced mental capacities; who knows what other wonderful things we are currently missing out on? It is an essential aspect of greater cognitive faculties that they facilitate new insights, inventions, and creative endeavors, as well as enabling new ways of thinking and experiencing. The uncertainty of the ultimate consequences of cognitive enhancement, far from being a sufficient ground for opposing them, is actually a strong consideration in their support.¹⁹

While some of the potential benefits might be hard to imagine, other benefits of greater cognitive faculties are quite plain. Diseases need cures, scientific questions need answers, poverty needs alleviation, and environmental problems need solutions. While a widespread increase in intelligence may not be sufficient to achieve all these goals, it could clearly help. Even the foreseeable benefits are very great.

One might object to this balancing of potential losses with potential gains by claiming that when it comes to the moral assessment of consequences, there is some normatively appropriate level of risk aversion which we must take into account. However, even if we accept such an account, and even if we completely disregard the possible gains mentioned above (both the unpredictable, and the more predictable ones),

19. Compare Nick Bostrom, "Transhumanist Values," in *Ethical Issues for the 21st Century*, ed. F. Adams (Charlottesville, VA: Philosophical Documentation Center, 2004).

it would still be difficult to make a case against cognitive enhancement. This is because while cognitive enhancement may create certain novel risks, it may also help to reduce many serious threats to humanity. In evaluating the riskiness of cognitive enhancement we must take into account both its risk-increasing and its risk-reducing effects. Mitigation of risk could result from a greater ability to protect ourselves from a wide range of natural hazards such as viral pandemics. There may also be threats to human civilization that we have not yet understood, but which greater intelligence would enable us to anticipate and counteract. The goal of reducing overall risk might turn out to be a strong reason for trying to develop ways to enhance our intelligence as soon as possible.²⁰

The Argument from Person-Affecting Ethics

Suppose that the cognitively enhanced would lead better lives. Does that give us a moral reason to enhance ourselves? Or to create cognitively enhanced people? It is possible to hold a person-affecting form of consequentialism according to which what we ought to do is to maximize the benefits we provide to people who either already exist or will come to exist independently of our decisions. On such views, there is no general moral reason to bring into existence people whose lives will be very good. By extension, there may be no moral reason to change ourselves into radically different sorts of people whose lives would be better than the ones we currently lead.²¹

Even if one accepts such a person-affecting ethics, one may still recognize moral reasons for supporting cognitive enhancement. In the case of bringing new people into existence, it would be difficult to deny that it would be a bad idea to deliberately select embryos with genetic disorders that cause severe retardation.²² This might indicate that one recognizes other types of moral considerations in addition to person-affecting ones or that one believes that selecting for mental retardation would adversely affect the existing population. Either way, the Reversal Test can be applied to put some pressure on those who hold these views to explain why the same kinds of reasons that make it a bad idea to

20. Compare Nick Bostrom, "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards," *Journal of Evolution and Technology* 9 (2002), <http://jetpress.org/contents.htm>.

21. See, e.g., Melinda A. Roberts, "A New Way of Doing the Best That We Can: Person-Based Consequentialism and the Equality Problem," *Ethics* 112 (2002): 315–50. Note that the idea of person-affecting ethics is not simply that what is good for one person may not be good for another. That the good for a person may partially depend on her preferences and other personal factors can of course be admitted by consequentialists who reject the person-affecting view.

22. Glover, *What Sort of People Should There Be?*

select for lower intelligence would not also make it a good idea to select for increased intelligence. For example, person-affecting reasons for bringing smarter children into the world could derive from many considerations, including the idea that present people might prefer to have such children or might benefit from being cared for in their old age by a more capable younger generation that could generate more economic resources for the elderly, invent more cures for diseases, and so on.

In the case of cognitively enhancing existing individuals, the consequentialist person-affecting reasons seem even stronger, at least for small or moderate enhancements. Practically everyone would agree that it would be the height of foolishness to set out to lower one's own intelligence, for instance, by deliberately ingesting lead paint. But if we think that becoming a little less intelligent would be bad for us, then we should either accept that becoming a little smarter would be good for us or else take on the burden of justifying the belief that we currently happen to have an optimal level of intelligence.

Very large cognitive enhancement for existing people is more problematic on a person-affecting view. A sufficiently radical enhancement might conceivably change an individual to such an extent that she would become a different person, an event that might be bad for the person that existed before. However, it is perhaps illuminating to make a comparison with children, whose cognitive capacities grow dramatically as they mature. Even though this eventually results in profound psychological changes, we don't think that it is bad for children to grow up. Similarly, it might be good for adults to continue to grow intellectually even if they eventually develop into rather different kinds of persons.²³

To summarize this section, we have proposed a heuristic for eliminating status quo bias, which transfers the onus of justification to those who reject both increasing and decreasing some human parameter. We illustrated this heuristic on the case of proposed cognitive enhancement. We considered four broad arguments by which one might attempt to carry the burden of justification, and we tried to show that, in regard to intelligence enhancement, these arguments do not succeed.

Our argument that status quo bias is widespread in bioethics thus proceeds in two steps. First, we note that the empirical literature shows that status quo bias affects many domains of human cognition, creating a *prima facie* reason for suspecting that it might affect some bioethical judgments in particular. Second, we apply the Reversal Test. Since the

23. However, in general, if the proposed change in a parameter is very large, the Reversal Test will tend to give a less definite verdict. This is because there is less *prima facie* implausibility in supposing that a larger interval of parameter values contains a local optimum than that a smaller interval does.

function of the Reversal Test is to remove whatever status quo bias is present, we infer that if our considered judgments change after the test has been applied, then our judgments prior to its implication were in fact affected by status quo bias. The four arguments we considered above are our best attempts, on behalf of the opponents of cognitive enhancement, to try to meet the burden of justification that the Reversal Test generates. We are not aware of any other arguments that have been advanced in the literature that could do this job. The test's challenge, of course, does not depend on its targets actually having offered these hypothetical arguments. If they have not and are not able to pass the test in some other way, then the indictment of status quo bias stands.

In order to further strengthen these conclusions, we will now present an extended version of the heuristic, which we call the Double Reversal Test. This version is especially useful in addressing the argument from transition costs and the argument from person-affecting ethics.

IV. THE DOUBLE REVERSAL TEST

Disaster! A hazardous chemical has entered our water supply. Try as we might, there is no way to get the poison out of the system, and there is no alternative water source. The poison will cause mild brain damage and thus reduced cognitive functioning in the current population. Fortunately, however, scientists have just developed a safe and affordable form of somatic gene therapy which, if used, will permanently increase our intellectual powers just enough to offset the toxicity-induced brain damage. Surely we should take the enhancement to prevent a decrease in our cognitive functioning.

Many years later it is found that the chemical is about to vanish from the water, allowing us to recover gradually from the brain damage. If we do nothing, we will become more intelligent, since our permanent cognitive enhancement will no longer be offset by continued poisoning. Ought we try to find some means of reducing our cognitive capacity to offset this change? Should we, for instance, deliberately pour poison into our water supply to preserve the brain damage or perhaps even undergo simple neurosurgery to keep our intelligence at the level of the status quo? Surely, it would be absurd to do so. Yet if we don't poison our water supply, the consequences will be equivalent to the consequences that would have resulted from performing cognitive enhancement in the case where the water supply hadn't been contaminated in the first place. Since it is good if no poison is added to the water supply in the present scenario, it is also good, in the scenario where the water was never poisoned, to replace that status quo with a state in which we are cognitively enhanced.

The argument contained in this thought experiment can be generalized into the following heuristic:

Double Reversal Test: Suppose it is thought that increasing a certain parameter and decreasing it would both have bad overall consequences. Consider a scenario in which a natural factor threatens to move the parameter in one direction and ask whether it would be good to counterbalance this change by an intervention to preserve the status quo. If so, consider a later time when the naturally occurring factor is about to vanish and ask whether it would be a good idea to intervene to reverse the first intervention. If not, then there is a strong *prima facie* case for thinking that it would be good to make the first intervention even in the absence of the natural countervailing factor.

The Double Reversal Test works by combining two possible perceptions of the status quo. On the one hand, the status quo can be thought of as defined by the current (average) value of the parameter in question. To preserve this status quo, we intervene to offset the decrease in cognitive ability that would result from exposure to the hazardous chemical. On the other hand, the status quo can also be thought of as the default state of affairs that results if we do not intervene. To preserve this status quo, we abstain from reversing the original cognitive enhancement when the damaging effects of the poisoning are about to wear off. By contrasting these two perceptions of the status quo, we can pin down the influence that status quo bias exerts on our intuitions about the expected benefit of modifying the parameter in our actual situation.

When this extended heuristic for assessing status quo bias can be applied, it accommodates a wider range of considerations than the simple Reversal Test. While the challenge posed by the Reversal Test can potentially be met in any of the several ways discussed above, the challenge posed by the Double Reversal Test already incorporates the possible arguments from evolutionary adaptation, transition costs, risk, and person-affecting morality into the overall assessment it makes. If we judge that, all things considered, it would be bad to reverse the original intervention when the natural factor disappears, this judgment already incorporates all these arguments.

The Double Reversal Test yields a particularly strong consequentialist reason for cognitive enhancement. While there may be a relevant difference between the two scenarios in terms of nonconsequentialist considerations (such as the distinction between acting and allowing), it is very difficult to find a difference in the expected consequences that could plausibly be thought of as decisive. Perhaps one could speculate that in the poisoning scenario, people would already have got used to

the idea of using a cognitive enhancement therapy, even though its effects were initially concealed by the presence of the natural factor. When the natural factor disappears, there might then be less psychological discomfort from allowing the enhancement to continue to operate. However, while such an effect is possible in principle, it seems unlikely that this speculative effect would be significant in realistic cases and utterly implausible to suppose that it could form a sufficient ground for opposing cognitive enhancement.²⁴

V. APPLYING THE REVERSAL TESTS TO OTHER CASES

We have illustrated the reversal heuristics on the hypothetical case of a medically safe and generally affordable enhancement of a population's cognitive capacity. The Reversal Tests, however, can be applied much more generally.

Consider a case where inequality and the distributional effects of an enhancement are concerns. Suppose, for example, that a cognitive enhancement could not be applied universally but only to some subset of the population. This might be because only the wealthy can afford to pay for it, or perhaps because certain groups decide not to avail themselves of the enhancement opportunity (perhaps on religious grounds). The development of such an enhancement would then potentially have negative consequences for social equality, and we may ask whether the benefits it would provide would be large enough to outweigh these potential inequality-increasing effects.

One way to approach this question would be to try to estimate the effects on social inequality that the development would have, come to some evaluative assessment of the severity of these effects, compare this assessment with an evaluation of the expected beneficial consequences that the enhancement technology would have, and then form a judgment of the overall expected goodness of the consequences based on

24. Alternatively, one could speculate that the direct enhancement of cognitive ability would set a different kind of precedent than either the "therapeutic enhancement" to compensate for a natural brain-damaging factor or the subsequent increase in cognitive ability that results when the natural factor disappears. But this speculation would have to be justified. If the idea is that direct cognitive enhancement would lead to further cognitive enhancement, it would have to be shown that (1) this is significantly more likely to result from direct cognitive enhancement than from therapeutic enhancement followed by a natural increase and (2) that further cognitive enhancement would be bad. But consider an iterated application of the Double Reversal Test: a series of disasters occur in which neurotoxins are released, each followed by a therapeutic enhancement to preserve the status quo and a subsequent elevation of cognitive ability when the neurotoxin disappears. At the end of the series, average cognitive ability is at a much higher level than it is today. Is there any point in this series where the brain damage ought not be compensated for by a therapeutic enhancement, or where it would be better to prevent the ensuing rise by preserving the brain-damaging factor?

this comparison. To consider the consequentialist grounds for enhancement means, of course, that one way or another we must make such a comparison. But realistically, there is no possibility of making this comparison in a completely scientifically rigorous way. Subjective intuitive judgment will inevitably enter into the assessment—both of what the likely consequences would be and of the goodness or badness of these consequences. We must therefore confront the possibility that these intuitions, which we perforce rely on, are biased in some way, and in particular the possibility that they are affected by status quo bias. This is where the Reversal Tests come in. Potential consequences that involve distributive concerns can be handled by the tests in the same way as other consequences.

In the case of cognitive enhancement, we can apply the simple Reversal Test and ask whether it would be a good thing if the treatment group (those who would be given the cognitive enhancement) instead had their cognitive capacity reduced. Are we prepared to claim that the status quo would be improved if the wealthy, say, suffered slight brain damage? If we are not prepared to make that claim, then the onus shifts to those who judge that the nonuniversal use of the cognitive enhancer would have on balance bad consequences: they need to explain why we should believe that the current cognitive ability of the potential enhancement users is at a local optimum such that both an increase and a decrease should be expected to make things worse on the whole.²⁵

We can also apply the Double Reversal Test. If the release of a hazardous chemical threatened to reduce cognitive ability among the potential enhancement users, would it be a good thing if they could use the permanent enhancement to stave off the impending decline? And if so, would it also be a good thing if, when the effects of the poison eventually started to wear off, the enhancement users refrained from taking steps to maintain their intellectual status quo (e.g., by injecting themselves with a neurotoxin)? If the answer to both these questions is yes, then there is a strong *prima facie* case for thinking that it would

25. Those (if any) who hold the opposite view should also address, e.g., whether the world would be better if nobody had access to expensive AIDS treatments, given that such treatments are not currently available to everybody. Or, to take a case more closely related to the one at hand, whether it would have been better in the past if nobody had been taught to read given that only elites had access to education. And considering that literacy is still far from universal, especially in the poorest countries, would it be better if nobody in those countries (or in developed countries?) were given this kind of cognitive enhancement unless and until everybody gets it? In such cases surely, *le mieux est l'ennemi du bien*.

be good overall—despite the assumed negative effect on equality—if the enhancement option is developed.²⁶

In a real-world situation, we are often interested in evaluating more than two alternatives. For example, we might conclude that even though it is better that the new enhancement option be allowed to reach the market than that it be banned, it is better still if its introduction be accompanied with inequality-reducing measures, for example, making the enhancement available via public health care at an affordable price. The Reversal Test could in principle be applied to evaluate each pair of policy options in turn. For instance, we could ask whether, given that the enhancement will be allowed on the market, it would be better if an inequality-*increasing* measure were implemented, and if the answer is no, we could place the onus on those who would maintain that neither inequality-increasing nor inequality-decreasing policies should be put in place to explain why we should think that the default degree of redistribution is optimal.²⁷ We can also apply the Reversal Test to cases of individual prudential decision making.

We have used the example of enhancement of cognitive ability, but the same considerations and the same heuristics can be applied to many other forms of human modification, such as proposed interventions to enhance the ability to concentrate, improve emotional well-being, reduce the need for sleep, or increase physical or sensory capacities. Those who deny that it would be a good thing for the healthy human life span to be extended may want to ask themselves whether they believe that it would be a good thing if health span were shortened, and if not, what reason there is for thinking that the current health span is optimal. They should also apply the Double Reversal Test and consider whether, for example, slowing the endogenous aging rate would be a good thing if it would serve to counterbalance some impending environmental factor that would otherwise shorten health spans; if this would be desirable,

26. There is a specific limitation when it comes to using the Reversal Test to address the issue of inequality. For the potential users who are already privileged in the status quo, the options of increasing the parameter and of decreasing it are opposites in terms of both equality and cognitive benefits. This allows our argument above to go through. However, for those who are at the average level of welfare in the society in the status quo this is not the case. While the options of increasing or decreasing cognitive ability will have opposite effects in the cognitive realm, both of these options will decrease social equality when applied to such a person. This is because in this situation any change to their well-being will create inequality: the equality of society is at a local optimum with respect to their welfare.

27. The Reversal Tests may sometimes appear to have less power to change opinion on matters of economic policy than on matters of human modification. If this is so, it might indicate that status quo bias is less pervasive in intuitions about economic policy, perhaps because we are more experienced in thinking about changes in economic policy than about changes in human nature.

then they should ask whether, were the environmental factor to be about to disappear, it would be desirable to take steps to preserve this damaging factor or else to adopt some alternative countermeasure (such as heavy smoking or an unhealthy diet) to retain the health-span status quo.

Beyond Therapy, the widely cited recent report produced by the President's Council on Bioethics, at one point comes tantalizingly close to considering a Reversal Test. After expressing many qualms and reservations about the consequences of using medical technology to extend the healthy human life span, the report reflects: "Yet if there is merit in the suggestion that too long a life, with its end out of sight and mind, might diminish its worth, one might wonder whether we have already gone too far in increasing longevity. If so, one might further suggest that we should, if we could, roll back at least some of the increases made in the average human lifespan over the past century."²⁸

In the next paragraph, the council makes clear that it does not favor such a rollback: "[Nothing] in our inquiry ought to suggest that the present average lifespan is itself ideal. We do not take the present (or any specific time past) to be 'the best of all possible worlds,' and we would not favor rolling back the average lifespan even if it were doable. Although we suggest some possible problems with substantially longer lifespans, we have not expressed, and would not express, a wish for shorter lifespans than are now the norm."²⁹

Having brought up the challenge, the council then unfortunately drops the subject after noting that while life expectancy (in the United States) has increased by thirty years in the last century, maximum life span has not changed much. But the reversal heuristic can be applied to hypothetical changes in either average or maximum life span. If the council believes that both shorter and longer maximum life spans would be worse than the present maximum life span, it owes us a convincing argument for why we should think this is so. It would have been interesting to know what conclusions the council members would have drawn if they had considered the reversal question more seriously. Would they have concluded that a shorter (maximum) life span would, after all, be preferable? Or that the current life span is exactly right? Or would they have changed their view and come out unambiguously in favor of a longer life span? Either way, the result would have been noteworthy and would have made it easier to assess the plausibility of the council's position.

The reversal heuristics do not indiscriminately favor all human en-

28. President's Council on Bioethics (U.S.), *Beyond Therapy: Biotechnology and the Pursuit of Happiness*, foreword by Leon Kass (Washington, DC: President's Council on Bioethics, 2004), 224.

29. Ibid.

hancements. For example, if we contemplate some intervention that would make everyone four inches taller, we may well come to the verdict that this would yield no net benefit. Clothes, buildings, vehicles, and so on, are designed for the current distribution of heights, so that changing average height would incur some costs. Since there are no obvious counterbalancing benefits from everybody being taller, these transition costs could justify the judgment that it is better to stick with the status quo. If we could safely and easily intervene to prevent an impending decrease in average height, say by administering growth hormone, we may have reason to do so; if whatever factor would otherwise have led to reduced height were to disappear, we might have reason to stop taking the growth hormone or to make some other intervention to prevent average height from increasing.³⁰

The Reversal Tests can be applied not only to choices affecting currently existing people but also to choices that affect what new types of people are brought into existence. Such choices, we may note, arise not only in the context of preimplantation genetic diagnosis, embryo screening, and possible future cases involving germ-line genetic modification but also in the contexts of maternal nutrition (e.g., whether to take a folic acid supplement to reduce the risk of neural tube defects), lifestyle (e.g., whether to abstain from heavy drinking during pregnancy), and timing (e.g., whether to conceive while suffering from rubella, or whether to postpone childbearing into one's forties). We can ask whether, from a consequentialist stance, it is better that a greater proportion of newborns are healthy or intelligent. Some critics of germ-line genetic enhancement have expressed doubt that it would be better if newborns had greater mental capacities. Applying the Reversal Test to this issue, we would ask whether it would be better if newborns had less intellectual capacity. If the answer is no, then we must ask for a strong justification for thinking that the current distribution of intellectual capacity in newborns is optimal.

Drawing moral conclusions about practices that influence what new types of people there should be may also require taking into account various deontological side constraints in addition to consequentialist considerations. Julian Savulescu, for example, has argued that parents have an obligation to select for the best children even if no net social

30. This example is not meant to be realistic. In the real world we have reason to celebrate the trend of increasing average height, as it is associated with beneficial developments resulting from improved nutrition. It is extremely implausible that the inconveniences of an increasing population height could ever be significant enough to outweigh the inevitable medical risks and costs of intervening to halt this trend (even setting aside important side constraints such as respect for individual autonomy).

benefit results.³¹ Others have opposed germ-line interventions on grounds that they involve an unjustifiable form of “tyranny” of the living over the unborn.³² While the heuristics offered in this article cannot fully address such deontological considerations, they may nevertheless be applied to check our intuitions for status quo bias insofar as consequentialist aspects feature in these deontological arguments.

For example, if the degree of control that present generations exercise over future ones is something that we should allegedly not increase by using germ-line therapy, we can apply the Reversal Test and ask whether we should instead reduce our control. Parents currently exert considerable influence over what new kinds of people there will be, through assortative mating, decisions whether to postpone pregnancies, the usage of preimplantation and embryonic screening, maternal nutrition, and child-rearing practices. If it is thought that it would be bad if parents had more influence over the traits of people-to-be, we should ask if it would be good if they had *less* influence. If this is not the case, we have reason for suspecting status quo bias. Michael Sandel, who argues against genetic enhancements on grounds that it is good for us to be “open to the unbidden,” seems to hold that it would be better if parents exerted less influence over their offspring than they currently do.³³ His view, therefore, may pass the Reversal Test.

The reversal heuristic is in principle applicable to any situation where we want to evaluate the consequences of some proposed change of a continuous parameter. However, its usefulness will vary from case to case. In many instances, it is possible to meet the challenge of the Reversal Tests: the method will certainly not always favor change over the status quo. The power of the heuristic lies in its ability to diagnose cases where status quo bias must be suspected and to challenge defenders of the status quo in these cases to provide further justification for their views. To what extent the example of cognitive enhancement generalizes to other issues remains to be seen, but the illustrations considered above suggest that the phenomenon is widespread. One might speculate that the popular intuition about the preferability of “the natural” might in part derive from a status quo bias. If so, then the manifestations of this bias may be endemic in human enhancement ethics and possibly in other parts of ethics as well.

A tool now exists for diagnosing status quo bias. While some reliance on intuitive judgment is unavoidable, there is no excuse for failing to test our intuitions with the most sophisticated methods available.

31. Julian Savulescu, “Procreative Beneficence: Why We Should Select the Best Children,” *Bioethics* 15 (2001): 413–26.

32. See, e.g., Jürgen Habermas, *The Future of Human Nature* (London: Blackwell, 2003).

33. Sandel, “The Case against Perfection.”

Annexe 6

Publication Specimen

“How Unlikely is a Domsday Catastrophe” (Tegmark & Bostrom)

How Unlikely is a Doomsday Catastrophe?

Max Tegmark¹ & Nick Bostrom²

¹*Department of Physics, Massachusetts Institute of Technology,
Cambridge, MA 02139, USA*

and

²*Future of Humanity Institute, Faculty of Philosophy,
Oxford University, OX14JJ, Oxford, UK*

(Dated: December 18, 2005. This paper is an extended version of the *Brief Communication* published in *Nature*, **438**, 754 [1].)

Numerous Earth-destroying doomsday scenarios have recently been analyzed, including breakdown of a metastable vacuum state and planetary destruction triggered by a “strangelet” or microscopic black hole. We point out that many previous bounds on their frequency give a false sense of security: one cannot infer that such events are rare from the fact that Earth has survived for so long, because observers are by definition in places lucky enough to have avoided destruction. We derive a new upper bound of one per 10^9 years (99.9% c.l.) on the exogenous terminal catastrophe rate that is free of such selection bias, using planetary age distributions and the relatively late formation time of Earth.

I. INTRODUCTION

As if we humans did not have enough to worry about, scientists have recently highlighted catastrophic scenarios that could destroy not only our civilization, but perhaps even our planet or our entire observable universe. For instance, fears that heavy ion collisions at the Brookhaven Relativistic Heavy Ion Collider (RHIC) might initiate such a catastrophic process triggered a detailed technical report on the subject [2], focusing on three risk categories:

1. Initiation of a transition to a lower vacuum state, which would propagate outward from its source at the speed of light, possibly destroying the universe as we know it [2, 3, 4].
2. Formation of a black hole or gravitational singularity that accretes ordinary matter, possibly destroying Earth [2, 4].
3. Formation of a stable “strangelet” that accretes ordinary matter and converts it to strange matter, possibly destroying Earth [2, 5].

Other catastrophe scenarios range from uncontroversial to highly speculative:

4. Massive asteroid impacts, nearby supernova explosions and/or gamma-ray bursts, potentially sterilizing Earth.
5. Annihilation by a hostile space-colonizing robot race.

The Brookhaven report [2] concluded that if 1-3 are possible, then they will with overwhelming likelihood be triggered not by RHIC, but by naturally occurring high-energy astrophysical events such as cosmic ray collisions. Risks 1-5 should probably all be treated as *exogenous*,

i.e., uncorrelated with human activities and our technical level of development. The purpose of the present paper is to assess the likelihood per unit time of exogenous catastrophic scenarios such as 1-5.

One might think that since life here on Earth has survived for nearly 4 Gyr (Gigayears), such catastrophic events must be extremely rare. Unfortunately, such an argument is flawed, giving us a false sense of security. It fails to take into account the observation selection effect [6, 7] that precludes any observer from observing anything other than that their own species has survived up to the point where they make the observation. Even if the frequency of cosmic catastrophes were very high, we should still expect to find ourselves on a planet that had not yet been destroyed. The fact that we are still alive does not even seem to rule out the hypothesis that the average cosmic neighborhood is typically sterilized by vacuum decay, say, every 10000 years, and that our own planet has just been extremely lucky up until now. If this hypothesis were true, future prospects would be bleak.

We propose a way to derive an upper bound on cosmic catastrophe frequency that is unbiased by such observer selection effects. We argue that planetary and stellar age distributions bound the rates of many doomsday scenarios, and that scenarios evading this bound (notably vacuum decay) are instead constrained by the relatively late formation time of Earth. The idea is that if catastrophes were very frequent, then almost all intelligent civilizations would arise much earlier than ours did. Using data on planet formation rates, it is possible to calculate the distribution of birth dates for intelligent species under different assumptions about the rate of cosmic sterilization. Combining this with the information about our own temporal location enables us to conclude that the cosmic sterilization rate for a habitable planet is at most of order one per Gigayear.

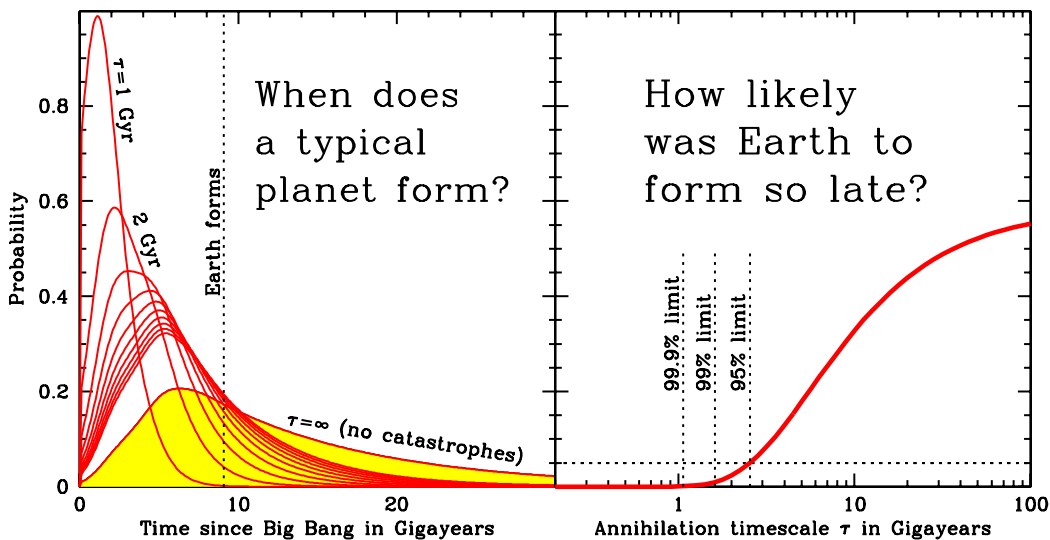


FIG. 1: The left panel shows the probability distribution for observed planet formation time assuming catastrophe timescales τ of ∞ (shaded), 10, 9, 8, 7, 6, 5, 4, 3, 2 and 1 Gyr, respectively (from right to left). The right panel shows the probability of observing a formation time ≥ 9.1 Gyr (that for Earth), *i.e.*, the area to the right of the dotted line in the left panel.

II. AN UPPER BOUND ON THE CATASTROPHE RATE

Suppose planets get randomly sterilized or destroyed at some rate τ^{-1} which we will now constrain. This means that the probability of a planet surviving a time t decays exponentially, as $e^{-t/\tau}$.

The most straightforward way of eliminating observer selection bias is to use only information from objects whose destruction would not yet have affected life on Earth. We know that no planets from Mercury to Neptune in our solar system have been converted to black holes or blobs of strange matter during the past 4.6 Gyr, since their masses would still be detectable via their gravitational perturbations of the orbits of other planets. This implies that the destruction timescale τ must be correspondingly large — unless their destruction is linked to ours, either by a common cause or by their implosion resulting in the emission of doomsday particles like black holes or strangelets that would in turn destroy Earth. This observer selection effect loophole is tightened if we consider extrasolar planets that have been seen to partially eclipse their parent star [8] and are therefore known not to have imploded. The doomsday particles discussed in the literature would be more readily captured gravitationally by a star than by a planet, in which case the observed abundance of very old ($\gtrsim 10$ Gyr) stars (*e.g.*, [9]) would further sharpen the lower bound on τ .

The one disaster scenario that exploits the remaining observer bias loophole and evades all these constraints is vacuum decay, either spontaneous or triggered by a high-energy event. Since the bubble of destruction expands with the speed of light, we are prevented from observing the destruction of other objects: we only see their destruction at the instant when we ourselves get destroyed.

In contrast, if scenarios 2 or 3 involved doomsday particle emission and proceed as a chain reaction spreading sub-luminally, we would observe spherical dark regions created by expanding destruction fronts that have not yet reached us. We will now show that the vacuum decay timescale can be bounded by a different argument.

The formation rate $f_p(t_p)$ of habitable planets as a function of time since the Big Bang is shown in Figure 1 (left panel, shaded distribution). This estimate is from [10], based on simulations including the effects of heavy element buildup, supernova explosions and gamma-ray bursts. If regions of space get randomly sterilized or destroyed at a rate τ^{-1} , then the probability that a random spatial region remains unscathed decays as $e^{-t/\tau}$. This implies that the conditional probability distribution $f_p^*(t_p)$ for the planet formation time t_p seen by an observer is simply the shaded distribution $f_p(t_p)$ multiplied by $e^{-t_p/\tau}$ and rescaled to integrate to unity, giving the additional curves in Figure 1 (left panel).¹ As we lower the catastrophe timescale τ , the resulting distributions (left panel) are seen to peak further to the left and

¹ Proof: Let $f_o(t_o)$ denote the probability distribution for the time t_o after planet formation when an observer measures t_p . In our case, $t_o = 4.6$ Gyr. We obviously know very little about this function f_o , but it fortunately drops out of our calculation. The conditional probability distribution for t_p , marginalized over t_o , is

$$f_p^*(t_p) \propto \int_0^\infty f_o(t_o) f_p(t_p) e^{-\frac{t_o+t_p}{\tau}} dt_o \propto f_p(t_p) e^{-\frac{t_p}{\tau}}, \quad (1)$$

independently of the unknown distribution $f_o(t_o)$, since $e^{-(t_o+t_p)/\tau} = e^{-t_o/\tau} e^{-t_p/\tau}$ and hence the entire integrand is separable into a factor depending on t_p and a factor depending on t_o .

the probability that Earth formed as late as observed (9.1 Gyr after the Big Bang) or later drops (right panel). The dotted lines show that we can rule out the hypothesis that $\tau < 2.5$ Gyr at 95% confidence, and that the corresponding 99% and 99.9% confidence limits are $\tau > 1.6$ Gyr and $\tau > 1.1$ Gyr, respectively.

Risk category 4 is unique in that we have good direct measurements of the frequency of impacts, supernovae and gamma-ray bursts that are free from observer selection effects. Our analysis therefore used the habitable planet statistics from [10] that folded in such category 4 measurements.

Our bounds do not apply in general to disasters of anthropogenic origin, such as ones that become possible only after certain technologies have been developed, *e.g.*, nuclear annihilation or extinction via engineered microorganisms or nanotechnology. Nor do they apply to natural catastrophes that would not permanently destroy or sterilize a planet. In other words, we still have plenty to worry about [11, 12, 13, 14]. However, our bounds do apply to exogenous catastrophes (*e.g.*, spontaneous or cosmic ray triggered ones) whose frequency is uncorrelated with human activities, as long as they cause permanent sterilization.

Our numerical calculations made a number of assumptions. For instance, we treated the exogenous catastrophe rate τ^{-1} as constant, even though one could easily imagine it varying by of order 10% over the relevant timescale, since our bound on τ is about 10% of the age of the Universe.² Second, the habitable planet formation rate involved several assumptions detailed in [10] which could readily modulate the results by 20%. Third, the risk from events triggered by cosmic rays will vary slightly with location if the cosmic ray rate does. Fourth, due to cosmological mass density fluctuations, the mass to scatter off of varies by about 10% from one region of size $c\tau \sim 10^9$ lightyear region to another, so the risk of cosmic-ray triggered vacuum decay will vary on the same order.

In summary, although a more detailed calculation could change the quantitative bounds by a factor of order unity, our basic result that the exogenous extinction rate is tiny on human and even geological timescales appears

rather robust.

III. CONCLUSIONS

We have shown that life on our planet is highly unlikely to be annihilated by an exogenous catastrophe during the next 10^9 years. This numerical limit comes from the scenario on which we have the weakest constraints: vacuum decay, constrained only by the relatively late formation time of Earth. conclusion also translates into a bound on hypothetical anthropogenic disasters caused by high-energy particle accelerators (risks 1-3).

This holds because the occurrence of exogenous catastrophes, *e.g.*, resulting from cosmic ray collisions, places an upper bound on the frequency of their anthropogenic counterparts. Hence our result closes the logical loop-hole of selection bias and gives reassurance that the risk of accelerator-triggered doomsday is extremely small, so long as events equivalent to those in our experiments occur more frequently in the natural environment. Specifically, the Brookhaven Report [2] suggests that possible disasters would be triggered at a rate that is at the very least 10^3 times higher for naturally occurring events than for high-energy particle accelerators. Assuming that this is correct, our 1 Gyr limit therefore translates into a conservative upper bound of $1/10^3 \times 10^9 = 10^{-12}$ on the annual risk from accelerators, which is reassuringly small.

² As pointed out by Jordi Miralda-Escude (private communication), the constraint from vacuum decay triggered by bubble nucleation is even stronger than our conservative estimate. The probability that a given point is not in a decayed domain at time t is the probability of no bubble nucleations in its backward light cone, whose spacetime 4-volume $\propto t^4$ for both matter-dominated and radiation-dominated expansion. A constant nucleation rate per unit volume per unit time therefore gives a survival probability $e^{-(t/\tau)^4}$ for some destruction timescale τ . Repeating our analysis with $e^{-t/\tau}$ replaced by the sharper cutoff $e^{-(t/\tau)^4}$ sharpens our constraint. Our quoted bound corresponds to the conservative case where τ greatly exceeds the age of the universe at the dark energy domination epoch, which gives a backward lightcone volume $\propto t$.

Acknowledgements:

The authors are grateful to Adrian Kent, Jordi Miralda-Escude and Frank Zimmermann for spotting loopholes in the first version of this paper, to the authors of [10] for use of their data, and to Milan Circovic, Hunter Monroe, John Leslie, Rainer Plaga and Martin Rees for helpful comments and discussions, Thanks to Paul

Davies, Charles Harper, Andrei Linde and the John Templeton Foundation for organizing a workshop where this work was initiated. This work was supported by NASA grant NAG5-11099, NSF CAREER grant AST-0134999, and fellowships from the David and Lucile Packard Foundation and the Research Corporation.

-
- [1] M. Tegmark and N. Bostrom, *Nature*, **438**, 754 (2005)
 - [2] R. L. Jaffe, W. Busza, Sandweiss J, and F. Wilczek, *Rev.Mod.Phys.*, **72**, 1125 (2000)
 - [3] P. Frampton, *Phys. Rev. Lett.*, **37**, 1378 (1976)
 - [4] P. Hut and M. J. Rees 1983, “How Stable Is Our Vacuum?”, *Nature*, **302**, 508 P. Hut 1984, *Nucl.Phys. A*, **418**, 301C
 - [5] A. Dar, A. De Rujula, and U. Heinz, *Phys.Lett. B*, **470**, 142 (1999)
 - [6] B. Carter 1974, in *IAU Symposium 63*, ed. M. S. Longair (Reidel: Dordrecht)
 - [7] N. Bostrom, *Anthropic Bias: Observation Selection Effects in Science and Philosophy* (Routledge: New York, 2002)
 - [8] F. Pont, astro-ph/0510846, 2005
 - [9] B. M. S Hansen *et al.*, *ApJ*, **574**, L155 (2002)
 - [10] C. H. Lineweaver, Y. Fenner, and B. K. Gibson, *Science*, **203**, 59 (2004)
 - [11] J. Leslie, *The End of the World: The Science and Ethics of Human Extinction* (Routledge: London, 1996)
 - [12] N. Bostrom, *Journal of Evolution and Technology*, **9**, 1 (2002)
 - [13] M. J. Rees, *Our Final Hour: How Terror, Error, and Environmental Disaster Threaten Humankind’s Future in This Century — On Earth and Beyond* (Perseus: New York, 2003)
 - [14] R. Posner, *Catastrophe: Risk and Response* (Oxford Univ. Press: Oxford, 2004)

Annexe 7

Publication Specimen

“The Wisdom of Nature” (Bostrom & Sandberg)

The Wisdom of Nature: An Evolutionary Heuristic for Human Enhancement

(2007)

Nick Bostrom
Anders Sandberg

Oxford Future of Humanity Institute
Faculty of Philosophy & James Martin 21st Century School
Oxford University

Forthcoming in *Enhancing Humans*, eds. Julian Savulescu and Nick Bostrom (Oxford: Oxford University Press)

www.nickbostrom.com

Abstract

Human beings are a marvel of evolved complexity. Such systems can be difficult to enhance. When we manipulate complex evolved systems which are poorly understood, our interventions often fail or backfire. It can appear as if there is a “wisdom of nature” which we ignore at our peril. Sometimes the belief in nature’s wisdom – and corresponding doubts about the prudence of tampering with nature, especially human nature – manifest as diffusely moral objections against enhancement. Such objections may be expressed as intuitions about the superiority of the natural or the troublesomeness of hubris, or as an evaluative bias in favor of the status quo. This paper explores the extent to which such prudence-derived anti-enhancement sentiments are justified. We develop a heuristic, inspired by the field of evolutionary medicine, for identifying promising human enhancement interventions. The heuristic incorporates the grains of truth contained in “nature knows best” attitudes while providing criteria for the special cases where we have reason to believe that it is feasible for us to improve on nature.

1. Introduction

1.1 The wisdom of nature, and the special problem of enhancement

We marvel at the complexity of the human organism, how its various parts have evolved to solve intricate problems: the eye to collect and pre-process visual information, the immune system to fight infection and cancer, the lungs to oxygenate the blood. The human brain – the focus of many of the most alluring proposed enhancements – is arguably the most complex thing in the known universe. Given how rudimentary is our understanding of the human organism, particularly the brain, how could we have any realistic hope of *enhancing* such a system?

To enhance even a system like a car or a motorcycle – whose complexity is trivial in comparison to that of the human organism – requires a fair bit of understanding of how

the thing works. Isn't the challenge we face in trying to enhance human beings so difficult as to be hopelessly beyond our reach, at least until the biological sciences and the general level of human abilities have advanced vastly beyond their present state?

It is easier to see how *therapeutic* medicine should be feasible. Intuitively, the explanation would go as follows: Even a very excellently designed system will occasionally break. We might then be able to figure out what has broken, and how to fix it. This seems much less daunting than to take a very excellently designed, unbroken system, and enhance it beyond its normal functioning.

Yet we know that even therapeutic medicine is very difficult. It has been claimed that until circa 1900, medicine did more harm than good.¹ And various recent studies suggest that even much of contemporary medicine is ineffectual or outright harmful.² Iatrogenic deaths account for 2-4% of all deaths in the US (the third leading cause of death according to one accounting³) and may correspond to a loss of life expectancy by 6-12 months.⁴ We are all familiar with nutritional advice, drugs, and therapies that were promoted by health authorities but later found to be damaging to health. In many cases, the initial recommendations were informed by large clinical trials. When even therapeutic medicine, based on fairly good data from large clinical trials, is so hard to get right, it seems that a prudent person has much reason to be wary of purported *enhancements*, especially as the case for such enhancements is often based on much weaker data. Evolution is a process powerful enough to have led to the development of systems – such as human brains – that are far more complex and capable than anything that human scientists or engineers have managed to design. Surely it would be foolish, absent strong supporting evidence, to suppose that we are currently likely to be able to do *better* than evolution, especially when so far we have not even managed to understand the systems that evolution has designed and when our attempts even just to repair what evolution has built so often misfire!

We believe that these informal considerations contain a grain of truth. Nonetheless, in many particular cases we believe it is practically feasible to improve human nature. The evolution heuristic is our explanation of why this is so. If the evolution heuristic works as we suggest, it shows that there is some validity to the widespread intuition that nature often knows best, especially in relation to proposals for human enhancement. But the heuristic also demonstrates that the validity of this intuition is limited, by revealing important exceptional cases in which we can hope to improve on nature using even our present or near-future science and technology.

The evolution heuristic might be useful for scientists working to develop enhancement technologies. It might also be useful in evaluating beliefs and arguments about the ethics of human enhancement. This is because intuitions about the wisdom of nature appear to play an important role in the cognitive ecology of many anti-enhancement advocates. While sophisticated bioconservatives (aware of the distinction between “is” and “ought”) may not *explicitly* base their arguments on the alleged wisdom in nature, we believe that such intuitions influence their evaluation of the plausibility of various empirical assumptions and mid-level moral principles that are invoked in the enhancement discourse; just as the opinions and practical judgments of the pro-

¹ (McKeown and Lowe 1974)

² (Newhouse and Group. 1993; Frech and Miller 1996; Kirsch, Moore, Scoboria and Nicholls 2002)

³ (Starfield 2000)

⁴ (Bunker 2001)

enhancement transhumanists look more plausible if one assumes that nature is generally unwise. Addressing such hidden empirical background assumptions may therefore help illuminate important questions in applied ethics.⁵

1.2 The evolution heuristic

The basic idea is simple. In order to decide whether we want to modify some aspect of a system, it is helpful to consider why the system has that aspect in the first place. Similarly, if we propose to introduce some new feature, we might ask why the system does not already possess it. The system of concern here is the human organism. The question why the human organism has a certain property can be answered on at least two different levels, ontogeny and phylogeny. Here the focus is on the phylogeny of the human organism.

We can conceive of a proposed enhancement as an ordered pair (α, A) , where α is some specific intervention (e.g. the administration of a drug) and A is the trait that the intervention is intended to realize (e.g. improved memory consolidation). We define an enhancement as an intervention that causes either an improvement in the functioning of some subsystem (e.g. long-term memory) beyond its normal healthy state in some individual or the addition of a new capacity (e.g. magnetic sense).

On this definition, an enhancement is not necessarily desirable, either for the enhanced individual or for society. For instance, we might have no reason to value an enhancement of our sweat glands that increases their ability to produce perspiration in response to heat stimuli. In other instances, we might benefit from increased functionality or a new capacity, and yet not benefit from the enhancement because the intervention also causes unacceptable side effects.⁶ The evolution heuristic is a tool to help us think through whether some proposed enhancement is likely to yield a net benefit.

The starting point of the heuristic is to pose the *evolutionary optimality challenge*:

(EOC) If the proposed intervention would result in an enhancement, why have we not already evolved to be that way?

Suppose that we liken evolution to a surpassingly great engineer. (The *limitations* of this metaphor are part of what makes it useful for our purposes.) Using this metaphor, the EOC can be expressed as the question, “How could we realistically hope to improve on evolution’s work?” We propose that there are three main categories of possible answers, which can be summarized as follows:

- *Changed tradeoffs*. Evolution “designed” the system for operation in one type of environment, but now we wish to deploy it in a very different type of environment. It is not surprising, then, that we might be able to modify the system better to meet the demands imposed on it by the new environment. Making such modifications need not require engineering skills on a par with those of evolution: consider that it is much harder to design and build a car from scratch than it is to fit an existing car with a new set of wheels or make some other tweaks to improve

⁵ On the role of mid-level principles in one area of applied ethics, see (Beauchamp and Childress 1979).

⁶ Which side effects are acceptable depends, of course, on the benefits resulting from the enhancement, and these may vary between subjects depending on their goals, life plans, and circumstances.

functioning in some particular setting, such as icy roads. Similarly, the human organism, whilst initially “designed” for operation as a hunter-gatherer on the African savannah, must now function in the modern world. We may well be capable of making some enhancing tweaks and adjustments to the new environment even though our engineering talent does not remotely approach that of evolution.

- *Value discordance.* There is a discrepancy between the standards by which evolution measured the quality of her work, and the standards that we wish to apply. Even if evolution had managed to build the finest reproduction-and-survival machine imaginable, we may still have reason to change it because what we value is not primarily to be maximally effective inclusive-fitness optimizers. This discordance in objectives is an important source of answers to the EOC. It is not surprising that we can modify a system better to meet our goals, if these goals differ substantially from the ones that (metaphorically might be seen as having) guided evolution in designing the system the way she did. Again, this explanation does not presuppose that our engineering talent exceeds Evolution’s. Compare the case to that of a mediocre technician, who would never be able to design a car, let alone a good one; but who may well be capable of converting the latest BMW model into a crude rain-collecting device, thereby *enhancing* the system’s functionality as a water collecting device.
- *Evolutionary restrictions.* We have access to various tools, materials, and techniques that were unavailable to evolution. Even if our engineering talent is far inferior to evolution’s, we may nevertheless be able to achieve certain things that stumped evolution, thanks to these novel aids. We should be cautious in invoking this explanation, for evolution often managed to achieve with primitive means what we are unable to do with state-of-the-art technology. But in some cases one can show that it is practically impossible to create a certain feature without some particular tool – no matter how ingenious the engineer – while the same feature can be achieved by any dimwit given access to the right tool. In these special cases we might be able to overcome evolutionary restrictions.

In the following three sections, we will explore each of these categories of possible answers to the EOC in more detail, and show how they can help us decide whether or not to go ahead with various potential human enhancements.

Our ideas about enhancement in many ways parallel earlier work in evolutionary medicine. Evolutionary medicine is based on using evolutionary considerations to understand aspects of human health.⁷ Hosts and parasites have adapted to one another, and analysis of the tradeoffs involved can reveal adaptations that contributed to fitness in the past but are maladaptive today, or symptoms that have been misdiagnosed as harmful but may actually aid recovery. Evolutionary medicine also helps explain the incidence of genetic diseases, which can be maintained in the population because of beneficial effects in historically normal environments. Another contribution of evolutionary medicine has been to draw attention to the fact that our modern environment may not always fit a biology designed for Pleistocene conditions, and how this mismatch can cause disease. These insights are recycled in our analysis of human enhancement.

⁷ (Williams and Nesse 1991; Trevathan, Smith and McKenna 1999)

Another strand of research relevant to our aims is evolutionary optimization theory, which seeks to determine the abilities and limitations of evolution in terms of producing efficient biological functions.⁸ While, naively, evolution might be thought to maximize individual fitness (the expected lifetime number of surviving offspring), there are many contexts in which this simplification leads to error. Sometimes it is necessary to focus on the concept of inclusive fitness, which takes into account the effects of a genotype on the fitness of blood-relatives other than direct decedents. Sometimes a gene-centric perspective is needed, to account for phenomena such as segregation distortion and junk DNA.⁹ There are also many other ways in which evolution routinely falls short of “optimality”, some of which we will be covered in later sections.

2. Changed tradeoffs

2.1 General remarks on tradeoffs

Evolutionary adaptation often involves striking a tradeoff between competing design criteria. Evolution has fine-tuned us for life in the ancestral environment, which, for the most part, was a life as a member of a hunter-gatherer tribe roaming the African savannah. Life in contemporary society differs in many ways from life in the environment of evolutionary adaptedness. Modern conditions are too recent for our species to have fully adapted to them, which means that the tradeoffs evolution struck may no longer be optimal today.

In evolutionary biology, the “environment of evolutionary adaptedness” (EEA) refers not to a particular time or place, but to the environment in which a species evolved and to which it is adapted.¹⁰ It includes both inanimate and animate aspects of the environment, such as climate, vegetation, prey, predators, pathogens, and the social environment of conspecifics. We can also think of the EEA as the set of all evolutionary pressures faced by the ancestors of the species over recent evolutionary time – in the case of humans, at least 200,000 years.¹¹ Hunting, gathering of fruits and nuts, courtship, parasites, and hand-to-hand combat with wild animals and enemy tribes were elements of the EEA; speeding cars, high levels of trans fats, concrete ghettos, and tax return forms were not.

The import of this for the evolution heuristic is that even if the human organism were a wonderfully well-designed system for life in the EEA, it may not in all respects be well designed for life in contemporary society. If we can identify specific changes to our environment that have shifted the optimal tradeoff point between competing design desiderata in a certain direction, we may be able to find relatively easy interventions that could “retune” the tradeoff to a point that is closer to its present optimum. Such retuning interventions might be among the low-hanging fruits on the enhancement tree, fruits within reach even in the absence of super-advanced medical technology.

Proposed enhancements aiming to retune altered tradeoffs can often meet the EOC. The new trait that the enhancement gives us might have been maladaptive in the EEA even though it would be adaptive now. Alternatively, the new trait might be

⁸ (Parker and Smith 1990)

⁹ (Dawkins 1976; Williams 1996/1966)

¹⁰ (Hagen 2002)

¹¹ (Hagen 2002)

intrinsically associated with another trait that was maladaptive in the EEA but has become less disadvantageous (or even beneficial) in the modern environment, so that the terms of the tradeoff have shifted. In either case, the enhancement could be adaptive in the current environment without having been so in the EEA, which would explain why we do not have that trait, allowing us to meet the EOC.

We can roughly distinguish two ways in which tradeoffs can change: new *resources* may have become available that were absent, or available only at great cost, in the EEA; or, the *demands* placed on one of the subsystems of the human organism may have changed since we left the EEA. Let us consider these two possibilities in turn and look at some examples.

2.2 Resources

One of the main differences between human life today (for most people in developed countries) and life in the EEA is the abundant availability of food independently of place and season. In the state of nature, food is relatively scarce much of the time, making energy conservation paramount and forcing difficult energy expenditure tradeoffs between metabolically costly tissues, processes, and behaviors. As we shall see, increased access to nutrients suggests several promising enhancement opportunities. We have also gained access to important new non-dietary resources, including improved protection against physical threats, obstetric assistance, better temperature control, and increased availability of information. Let us examine how these new resources are relevant to potential enhancements of the brain and the immune system.

2.2.1 The brain

The human brain constitutes only 2% of body mass yet accounts for about 20% of total energy expenditure. Combined, the brain, heart, gastrointestinal tract, kidneys, and liver consume 70% of basal metabolism. This forces tradeoffs between the size and capacity of these organs, and between allocation of time and energy to activities other than searching for food in greater quantity or quality.¹²

Unsurprisingly, we find that, in evolutionary lineages where nutritional demands are high and cognitive demands low (such as bats hunting in uncluttered environments), relative brain size is correspondingly smaller.¹³ In humans, brain size correlates positively with cognitive capacity (≈ 0.33).¹⁴

Holding brain mass constant, a greater level of mental activity might also enable us to apply our brains more effectively to process information and solve problems. The brain, however, requires extra energy when we exert mental effort, reducing the normally tightly regulated blood glucose level by about 5% (0.2 mmol/l) for short (<15 min) efforts and more for longer exertions.¹⁵ Conversely, increasing blood glucose levels has been shown to improve cognitive performance in demanding tasks.¹⁶

¹² (Aiello, Bates and Joffe 2001; Fish and Lockwood 2003)

¹³ (Niven 2005)

¹⁴ (McDaniel 2005)

¹⁵ (Scholey, Harper and Kennedy 2001; Fairclough and Houston 2004)

¹⁶ (Korol and Gold 1998; Manning, Stone, Korol and Gold 1998; Martin and Benton 1999; Meikle, Riby and Stollery 2005). Increasing oxygen levels (another requirement for metabolism) also improves cognition (Winder and Borrill 1998).

The metabolic problem is exacerbated during prenatal and early childhood growth where brain development requires extra energy. Brain metabolism accounts for a staggering 60% of total metabolism in newborns,¹⁷ exacerbating the competitive situation between mother and child for nutritional resources – an unpleasant tradeoff.¹⁸ Children with greater birth weight have a cognitive advantage.¹⁹

Another constraint on prenatal cerebral development is the size of the human birth canal (itself constrained by bipedalism), which historically placed severe constraints on the head size of newborns.²⁰ These constraints are partly obviated by modern obstetrics and the availability of caesarian section. One way of reducing head size at birth and perinatal energy demands would be to extend the period of postnatal maturation. However, delayed maturation was vastly riskier in the EEA than it is now.

What all this suggests is that cognitive enhancements might be possible if we can find interventions that recalibrate these legacy tradeoffs in ways that are more optimal in the contemporary world. For example, suppose we could discover interventions that moderately increased brain growth during gestation, or slightly prolonged the period of brain growth during infancy, or that triggered an increase in available mental energy. Applying the EOC to these hypothetical interventions, we get a green light. We can see why these enhancements would have been maladaptive in the EEA, and why they may nevertheless have become entirely beneficial now that the underlying tradeoffs have changed as a result of the availability new resources. If the “downside” of getting more mental energy is that we would burn more calories, many of us would pounce at the opportunity.

Not all cognitive enhancement interventions get an immediate green light from the above argument. Stimulants like caffeine and Modafinil enable increased wakefulness and control over sleep patterns.²¹ But sleep serves various (poorly understood) functions other than to conserve energy.²² If the explanation for why we do not sleep less than we do has to do with these other functions, then reducing sleep might well have more problematic side-effects than increasing the amount of calories we need to consume. For any particular intervention, such as the administration of some drug, we also of course need to consider the possibility of contingent side-effects, i.e. that the drug might have effects on the body other than simply retuning the target tradeoff.

2.2.2 The immune system

While the immune system serves an essential function by protecting us from infection and cancer, it also consumes significant amounts of energy.²³ Experiments have found direct energetic costs of immune activation.²⁴ In birds immune activation corresponded to a 29% rise of resting metabolic rate²⁵ and in humans the rate increases by 13% per degree

¹⁷ (Holliday 1986)

¹⁸ (Martin 1996)

¹⁹ (Matte 2001)

²⁰ (Trevathan 1987)

²¹ (Caldwell 2001)

²² (Siegel 2005)

²³ (McDade 2003)

²⁴ (Demas, Chefer, Talan and Nelson 1997; Moret and Schmid-Hempel 2000; Ots, Kerimov, Ivankina, Ilyina and Horak 2001)

²⁵ (Martin, Scheuerlein and Wikelski 2003)

centigrade of fever.²⁶ In addition, the protein synthesis demands of the immune system are sizeable yet prioritized, as evidenced by a 70% increase in protein turnover in children during infection despite a condition of malnourishment.²⁷ One would expect the immune system to have evolved a level of activity that strikes a tradeoff between these and other requirements – a level optimized for life in the EEA but perhaps no longer be ideal.

Such a tradeoff has been proposed as part of an explanation of the placebo effect.²⁸ The placebo effect is puzzling because it apparently involves getting something (accelerated recovery from disease or injury) for nothing (merely having a belief). If the subjective experience of being treated causes a health-promoting response, why are we not always responding that way? Studies have shown that it is possible chemically to modulate the placebo response down²⁹ or up.³⁰

One possible explanation is that mobilizing the placebo effect consumes resources, perhaps through activation of the immune system or other forms of physiological health investment. Also, to the extent that the placebo response reduces defensive reactions (such as pain, stiffness, and inflammation), it might increase our vulnerability to future injury and microbial assaults. If so, one might expect that natural selection would have made us such that the placebo response would be triggered by signals indicating that that in the near future we will (a) recover from our current injury of disease (in which case there is no need to conserve resources to fight a drawn-out infection and less need to maintain defensive reactions), (b) have good access to nutrients (in which case, again, there is no need to conserve resources), and (c) be protected from external threats (in which case there is less need to keep resources in reserve for immediate action readiness). Consistent with this model, the evidence does indeed show that the healing system is activated not only by the expectation that we will get well soon but also by the impression that external circumstances are generally favorable. For example, social status,³¹ success, having somebody looking after us,³² sunshine, and regular meals might all indicate that we are in circumstances where it is optimal for the body to invest in healing and long-term health, and they do seem to prompt the body to do just that. By contrast, conflict,³³ stress, anxiety, uncertainty,³⁴ rejection, isolation, and despair appear to shift resources towards immediate readiness to face crises and away from building long-term health.

If this model of the placebo response is correct, several potential avenues of enhancement are worth exploring. One is that since physical safety and reliable access to food are much improved compared to the EEA, it might now be beneficial to invest more in biological processes that build long-term health than was usually optimal in the EEA. We might thus inquire whether the placebo effect and other evolved responses are flexible enough to have adjusted the level of health investment to a level that is optimal under modern conditions. If not, we could benefit from an intervention that triggers a placebo-like response or otherwise increases the body's health investment.

²⁶ (Elia 1992)

²⁷ (Waterlow 1984; McDade 2003)

²⁸ (Humphrey 2000)

²⁹ (Sauro and Greenberg 2005)

³⁰ (Colloca and Benedetti 2005)

³¹ (Sapolsky 2005)

³² (House, Landis and Umberson 1988)

³³ (Kiecolt, Glaser, Cacioppo, MacCallum, Snyder-Smith, Kim and Malarkey 1997)

³⁴ (McDade 2002)

However, while external stresses and resource constraints are reduced in the modern environment, the danger of autoimmune reactions remains. We would therefore have to be careful not to overshoot the target. It is possible that we would benefit from a *lower* baseline immune activity in some parts of the immune system since we are now less at risk of dying from infectious diseases. As an example, the hygiene theory of allergic diseases claims that the reduction in immunological challenge in particular from helminth parasites during early life increases the risk of allergic disease later in life.³⁵ If true, then a down-regulation of a particular dendritic cell subpopulation (DC2) sensitive to helminthes but causing allergic reactions might be desirable. Alternatively, an up-regulation of regulatory (DCreg) cells that tend to be lost in unstimulated immune systems might be used to control the DC2 cells.

The evolution heuristic also leads us to consider other potential immune system enhancements. Even if the average activation level of our immune systems were still optimal in the modern era, we now possess more information (a new resource) about the detailed requirements in specific situations. We can use this information to override our bodies' natural response tendencies. For example, recipients of donated organs can benefit from immunosuppressant drugs. Conversely, a patient with early-stage cancer might be better off if her immune system could be induced to mount an immediate all-out assault on the incipient tumor instead of conserving resources in for hypothetical future challenges.³⁶

A more radical enhancement would be to improve DNA repair, which would reduce cancer-causing mutations and improve radiation resistance, at the price of increasing metabolic needs. The modification could be achieved through overexpression of existing DNA repair genes³⁷ or perhaps even by transgenic incorporation of the unique abilities of *Deinococcus radiodurans*.³⁸ Increased repair would have to be balanced with apoptosis and replacement of irreparably damaged cells (another energy cost). Until recently, increased DNA repair activity might have been too metabolically costly and mutation-prone for evolution to consider it a worthwhile bargain. One of the most well studied pathways, the PARP-1 pathway, protects the genome from damage but requires so much energy that it can damage cells through energy depletion.³⁹

Since the objective of the interventions suggested above is to restore health, one could argue that they should be regarded as therapeutic rather than enhancing. But these classifications are not necessarily incompatible. We could regard the interventions as therapeutic for the subsystems whose functioning has been deteriorated by disease, yet enhancing for the immune system, whose functioning is improved beyond its normal state.⁴⁰

³⁵ (Yazdanbakhsh, Kremsner and van Ree 2002; Maizels 2005)

³⁶ (Boon and van Baren 2003; Dunn, Old and Schreiber 2004)

³⁷ (Wood, Mitchell, Sgouros and Lindahl 2001)

³⁸ (Battista, Earl and Park 1999; Venkateswaran, McFarlan, Ghosal, Minton, Vasilenko, Makarova, Wackett and Daly 2000)

³⁹ (de Murcia and Shall 2000; Skaper 2003).

⁴⁰ In like manner, we can view vaccinations as both therapeutic (or more accurately, prophylactic) and as enhancing.

2.3 Demands

Just as we have many resources that were denied our hunter-gatherer ancestors, we also face a different set of demands than they did. This suggests further opportunities for enhancement.

Changes in demands on the human organism occur when old demands disappear or are reduced (e.g. less need for long treks to get food; hygienic surroundings reducing demands on the immune system), and when demands grow in strength or new demands arise (e.g. greater need to be able to concentrate on abstract material for long periods; new pathogens spreading in larger societies). The source and nature of a particular demand may also change. For instance, exercise is no longer necessary to gain sustenance, but is instead needed to maintain the body in good shape.

Many “diseases of civilization” are due to these changed demands. For example, our ancestors needed to exert themselves physically to secure adequate nutrition, whereas our easy access to abundant food can lead to obesity. People working indoors do not get the sun exposure that our ancestors had, leading to vitamin D deficiency;⁴¹ yet we risk skin cancer when we expose pale skin to the sun during occasional recreational activities. Rapid blood coagulation was beneficial in the past, when there was a high risk of wounding. The increased risk for cardiovascular problems and embolisms was an acceptable tradeoff. Today, the risk of wounding has sharply decreased, making the downsides relatively more important. Reducing coagulation, e.g. by taking low-dose aspirin, can be beneficial given these changed demands,⁴² although we risk incidental side-effects such as stomach irritation.

While the change in demands can cause or exacerbate problems, it can also alleviate them. The recent emergence of the IT industry appears to have produced a refuge for people with Asperger’s syndrome where their preference for structure and detail becomes a virtue and their problems with face-to-face communication less of a disadvantage.⁴³ Deliberate fitting of environments to human evolutionary adaptations and individual idiosyncrasies is a promising adjunct to direct human enhancement for improving human performance and wellbeing.

2.3.1 Literacy and numeracy

Intellectual capacity, or at least some specific forms of it, seem to have become more rewarded in contemporary society than they were in the EEA. There is a positive correlation in Western society between IQ and income.⁴⁴ Higher levels of general cognitive ability are important not just for highly demanding high status jobs, but also for success in everyday life, such as being able to fill out forms, understand news, and maintain health. As society becomes more complex, these demands increase, placing people of low cognitive ability at a greater disadvantage.⁴⁵ While general cognitive ability

⁴¹ (Thomas, Lloyd-Jones, Thadhani, Shaw, Deraska, Kitch, Vamvakas, Dick, Prince and Finkelstein 1998)

⁴² (Force 2002)

⁴³ (Silberman 2001)

⁴⁴ (Neisser, Boodoo, Bouchard, Boykin, Brody, Ceci, Halpern, Loehlin, Perloff, Sternberg and Urbina 1996; Gottfredson 1997; Bersaglieri, Sabeti, Patterson, Vanderploeg, Schaffner, Drake, Rhodes, Reich and Hirschhorn 2004)

⁴⁵ (Gottfredson 1997; Gottfredson 2004)

may have been advantageous (and selected for) in our evolutionary past,^{46,47} numeracy and literacy represent more specific abilities whose utility has increased dramatically in recent times.

Before the invention of writing, the human brain faced no pressure to be literate. In the current age, however, literacy is in very high demand. Failing to meet this demand places an individual at a severe disadvantage in modern society. Since writing is a relatively recent invention (3,500 BC), and since it is even more recently that written language has become such a dominant mode of communication, it is plausible that the human brain is not optimized for modern conditions. The fact that the neural machinery needed for writing and reading largely overlaps with that needed to produce and interpret oral communication means that the mismatch between evolved capacity and present demands is not as great as it might have been. Nevertheless, as the phenomenon of dyslexia demonstrates, it is possible to have deficits in language processing that are relatively specific to written language, possibly arising from minor variations in phonological processing.⁴⁸ Dyslexia also appears to be linked to enhanced or atypical visuospatial abilities.⁴⁹ These abilities might have been useful in the EEA, but today literacy is usually more important for achieving life goals. If our species had been using written language for a couple of million years and reproductive fitness had depended on literacy, dyslexia might have been much rarer than it is.

Modern society also places much greater demands on advanced numerical skills than we faced in the EEA. In hunter-gatherer societies, numeracy demands appear to have been limited to being able to count to five or ten.⁵⁰ In the modern world, one is at a major disadvantage if one cannot understand at least basic arithmetic. Many occupations require a grasp of statistics, calculus, geometry, or higher mathematics. Programming skills open up additional employment possibilities. Good logical and analytical skills create further opportunities in our information-dense, technology-mediated, and generally formalized modern society. These skills were much less useful in the Pleistocene.

The altered nature of the demands we face suggests opportunities for enhancement by readjusting tradeoffs that are no longer optimal. For example, number relations appear to be handled by brain circuits closely linked to spatial cognition of external objects, and affected by spatial attention abilities.⁵¹ Hence enhancement of this type of spatial attention,⁵² possibly at the expense of remote or peripheral attention, could be a useful enhancement. Similarly, enhancements in reading ability at the expense of the dyslexia-related visuospatial abilities might gain support from the EOC.

2.3.2 Concentration

The importance of being able to concentrate on abstract thinking and tasks with little sensory feedback has increased significantly in modern times relative to the importance of

⁴⁶ (Gottfredson 2007)

⁴⁷ It should be noted that IQ correlates negatively with fertility in many modern societies (Udry 1978; Vancourt and Bean 1985; Vining, Bygren, Hattori, Nystrom and Tamura 1988) This might be an example of value discordance between human values and evolutionary fitness.

⁴⁸ (Goulandris, Snowling and Walker 2000)

⁴⁹ (von Karolyi, Winner, Gray and Sherman 2003; Brunswick, Martin, Marzano and Savill 2007)

⁵⁰ (Pica, Lemer, Izard and Dehaene 2004)

⁵¹ (McCord 2000; Hubbard, Piazza, Pinel and Dehaene 2005)

⁵² (Green and Bavelier 2006)

peripheral awareness. In the EEA, peripheral awareness was crucial for detecting predators and enemies, while an ability to exclude other stimuli had few applications. We may hence have evolved attention systems with a tendency to be too easily distracted in a modern setting. It has been suggested that ADHD is a form of “response-readiness” that was more adaptive in past environments.⁵³ Concentration enhancers may therefore be feasible and promising in modern settings, enabling users to meet high demands for sustained attention. Drugs such as methylphenidate (Ritalin) are already used to treat ADHD and occasionally also for enhancement purposes.⁵⁴

2.3.3 Dietary preferences and fat storage

One tradeoff involving food availability relates to the question of how much nutrition the body should store in fatty deposits. If high-calorie foods are scarce and food availability highly variable, it is optimal for an individual to crave high-calorie foods and to store lots of energy in fat deposits as insurance against lean times. We still need an appetite today, and we still need fat deposits, but – at least in the developed world – they are much less important now than in the past. Many people’s natural set-points of appetite and body fat are higher than optimal, leading to increased morbidity. In wealthy modern societies, where a Mars bar is never far away, the risks of obesity and diabetes outweigh the risk of under-nutrition,⁵⁵ and a sweet tooth is maladaptive.

This suggests that it might be possible to enhance human health by finding effective ways to down-regulate our cravings for fat and sugar, or by reducing the absorption and storage of these calories in fatty tissues. Such an enhancement might take various forms: nutritional advice, diet pills, artificial sweeteners, indigestible substances that taste like fat, weight-loss clubs, hypnotherapy, and, in the future, gene therapy. The evolution heuristic suggests that our natural proclivities to consume and store nutrients might be a case where we could benefit from going against the wisdom of nature. Independent considerations and possibly further research would be needed to determine the most effective way of doing this, given that weight loss itself is a longevity risk factor⁵⁶ and that those who are mildly overweight have lower mortality than those who are underweight or obese.⁵⁷ Possibly an aversion to unhealthy foods and eating habits would be more effective and safer than a general down-regulation of appetite. The heuristic tells us only that there are no general “wisdom of nature” reasons to retain our current bodyweight set-points; it does not by itself tell us which approaches to changing them would be safest.

2.4 The interplay between resources and demands

The picture is complicated by the fact that some phenomena zigzag across the two subcategories of changed tradeoffs (resources and demands). Transport vehicles and machinery are new resources that reduce the demand for physical exertion. The effect is that most of us get less exercise in the course of our daily routines. Yet our bodies appear

⁵³ (Jensen, Mrazek, Knapp, Steinberg, Pfeffer, Schowalter and Shapiro 1997)

⁵⁴ (Farah, Illes, Cook-Deegan, Gardner, Kandel, King, Parens, Sahakian and Wolpe 2004)

⁵⁵ (Fontaine, Redden, Wang, Westfall and Allison 2003)

⁵⁶ (Gaesser 1999)

⁵⁷ (Flegal, Graubard, Williamson and Gail 2005)

to be designed for physical activity, so a sedentary life causes a variety of health problems. New resources (gyms, exercise equipment, parks, jogging clubs) have been developed to help us overcome the problems of a sedentary lifestyle. But now a new demand arises: we need the energy and self-motivation to make use of these resources – a demand that many find it difficult to meet.

In a case like this, there are multiple potential intervention points where a change could result in an improvement of our lives. One approach would be to design our environment in such a way as to force us to be more physically active. Elevators could be removed, motor vehicles banned from certain areas, and so forth. Another approach would be to attempt to redesign our bodies so that they would not be dependent on frequent physical exertion to remain healthy. On this approach, we might try to develop pharmaceuticals that trigger effects in the body similar to those normally caused by exercise (such as the IGF-1/MGF signaling pathways, which are stimulated by exercise or muscle damage⁵⁸). Yet another approach would be to attempt interventions that increase our energy and self-motivation, thereby making it easier for us to exercise on our own initiative. For instance, there might be pharmaceuticals that would give us more energy or strengthen our willpower, or perhaps a habit of regular workouts instilled in childhood would carry over into adult life.

Whether any of these interventions will work, and, if so, which one would be the most effective and have the best balance of benefits over burdens, cannot be determined *a priori*. This is an empirical question, whose answer may depend on changing social circumstances, levels of technology, personal preferences, and other factors. One should note that it is not only biological interventions which can have undesirable side-effects. Removing elevators might cause some health benefits for people forced to climb the stairs, but it may also deny access for people with mobility impairments and cause unnecessary inconvenience to others. Encouraging high levels of physical activity in children might have overall health benefits but it might also lead to more injuries, more worn-out knees and hip joints later in life, and less time for non-physical activities.

Another illustration of the complex interplay between new resources and new demands is offered by the case of addictive drugs. Alcohol, heroin, and crack cocaine are comparatively novel resources. The availability of these resources create a new demand on the human organism: the ability to avoid becoming addicted to harmful drugs that hijack the brain's reward system. Individuals vary in how they metabolize these drugs and how their brains react to exposure. Again, the solution might be to develop new resources (e.g. detox clinics), temporary pharmacological interventions (methadone), permanent biological modifications (vaccines), educational initiatives (drug awareness programs), or social policies (criminalization). Alternatively, one might attempt to develop safer, non-addictive substitutes for harmful drugs.⁵⁹ There are many possible ways to defy or to work around the wisdom of nature.

⁵⁸ (Baldwin and Haddad 2002; Goldspink 2005)

⁵⁹ (Nutt 2006)

3. Value discordance

3.1 General remarks on value discordance

We have discussed opportunities for enhancement arising from the changed tradeoffs we face in the modern world compared to those of the EEA. (A great engineer built a system for use in a certain environment; we adapt it for use in a different environment.) In this section, we discuss another source of enhancement opportunities: the discordance between evolutionary fitness and human values. (A great engineer built a system that efficiently serves one purpose; we tinker with it to make it serve a different purpose.)

While our goals are not identical to those of Evolution, there is considerable overlap. We value health, and health increases inclusive fitness. We value good eyesight, and good eyesight is useful for survival. We value musicality and artistic creativity, and these talents helped to attract mates in the EEA. If we are hoping to enhance some attribute for which the concordance in objectives is perfect, the present category will not give any help in meeting the EOC. We then either have to find an answer from one of the other categories or else suspect that what appears to be an easy enhancement will in fact come at a large hidden cost.

Whilst some of our traits are both valuable to us and conducive to fitness, many attributes that we value would either not have promoted inclusive fitness in our natural environment, or else would not have been fitness-promoting to a sufficient extent to result in a profile of traits that is optimal from the perspective of our own values. There is a plethora of capacities or characteristics to which we assign a value that exceeds the contribution these characteristics made to survival and reproduction.

One obvious example is contraceptive technology. Vasectomy, birth control pills, and other contraceptive methods enhance our control over our reproductive systems, severing the link between sex and reproduction. We may value such enhancements because they make family planning easier and increase choice. But Evolution would frown on these practices. The great engineer would not regard the absence of an easy reproductive off-switch as a defect. When our goals differ from hers, it is unsurprising that we are able to modify her design in ways that make it better (by our lights) even if our design skills fall far short of hers.⁶⁰

We can distinguish (at least) two distinct sources of such value discordance. The first is that the characteristics that would maximize an individual's inclusive fitness are not always identical to the characteristics that would be best for her. The other is that the characteristics that would maximize an individual's inclusive fitness are not always identical to those that would be best for society, or impersonally best. If our goal is to identify potential interventions that individuals would have prudential reasons for wanting, then we may perhaps set aside the second source of value discordance. If, however, we are interested in addressing ethical and public policy matters, then it is relevant to consider value discordance arising from either of these two sources. Let us consider each in turn.

⁶⁰ Evolution might still have the last laugh if in the long-run she redesigns our species to directly desire to have as many children as possible, or to have an aversion against contraceptives. Cultural "evolution" might beat biological evolution to the punch.

3.2 Good for the individual

What characteristics promote individual well-being? There is a vast ethical and empirical literature on this question, which we shall not attempt to review here. For our purposes, it will suffice to list (table 1) some candidate characteristics, ones which may with some plausibility be taken to be among those that contribute to individual well-being in a wide range of circumstances. This list is for illustration only. Other lists could be substituted without affecting the structure of our argument.⁶¹

Table 1: Some traits that may promote individual well-being

- | |
|--|
| <ul style="list-style-type: none">• Emotional well-being• Freedom from severe or chronic pain• Friendship and love• Long-term memory• Mathematical ability• Awareness and consciousness• Musicality• Artistic appreciation and creativity• Literary appreciation• Confidence and self-esteem• Healthy pleasures• Mental energy• Ability to concentrate• Abstract thinking• Longevity• Social skills |
|--|

To illustrate the idea, take mathematical ability. Suppose that we believe that having greater mathematical ability would tend to make our lives go better – perhaps because it would give us competitive advantages in the job market, perhaps because appreciating mathematical beauty is a value in itself, or perhaps because we believe that mathematical ability is linked to other abilities that would increase our well-being. We then pose the EOC: Why has evolution not already endowed us with more mathematical ability than we have?

It is possible that answers to this EOC may be found in the other categories we discuss in this paper (changed tradeoffs or evolutionary restrictions). Yet suppose that is not so. We may then appeal to an answer in the value discordance category. Even if greater mathematical capacity would have been maladaptive in the EEA and even if it would still be maladaptive today, it may nevertheless be good for us, because the good for humans is different from what maximizes our fitness.

But we are not yet done. What the evolution heuristic teaches us in this case is that we must expect that the intervention will have some effect that reduces fitness. If we

⁶¹ The items in the list need not be final goods. Characteristics that are mere *means* to more fundamental goods can be included. For example, even if one thinks that musicality or musical appreciation is not intrinsically good, one can still include them in the list if one believes that they tend – as a matter of empirical fact – to promote well-being (for example, by creating opportunities for enjoyment).

cannot form any plausible idea of what sort of effect the intervention might produce that would reduce fitness, then we must suspect that the intervention will have important effects that we have not understood. That should give us pause. A fitness-reducing effect that we have not anticipated might be something very bad, such as a serious medical side-effect. The EOC hoists a warning flag. If, however, we can give a plausible account of why the proposed intervention to increase mathematical ability would reduce fitness, *and yet we judge this fitness-reducing effect as desirable or at least worth enduring for the sake of the benefit*, then we have met the EOC.

This does not guarantee that the enhancement will succeed. It is still possible that the intervention will fail to produce the desired result or that it would have some unforeseen side-effect. There might be more than one sufficient reason why evolution did not already make this intervention to enhance mathematical ability. But once we have identified at least one sufficient reason, the warning flag raised by the EOC comes down. We have shown that one potential reason for thinking that the enhancement will fail (the “wisdom of nature” reason) does not apply to the present case.

As an example, evolution has not optimized us for happiness and has instead led to a number of adaptations that cause psychological distress and frustration.⁶² The “hedonic-treadmill” causes us quickly to adapt to positive experiences and to seek more, as goods we have gained become taken for granted as a new status quo.⁶³ Sexual jealousy, romantic heartaches, status envy, competitiveness, anxiety, boredom, sadness and despair may have been essential for survival and reproductive success in the EEA, but they take a toll in terms of human suffering and may substantially reduce our well-being. An intervention that caused an upward shift in hedonic set-point, or that down-regulated some of these negative emotions, would hence meet the EOC: we can see why the effect would have been maladaptive in the EEA, and yet believe that we would benefit from these effects because of a discordance between inclusive fitness and individual well-being.

3.3 Good for society

Many characteristics that promote individual well-being also promote the social good, but the two lists are unlikely to be identical. Table 2 lists some candidate traits that might contribute to the good of society.

Table 2: Some traits that may promote the social good

- | |
|---|
| <ul style="list-style-type: none"> • Extended altruism • Conscientiousness and honesty • Modesty and self-deprecation • Originality, inventiveness, and independent thinking • Civil courage • Knowledge and good judgment about public affairs • Empathy and compassion • Nurturing emotions and caring behavior • Just admiration and appreciation |
|---|

⁶² (Buss 2000)

⁶³ (Diener, Suh, Lucas and Smith 1999)

- Self-control, ability to control violent impulses
- Strong sense of fairness
- Lack of racial prejudice
- Lack of tendency to abuse drugs
- Taking joy in others' successes and flourishing
- Useful forms of economic productivity
- Healthy longevity

As with the list for individual well-being, this one is for illustration only. One could create alternative lists for various related questions, such as traits that are good for humanity as a whole, or for sentient life, or for a particular community, or traits that specifically help us become better moral agents. While the lists may overlap, they will likely disagree about some characteristics or their relative importance. The evolution heuristic can be applied using any such list as input.

To use such a list with the EOC, we proceed in the same way as with the “good for the individual” source of value discordance. For example, we might have a drug that appears to make those who take it more compassionate. This might seem like a good thing, but why has evolution not already made us more compassionate? Presumably, evolution could easily have produced an endogenous substance with similar effects to the drug; so the likely explanation is that a higher level of compassionateness would not have increased inclusive fitness in the EEA. We may press on and ask *why* it is that greater compassionateness would have been maladaptive in the EEA. One may surmise that such a trait would have been associated with evolutionary downsides – such as reduced ability credibly to threaten savage retaliation, a tendency to spare the lives of enemies allowing them to come back another day and reverse their defeat, an increased propensity to offer help to those in need beyond what is useful for reciprocity and social acceptance, and so forth. But these very effects, which would have made heightened compassionateness maladaptive for an individual in the EEA, are precisely the kinds of effects which we might believe would be beneficial for the common good today. We do not have to assume that the relevant trade-offs have changed since the EEA. Even in the EEA, it might have had net good effects for a local population of hunter-gathers if one person was born with a mutation causing an unusually high level of compassionateness, even though that individual himself might have suffered a fitness penalty. If we accept these premises, then the hypothetical drug that increases compassionateness would pass the EOC. It would be a case where we have reason to think that the wisdom of nature has not achieved what would be best for society and that we could feasibly do better.

4. Evolutionary restrictions

4.1 General remarks on evolutionary restrictions

The final category of answers to the EOC focuses on the fact that there are certain limitations in what evolution can do. Using the “great engineer” metaphor, we may say that we can hope to achieve certain things with our ham-handed tinkering that stumped Evolution, because we have access to tools, materials, and techniques that the great ingenious engineer lacked.

Metaphors aside, we can identify several restrictions of evolution's ability to achieve fitness-maximizing phenotypes even in the EEA. These are important, because in some cases they will indicate clear limitations in the "wisdom of nature", and *a fortiori* cases where there is room for potentially easy improvements. At a high level of abstraction, we can divide these restrictions into three classes:

- *Fundamental inability*: evolution is fundamentally incapable of producing a trait *A*
- *Entrapment in local optimum*: evolution is stuck in a local optimum that excludes trait *A*
- *Evolutionary lag*: evolution of trait *A* takes so many generations that there has not yet been enough time for it to develop

These three classes, which are discussed in more detail in the following three subsections, are not sharply separate. For example, one reason why a trait may take a vast number of generations to develop is that it requires escaping from one or more local optima. And given truly astronomical time scales, even some traits that we shall regard as fundamentally beyond evolution's reach might conceivably have evolved. However, the three classes are distinct enough to deserve individualized attention.

4.2 Fundamental inability

Biology is limited in what it can build. DNA can only code for proteins, which have to act on moieties in a water-based cellular environment using the relatively weak chemical forces that a protein can muster. This makes it very unlikely that any terrestrial organism could produce diamond, for instance, since the synthesis of diamondoid structures requires significant energy.⁶⁴ And while bacteria can produce microscopic metal crystals,⁶⁵ there is no way to unite them into contiguous metal. Hence evolution cannot achieve diamond tooth enamel or a titanium skeleton, even if these traits would have improved fitness.

Examples can be multiplied. It is unlikely that evolution could have evolved high-performance silicon chips to augment neural computation, even though such augmentations might have provided important benefits. A theoretical design of artificial red blood cells has been published, calculating the performance of a potentially feasible physical structure for transporting oxygen and carbon dioxide in the blood.⁶⁶ This design, which is not limited by the materials and pressures that can be achieved using biology, would enable performance far outside the range of natural red blood cells.

Radical departures from nature are apt to raise a host of separate questions regarding biocompatibility and functional integration with evolved systems. But at least there is no mystery as to why we would not already have evolved these enhancements even if they would have increased inclusive fitness in the EEA.

⁶⁴ Adding a carbon dimer to a diamond surface using a nanotechnological tool would take more than 6.1 eV (Merkle and Freitas 2003), about 20 times more energy than is released by the ATP hydrolysis that powers most enzymatic actions.

⁶⁵ (Klaus, Joerger, Olsson and Granqvist 1999)

⁶⁶ (Freitas 1998)

Enhancements that evolution is fundamentally incapable of producing can therefore meet the EOC. When invoking “fundamental inability”, it is important to determine that the inability does not pertain merely to the specific means one intends to use to effect the enhancement. If evolution would have been able to employ some different means to achieve the same effect, the challenge would remain to explain why evolution has not achieved the enhancement using that alternative route.

4.3 Entrapment in local optimum

Evolution sometimes get stuck on solutions that are locally but not globally optimal. A locally optimal solution is one where any small change would make the solution worse, even if some big changes might make it better.

Being trapped in a local optimum is especially likely to account for failure to evolve polygenic traits that are adaptive only once fully developed and incur a fitness penalty in their intermediary stages of evolution. In some cases, the evolution of such traits may require an improbable coincidence of several simultaneous mutations that might simply not have occurred among our finite number of ancestors. A crafty genetic engineer may be able to solve some of the problems that were intractable to blind evolution. A human engineer can think backwards, starting with a goal in mind, working out what genetic modifications are necessary for its attainment.

The human appendix, a vestigial remnant of the caecum in other mammals, whilst having some limited immunological function,⁶⁷ easily becomes infected. In the natural state appendicitis is a life-threatening condition, and is especially likely to occur at a young age. There is also evidence that surgical removal of the appendix reduces the risk of ulcerative colitis.⁶⁸ It appears that removal of the appendix would have increased fitness in the EEA. However, *a smaller* appendix increases the risk of appendicitis. Carriers of genes predisposing for small appendices have higher risks of appendicitis than non-carriers, and, presumably, lower fitness.⁶⁹ Therefore, unless evolution could find a way of doing away with the appendix entirely in one fell swoop, it might be unable to get rid of the organ; whence it remains. An intervention that safely and conveniently removed it might be an enhancement, increasing both fitness and quality of life.

Another source of evolutionary lock-in is antagonistic pleiotropy, referring to a situation in which a gene affects multiple traits in both beneficial and harmful ways. If one trait is strongly fitness-increasing and the other mildly fitness-decreasing, the overall effect is positive selection for the gene.⁷⁰ One example is the $\epsilon 4$ allele of apolipoprotein E. Having one or two copies of the allele increases the risk of Alzheimer disease in middle age but lowers the incidence of childhood diarrhea and may protect cognitive development.⁷¹ Antagonistic pleiotropy has also been discussed in relation to theories of ageing. The local optimum here is to retain the genes in question, but the global optimum would be to eliminate the antagonistic pleiotropy by evolving genes that specifically produced the beneficial traits without detrimental effects on other traits. Over longer

⁶⁷ (Fisher 2000)

⁶⁸ (Koutroubakis and Vlachonikolis 2000; Andersson, Olaison, Tysk and Ekblom 2001)

⁶⁹ (Nesse and Williams 1998)

⁷⁰ (Leroi, Bartke, De Benedictis, Franceschi, Gartner, Gonos, Fedei, Kivisild, Lee, Kartaf-Ozer, Schumacher, Sikora, Slagboom, Tatar, Yashin, Vijg and Zwaan 2005)

⁷¹ (Oria, Patrick, Zhang, Lorntz, Costa, Brito, Barrett, Lima and Guerrant 2005)

timescales, evolution usually gets around antagonistic pleiotropy, for example by evolving modifier genes that counteract the negative effects,⁷² but such developments can take a long time and in the meanwhile a species remains trapped in a local optimum.

Yet another way in which evolution can get locked into a suboptimal state is exemplified by the phenomenon of heterozygote advantage. This refers to the common situation where individuals who are heterozygous for a particular gene (i.e. have two different alleles of that gene) have an advantage over homozygote individuals (who have two identical copies of the gene). Heterozygote advantage is responsible for many cases of potentially harmful genes being maintained at a finite frequency in a population.

The classic example of heterozygote advantage is sickle-cell gene, where homozygote individuals suffer anemia while heterozygote individuals benefit from improved malaria resistance⁷³. Heterozygotes have greater fitness than both types of homozygote (those lacking the sickle-cell allele and those having two copies of it). Balancing selection preserves the sickle-cell gene in populations (at a frequency that varies geographically with the prevalence of malaria). The “optimum” that evolution selects is one in which, by chance, some individuals will be born homozygous for the gene, resulting in sickle-cell anemia, a potentially fatal blood disease. The “ideal optimum” – everybody being heterozygous for the gene – is unattainable by natural selection because of Mendelian inheritance, which gives each child born to heterozygote parents a 25% chance of being born homozygous for the sickle-cell allele.

Heterozygote advantage suggests an obvious enhancement opportunity. If possible, the variant allele could be removed and its gene product administered as medication. Alternatively, genetic screening could be used to guarantee heterozygosity, enabling us to reach the ideal optimum that eluded natural selection.

The phenomenon of heterozygote advantage points to potential enhancements beyond reducing susceptibility to diseases such as malaria and sickle-cell anemia. For instance, there is some indirect evidence that at least Type I Gaucher’s Disease (and possibly other sphingolipid storage diseases) is linked to improved cognition, given the significantly higher proportion of sufferers in occupations correlated with high IQ.⁷⁴ This, and other circumstantial evidence, is used by the authors of the cited study to argue that heterozygote advantage can explain the high IQ test scores and the high prevalence of Type I Gaucher’s Disease among Ashkenazi Jews. Should this prediction be borne out by finding an IQ advantage for heterozygote carriers of the diseases, it would suggest that screening to promote heterozygosity, or genetic interventions to induce it, would be viable forms of cognition enhancement that meet the EOC.

One other kind of evolutionary entrapment is worth noting here, that of an evolutionarily stable strategy (ESS), “a strategy such that, if all the members of a population adopt it, no mutant strategy can invade”.⁷⁵ One way in which a species can become trapped in an ESS is through sexual selection. In order to be successful at wooing peahens, peacocks have to produce extravagant tails which serve to advertise the male’s genetic quality. Only healthy peacocks can afford to produce and carry top-notch tails. It is adaptive for peahens to prefer to mate with peacocks that sport an impressive tail; and given this fact, it is also adaptive for peacocks to invest heavily in their plumage. It is

⁷² (Hammerstein 1996)

⁷³ (Allison 1954; Cavalli-Sforza and Bodmer 1999)

⁷⁴ (Cochran, Hardy and Harpending 2006)

⁷⁵ (Smith 1982)

likely that the species would have been better off if it had evolved some less costly way for males to signal fitness. Yet no individual peacock or peahen is able to defect from the ESS without thereby removing themselves from the gene pool. If there had been a United Nations of the peafowl, through which the birds could have adopted a coordinated millennium plan to overcome their species' vanity, the peacocks would surely soon be wearing a more casual outfit.

The concept of an ESS can be generalized to that of an evolutionarily stable state. A population is said to be in an evolutionarily stable state if its genetic composition is restored by selection after a disturbance, provided the disturbance is not too large.⁷⁶ Such a population can be genetically monomorphic or polymorphic. Thus, while ESS refers to a specific strategy that is stable if everybody adopts it, an evolutionary stable state can encompass a set of strategies whose distribution is stable under small perturbations. It has been suggested that the human population has been in a stable state in the EEA with regard to sociopathy, which can be seen as a defector strategy which can prosper when it is rare but becomes maladaptive when it is more common.⁷⁷

Another way in which evolution can fail to produce solutions that are fitness-maximizing for organisms is intragenomic conflict, in which phenomena such as meiotic drive, transposons, homing endonuclease genes, B-chromosomes, and plasmids result from natural selection among lower-level units such as individual genes.⁷⁸ In cases where we can identify intragenomic conflict as responsible for a suboptimal outcome, there is an opportunity for enhancement that can meet the EOC (provided we have the technological means to make the requisite interventions). Genes or traits that would not have evolved, or which would not have been stable against intragenomic competition, could be inserted, possibly supported by interventions removing some of the competing genetic elements.

4.4 Evolutionary lag

Evolution takes time – often, a long time. If conditions change rapidly, the genome will lag. Given that conditions for humanoid ancestors were quite variable – due to migration into new regions, climate change, social dynamics, advances in tool use, and adaptation in pathogens, parasites, predators, and prey – our species has never been perfectly adapted to its environment. Evolution is running up fitness slopes, but when the fitness landscape keeps changing under its feet, it may never reach a peak. Even if beneficial alleles or allele combinations exist, they may not have had the time to diffuse across human populations. For some proposed enhancements, evolutionary lag can therefore provide an answer to the EOC.

This source of answers to the EOC is related to the changed tradeoffs category, but with the difference that here we are focusing on ways in which even during the EEA we were not perfectly adapted to our environment. Even if we set aside the dramatic ways in which resources and demands have changed since the introduction of agriculture, there may still be instances of earlier evolutionary lags that have not yet been truncated and which may point to opportunities for enhancement.

⁷⁶ (Smith 1982)

⁷⁷ (Mealey 1995)

⁷⁸ (Burt and Trivers 2006)

There are many factors limiting the speed of evolution.⁷⁹ Some are inherent in the process itself, such as the mutation rate, the need for sufficient genetic diversity, and the constraint that selection can only encode a few bits into the genome per generation.⁸⁰ A recessive beneficial mutation will spread to an appreciable fraction of a fixed well-mixed population in time inversely proportional to its selective advantage. For example, if the mutation gives a 0.1% increase in fitness, it will take 9,200 generations (230,000 years assuming 25 years per generation) to reach 50% of the population from a starting level of 0.01%. For a 10% fitness-advantage, just 92 generations (2,300 years) are needed⁸¹. Population structure and especially low-population bottlenecks can accelerate the spread significantly.

In nature, the strength of selection for a trait is generally quite weak. A review of published studies⁸² found the distribution of selection strengths across species to be exponential, with a small median magnitude: for most traits and in most systems directional selection is fairly weak. Selection via survival appears to be weaker than selection through mating success, making sexual selection a big factor. Quadratic selection gradients, indicating the “sharpness” of fitness peaks, were also found to be exponentially distributed and with small median. This implies that stabilizing selection (reducing genetic diversity once a population has reached a local fitness peak) is often fairly weak. Indirect selection (where trait fitness depends on another correlated trait) also appears to be playing only a minor role.⁸³ These results suggest that beneficial new traits are likely to spread slowly.

A population living in a heterogeneous or changeable environment may not be able to converge on a single fitness peak but will be spread out around it. This might reduce extinction risks for the lineage, since there will always be some individuals that are well adapted if the conditions change and the lineage will survive more easily than if a less dispersed population had to ascend the current gradient towards the top through a region of low survivability.

It is possible to detect empirically the presence of genetic variations under positive fitness pressure through their signatures.⁸⁴ These signatures range from multimillion year timescale changes in gene sequence (mostly useful to point out ongoing or recurrent selection), to changes in genetic diversity caused by the rapid spread of a beneficial mutation in the past 250,000 years, to the differences between human populations which can indicate genetic selection over the last 50,000-75,000 years. Such long-term selection evidence is mainly useful for understanding the selection pressures in the EEA.

There is evidence for recent positive selection in humans.⁸⁵ Some of it may be in response to climate variations, producing a wide range of variation in salt-regulating genes in populations far from the equator.⁸⁶ Genes involved in brain development have

⁷⁹ (Barton and Partridge 2000)

⁸⁰ (Worden 1995)

⁸¹ (Cavalli-Sforza and Bodmer 1999)

⁸² (Hoekstra, Hoekstra, Berrigan, Vignieri, Hoang, Hill, Beerli and Kingsolver 2001)

⁸³ (Hoekstra, Hoekstra, Berrigan, Vignieri, Hoang, Hill, Beerli and Kingsolver 2001)

⁸⁴ (Sabeti, Schaffner, Fry, Lohmueller, Varilly, Shamovsky, Palma, Mikkelsen, Altshuler and Lander 2006)

⁸⁵ (Voight, Kudaravalli, Wen and Pritchard 2006)

⁸⁶ (Thompson, Kuttub-Boulos, Witonsky, Yang, Roe and Di Rienzo 2004)

also been shown to be under strong positive selection with new variants emerging over the last 37,000 years⁸⁷ and 5,800 years.⁸⁸

There is evidence that genes related to the brain have evolved more quickly in the human lineage than in other primates and rodents.⁸⁹ The rapid growth of the brain in the human lineage also suggests that its size must be controlled by relatively simple genetic mechanisms.⁹⁰ Despite this, it should be noted that the selection differential per generation for human brain weight during the Pleistocene was only 0.0004 per generation:⁹¹ even under fast evolution brain size was limited by tradeoffs.

If we find a gene that has a desirable effect, and that evolved recently and has not yet spread far despite showing evidence of positive selection, interventions that insert it into the genome or mimic its effects would likely meet the EOC. A simple example would be lactose tolerance. While development of lactose intolerance is adaptive for mammals since it makes weaning easier, dairy products have stimulated selection for lactase in humans over the last 5,000-10,000 years.⁹² This is so recent that there has not been time for the trait to diffuse to all human populations. (Populations that have domesticated cattle but do not have lactose tolerance instead make use of fermented milk or cheese.) Taking lactase pills enables lactose-intolerant people to digest lactose, widening the range of food they can enjoy. This enhancement clearly passes the EOC.

5. Discussion

The evolution heuristic instructs us to consider, for an apparently attractive enhancement, why we have not already evolved the intended trait if it is really such a good idea. We called this question the Evolutionary Optimality Challenge, and we have described three broad categories of possible answers, and given some examples of particular enhancements for which it is possible to meet the EOC, and which, therefore, seem comparatively promising as intervention targets that may be feasible in the relatively near term and which may have on balance beneficial effects.

In general, when we pose the EOC for some particular proposed enhancement, we might discover one of several things:

1. Current ignorance prevents us from forming any plausible idea about the evolutionary factors at play.
2. We come up with a plausible idea about the relevant evolutionary factors, and this reveals that the proposed modification would likely not be a net benefit.
3. We come up with a plausible idea about the relevant evolutionary factors, and this reveals why we would not already have evolved to have the enhanced capacity even if it would be a net benefit.

⁸⁷ (Evans, Gilbert, Mekel-Bobrov, Vallender, Anderson, Vaez-Azizi, Tishkoff, Hudson and Lahn 2005)

⁸⁸ (Mekel-Bobrov, Gilbert, Evans, Vallender, Anderson, Hudson, Tishkoff and Lahn 2005)

⁸⁹ (Dorus, Vallender, Evans, Anderson, Gilbert, Mahowald, Wyckoff, Malcom and Lahn 2004)

⁹⁰ (Roth and Dicke 2005)

⁹¹ (Cavalli-Sforza and Bodmer 1999, p. 692)

⁹² (Bersaglieri, Sabeti, Patterson, Vanderploeg, Schaffner, Drake, Rhodes, Reich and Hirschhorn 2004; Tishkoff, Reed, Ranciaro, Voight, Babbitt, Silverman, Powell, Mortensen, Hirbo, Osman, Ibrahim, Omar, Lema, Nyambo, Ghorri, Bumpstead, Pritchard, Wray and Deloukas 2007)

4. We come up with several plausible but mutually inconsistent ideas about the relevant evolutionary factors.

The first possibility means that we have no clear idea about why, from a phylogenetic perspective, the trait that is the target of the proposed enhancement is the way it is. This should give us pause. If we do not understand why a very complex evolved system has a certain property, there is a considerable risk that something will go wrong if we try to modify it. The case might be one of those where nature does know best. Like an over-ambitious tinkerer with merely superficial understanding of what he is doing while he is making changes to the design of a master craftsman, the potential for damage is considerable and the chances of producing an all-things-considered improvement are small.

We are not claiming that it is always inadvisable to try an intervention when we have no adequate understanding of the subsystem we intend to enhance. We might have other sources of evidence that afford us sufficient assurance that the intervention will work and will not cause unacceptable side effects, even without understanding the evolutionary functions involved. For example, we might have used the intervention many times before and found that it works well. Alternatively, we might have evidence from a closely analogous subsystem, such as an animal model, that suggests that the intervention should work in humans too. In such cases, the evolution heuristic delivers only a weak recommendation: that absent any good answer to the EOC, we should proceed only with great caution. In particular, we should be alert to the possibility that the proposed intervention will turn out to have significant (but perhaps subtle) side effects.

The second possibility is that we succeed in developing a plausible understanding of the pertinent evolutionary factors, and, having done so, we find our initial hopes about the proposed modification undermined. None of the three categories we have described yields a satisfactory answer to the EOC: the relevant tradeoffs have not changed since the EOC, there is no relevant value discordance, and no evolutionary restriction would have prevented the modification from already having evolved by now. In this case we have strong reason for thinking that the enhancement intervention will fail or backfire. If we proceed, the wisdom of nature will bite us.

The fourth possibility is that we come up with two or more plausible but incompatible evolutionary accounts of the evolutionary factors at play. In this case, we can consider the implications of each of the different evolutionary accounts separately according to the above criteria. If all yield green lights, we are encouraged to proceed. If some of the evolutionary accounts yield green lights but others yield red lights, then we face a situation of uncertainty. We can use standard decision theory to determine how to proceed – we can take a gamble if we feel that the balance of probabilities sufficiently favor the green lights; if not, we can attempt to acquire more information in order to reduce the uncertainty, or forego the potential enhancement and try something else.

The evolution heuristic is not a rival method to the more obvious way of determining whether some enhancement intervention works: testing it in well-designed clinical trials. Instead, the heuristic is complementary. It helps us ask some useful questions. By posing the EOC, and carefully searching for and evaluating possible answers in each of the three categories we described, we can (a) identify promising candidate enhancement interventions, to be explored further in laboratory and clinical studies, and (b) better evaluate the likelihood that some intervention which has shown

seemingly positive results in clinical studies will actually work as advertised and will not have unacceptable side-effects of a hidden, subtle, or long-term nature.

6. Conclusion

There is a widespread belief in some kind of “wisdom of nature”. Many people prefer “natural” remedies, “natural” food supplements, and “natural” ways of improving human capacities such as training, diet, and grooming. “Unnatural” interventions are often viewed with suspicion, and this attitude seems to be especially pronounced in relation to unnatural ways of enhancing human capacities, which are viewed as unwise, short-sighted, and hubristic. We believe that such attitudes also exert an influence on beliefs about the kind of matters that arise in bioethical discussions of human enhancement.

While it is tempting to dismiss intuitions about the wisdom of nature as vulgar prejudice, we have suggested that these intuitions contain a grain of truth, especially as they pertain to human enhancement. We have attempted to explicate this grain of truth as the Evolutionary Optimality Challenge.

After posing this challenge, the evolution heuristic instructs us to examine three broad categories of potential ways of meeting the challenge: changed tradeoffs, value discordance, and evolutionary restrictions. These categories correspond to systematic limitations in the wisdom of nature idea. For some potential enhancement interventions, the challenge can be met with an answer from one of these categories; for other potential interventions, the challenge cannot be met. The latter interventions merit suspicion, and attempting them may indeed be unwise, short-sighted, and hubristic. The former interventions, in contrast, do not defy the wisdom of nature and have a better chance of working.

By understanding both the sense in which there is validity in the idea that nature is wise and the limits beyond which the idea ceases to be valid, we are in a better position to identify promising human enhancements and to evaluate the risk-benefit ratio of extant enhancements. If we are right in supposing that intuitions about the wisdom of nature exert an inarticulate influence on opinion in contemporary bioethics of human enhancement, then the evolution heuristic – while primarily a method for addressing empirical questions – may also help to inform our assessments of more normatively loaded items of dispute.⁹³

References

- Aiello, L. C., N. Bates and T. Joffe. 2001. In defense of the Expensive Tissue Hypothesis. *Evolutionary Anatomy of the Primate Cerebral Cortex*. Falk, D. and K. R. Gibson, Cambridge University Press: 57–78.
- Allison, A. C. 1954. Protection Afforded by Sick Cell Trait Against Subtertian Malarial Infection. *British Medical Journal* 1: 290 - 294.

⁹³ For their comments, we are grateful to Rebecca Roache and [your name goes here!] for helpful comments on an earlier draft of this paper, and to the audience at the *TransVision 2006* conference in Helsinki, Finland, for useful questions.

- Andersson, R. E., G. Olaison, C. Tysk and A. Ekbom. 2001. Appendectomy and protection against ulcerative colitis. *New England Journal of Medicine* 344(11): 808-814.
- Baldwin, K. M. and F. Haddad. 2002. Skeletal muscle plasticity - Cellular and molecular responses to altered physical activity paradigms. *American Journal of Physical Medicine & Rehabilitation* 81(11): S40-S51.
- Barton, N. and L. Partridge. 2000. Limits to natural selection. *Bioessays* 22(12): 1075-1084.
- Battista, J. R., A. M. Earl and M. J. Park. 1999. Why is *Deinococcus radiodurans* so resistant to ionizing radiation? *Trends in Microbiology* 7(9): 362-5.
- Beauchamp, T. L. and J. F. Childress. 1979. *Principles of Biomedical Ethics*. New York & Oxford, Oxford University Press.
- Bersaglieri, T., P. C. Sabeti, N. Patterson, T. Vanderploeg, S. F. Schaffner, J. A. Drake, M. Rhodes, D. E. Reich and J. N. Hirschhorn. 2004. Genetic signatures of strong recent positive selection at the lactase gene. *American Journal of Human Genetics* 74(6): 1111-1120.
- Boon, T. and N. van Baren. 2003. Immunosurveillance against cancer and immunotherapy--synergy or antagonism?[comment]. *New England Journal of Medicine* 348(3): 252-4.
- Brunswick, N., G. N. Martin, L. Marzano and N. Savill. 2007. Visuo-spatial ability, handedness and developmental dyslexia: Just how sinister was Andy Warhol?
- Bunker, J. P. 2001. The role of medical care in contributing to health improvements within societies. *Int J Epidemiol* 30(6): 1260-3.
- Burt, A. and R. L. Trivers. 2006. *Genes in Conflict : The Biology of Selfish Genetic Elements*. Harvard, Belknap Press.
- Buss, D. M. 2000. The evolution of happiness. *American Psychologist* 55(1): 15-23.
- Caldwell, J. A. 2001. Efficacy of stimulants for fatigue management: The effects of Provigil and Dexedrine on sleep-deprived aviators. *Fatigue in Transportation*(Part F): 19-37.
- Cavalli-Sforza, L. L. and W. F. Bodmer. 1999. *The Genetics of Human Populations*, Dover Publications.
- Cochran, G., J. Hardy and H. Harpending. 2006. Natural History of Ashkenazi Intelligence. *Journal of Biosocial Science*.
- Colloca, L. and F. Benedetti. 2005. Placebos and painkillers: is mind as real as matter? *Nature Reviews Neuroscience* 6(7): 545-552.
- Dawkins, R. 1976. *The Selfish Gene*. Oxford, Oxford University Press.
- de Murcia, G. and S. Shall, Eds. 2000. *From DNA Damage and Stress Signaling to Cell Death: Poly ADP-Ribosylation Reactions*. Oxford, Oxford University Press.
- Demas, G. E., V. Chefer, M. I. Talan and R. J. Nelson. 1997. Metabolic costs of mounting an antigen-stimulated immune response in adult and aged C57BL/6J mice. *American Journal of Physiology-Regulatory Integrative and Comparative Physiology* 42(5): R1631-R1637.
- Diener, E., E. M. Suh, R. E. Lucas and H. L. Smith. 1999. Subjective well-being: Three decades of progress. *Psychological Bulletin* 125(2): 276-302.
- Dorus, S., E. J. Vallender, P. D. Evans, J. R. Anderson, S. L. Gilbert, M. Mahowald, G. J. Wyckoff, C. M. Malcom and B. T. Lahn. 2004. Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell* 119(7): 1027-1040.

- Dunn, G. P., L. J. Old and R. D. Schreiber. 2004. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity* 21(2): 137-48.
- Elia, M. 1992. Organ and tissue contribution to metabolic rate. *Energy metabolism: tissue determinants and cellular corollaries*. McKinney, J. M. and H. N. Tucker. New York, Raven: 61-79.
- Evans, P. D., S. L. Gilbert, N. Mekel-Bobrov, E. J. Vallender, J. R. Anderson, L. M. Vaez-Azizi, S. A. Tishkoff, R. R. Hudson and B. T. Lahn. 2005. Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science* 309(5741): 1717-1720.
- Fairclough, S. H. and K. Houston. 2004. A metabolic measure of mental effort. *Biological Psychology* 66(2): 177-190.
- Farah, M. J., J. Illes, R. Cook-Deegan, H. Gardner, E. Kandel, P. King, E. Parens, B. Sahakian and P. R. Wolpe. 2004. Neurocognitive enhancement: what can we do and what should we do? *Nature Reviews Neuroscience* 5(5): 421.
- Fish, J. L. and C. A. Lockwood. 2003. Dietary constraints on encephalization in primates. *American Journal of Physical Anthropology* 120(2): 171-181.
- Fisher, R. E. 2000. The primate appendix: A reassessment. *Anatomical Record* 261(6): 228-236.
- Flegal, K. A., B. I. Graubard, D. F. Williamson and M. H. H. Gail. 2005. Excess deaths associated with underweight, overweight, and obesity. *Jama-Journal of the American Medical Association* 293(15): 1861-1867.
- Fontaine, K. R., D. T. Redden, C. X. Wang, A. O. Westfall and D. B. Allison. 2003. Years of life lost due to obesity. *Jama-Journal of the American Medical Association* 289(2): 187-193.
- Force, U. S. P. S. T. 2002. Aspirin for the primary prevention of cardiovascular events: recommendation and rationale. *Annals of Internal Medicine* 136(2): 157-60.
- Frech, H. E. and R. D. Miller. 1996. The Productivity of Health Care and Pharmaceuticals: An International Comparison. UCLA Research Program in Pharmaceutical Economics and Policy 97-1. <http://repositories.cdlib.org/pep/97-1/>
- Freitas, R. A., Jr. 1998. Exploratory Design in Medical Nanotechnology: A Mechanical Artificial Red Cell. *Artificial Cells, Blood Substitutes, and Immobilization Biotechnology* 26: 411-430.
- Gaesser, G. A. 1999. Thinness and weight loss: beneficial or detrimental to longevity? *Medicine and Science in Sports and Exercise* 31(8): 1118-1128.
- Goldspink, G. 2005. Mechanical signals, IGF-I gene splicing, and muscle adaptation. *Physiology* 20: 232-238.
- Gottfredson, L. S. 1997. Why g matters: The complexity of everyday life. *Intelligence* 24(1): 79-132.
- Gottfredson, L. S. 2004. Life, death, and intelligence. *Journal of Cognitive Education and Psychology* 4(1): 23-46.
- Gottfredson, L. S. 2007. Innovation, fatal accidents, and the evolution of general intelligence. *Integrating the mind*. Roberts, M. J. Hove, UK, Psychology Press.
- Goulandris, N. K., M. J. Snowling and I. Walker. 2000. Is dyslexia a form of specific language impairment? A comparison of dyslexic and language impaired children as adolescents. *Annals of Dyslexia* L: 103-120.
- Green, C. S. and D. Bavelier. 2006. Enumeration versus multiple object tracking: the case of action video game players. *Cognition* 101(1): 217-245.

- Hagen, E. H. (2002). What is the EEA? (detailed answer). *The Evolutionary Psychology FAQ* Retrieved 02 July, 2006, from <http://www.anth.ucsb.edu/projects/human/epfaq/eea2.html>.
- Hammerstein, P. 1996. Darwinian adaptation, population genetics and the streetcar theory of evolution. *Journal of Mathematical Biology* 34(5-6): 511-532.
- Hoekstra, H. E., J. M. Hoekstra, D. Berrigan, S. N. Vignieri, A. Hoang, C. E. Hill, P. Beerli and J. G. Kingsolver. 2001. Strength and tempo of directional selection in the wild. *Proc Natl Acad Sci U S A* 98(16): 9157-9160.
- Holliday, M. A. 1986. Body composition and energy needs during growth. *Human Growth: A Comprehensive Treatise*. Falkner, F. and J. M. Tanner. New York, Plenum Press: 101-117.
- House, J. S., K. R. Landis and D. Umberson. 1988. Social relationships and health. *Science* 241(4865): 540-5.
- Hubbard, E. M., M. Piazza, P. Pinel and S. Dehaene. 2005. Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience* 6(6): 435-448.
- Humphrey, N. 2000. Great expectations: the evolutionary psychology of faith-healing and the placebo effect. *International Journal of Psychology* 35(3-4): 109-109.
- Jensen, P. S., D. Mrazek, P. K. Knapp, L. Steinberg, C. Pfeffer, J. Schowalter and T. Shapiro. 1997. Evolution and revolution in child psychiatry: ADHD as a disorder of adaptation. *Journal of the American Academy of Child and Adolescent Psychiatry* 36(12): 1672-1679.
- Kiecolt, J. K., R. Glaser, J. T. Cacioppo, R. C. MacCallum, M. Snyder-Smith, C. Kim and W. B. Malarkey. 1997. Marital conflict in older adults: Endocrinological and immunological correlates. *Psychosomatic Medicine* 59(4): 339-349.
- Kirsch, I., T. J. Moore, A. Scoboria and S. S. Nicholls. 2002. The Emperor's New Drugs: An Analysis of Antidepressant Medication Data Submitted to the U.S. Food and Drug Administration. *Prevention & Treatment* 5.
- Klaus, T., R. Joerger, E. Olsson and C. G. Granqvist. 1999. Silver-based crystalline nanoparticles, microbially fabricated. *Proc Natl Acad Sci U S A* 96(24): 13611-13614.
- Korol, D. L. and P. E. Gold. 1998. Glucose, memory, and aging. *American Journal of Clinical Nutrition* 67(4): 764S-771S.
- Koutroubakis, I. E. and I. G. Vlachonikolis. 2000. Appendectomy and the development of ulcerative colitis: Results of a metaanalysis of published case-control studies. *American Journal of Gastroenterology* 95(1): 171-176.
- Leroi, A. M., A. Bartke, G. De Benedictis, C. Franceschi, A. Gartner, E. S. Gonos, M. E. Fedei, T. Kivisild, S. Lee, N. Kartaf-Ozer, M. Schumacher, E. Sikora, E. Slagboom, M. Tatar, A. I. Yashin, J. Vijg and B. Zwaan. 2005. What evidence is there for the existence of individual genes with antagonistic pleiotropic effects? (vol 126, pg 421, 2005). *Mechanisms of Ageing and Development* 126(8): 906-906.
- Maizels, R. M. 2005. Infections and allergy - helminths, hygiene and host immune regulation. *Current Opinion in Immunology* 17(6): 656-661.
- Manning, C. A., W. S. Stone, D. L. Korol and P. E. Gold. 1998. Glucose enhancement of 24-h memory retrieval in healthy elderly humans. *Behavioural Brain Research* 93(1-2): 71-6.

- Martin, L. B., 2nd, A. Scheuerlein and M. Wikelski. 2003. Immune activity elevates energy expenditure of house sparrows: a link between direct and indirect costs? *Proc Biol Sci* 270(1511): 153-8.
- Martin, P. Y. and D. Benton. 1999. The influence of a glucose drink on a demanding working memory task. *Physiology & Behavior* 67(1): 69-74.
- Martin, R. D. 1996. Scaling of the mammalian brain: The maternal energy hypothesis. *News in Physiological Sciences* 11: 149-156.
- Matte, T. D. 2001. Influence of variation in birth weight within normal range and within sibships on IQ at age 7 years: cohort study (vol 323, pg 310, 2001). *British Medical Journal* 323(7314): 684-684.
- McCord, J. M. 2000. The evolution of free radicals and oxidative stress. *Am J Med* 108(8): 652-9.
- McDade, T. W. 2002. Status incongruity in Samoan youth: a biocultural analysis of culture change, stress, and immune function. *Medical Anthropology Quarterly* 16(2): 123-50.
- McDade, T. W. 2003. Life History Theory and the Immune System: Steps Toward a Human Ecological Immunology. *Yearbook of Physical Anthropology* 46: 100-125.
- McDaniel, M. A. 2005. Big-brained people are smarter: A meta-analysis of the relationship between in vivo brain volume and intelligence. *Intelligence* 33(4): 337-346.
- McKeown, T. and C. R. Lowe. 1974. *An Introduction to Social Medicine*. Oxford, Blackwell Scientific.
- Mealey, L. 1995. The Sociobiology of Sociopathy - an Integrated Evolutionary Model. *Behavioral and Brain Sciences* 18(3): 523-541.
- Meikle, A., L. M. Riby and B. Stollery. 2005. Memory processing and the glucose facilitation effect: the effects of stimulus difficulty and memory load. *Nutritional Neuroscience* 8(4): 227-32.
- Mekel-Bobrov, N., S. L. Gilbert, P. D. Evans, E. J. Vallender, J. R. Anderson, R. R. Hudson, S. A. Tishkoff and B. T. Lahn. 2005. Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. *Science* 309(5741): 1720-1722.
- Merkle, R. C. and R. A. Freitas. 2003. Theoretical analysis of a carbon-carbon dimer placement tool for diamond mechanosynthesis. *Journal of Nanoscience and Nanotechnology* 3(4): 319-324.
- Moret, Y. and P. Schmid-Hempel. 2000. Survival for immunity: the price of immune system activation for bumblebee workers. *Science* 290(5494): 1166-8.
- Neisser, U., G. Boodoo, T. J. Bouchard, A. W. Boykin, N. Brody, S. J. Ceci, D. F. Halpern, J. C. Loehlin, R. Perloff, R. J. Sternberg and S. Urbina. 1996. Intelligence: Knowns and unknowns. *American Psychologist* 51(2): 77-101.
- Nesse, R. M. and G. C. Williams. 1998. Evolution and the origins of disease. *Scientific American* 279(5): 86-93.
- Newhouse, J. P. and T. I. E. Group. 1993. *Free for All? Lessons from the RAND Health Insurance Experiment*. Cambridge, Massachusetts, Harvard University Press.
- Niven, J. E. 2005. Brain evolution: Getting better all the time? *Current Biology* 15(16): R624-R626.
- Nutt, D. J. 2006. Alcohol alternatives - a goal for psychopharmacology? *Journal of Psychopharmacology* 20(3): 318-320.

- Oria, R. B., P. D. Patrick, H. Zhang, B. Lorntz, C. M. D. Costa, G. A. C. Brito, L. J. Barrett, A. A. M. Lima and R. L. Guerrant. 2005. APOE4 protects the cognitive development in children with heavy diarrhea burdens in northeast Brazil. *Pediatric Research* 57(2): 310-316.
- Ots, I., A. B. Kerimov, E. V. Ivankina, T. A. Ilyina and P. Horak. 2001. Immune challenge affects basal metabolic activity in wintering great tits. *Proc Biol Sci* 268(1472): 1175-81.
- Parker, G. A. and J. M. Smith. 1990. Optimality Theory in Evolutionary Biology. *Nature* 348(6296): 27-33.
- Pica, P., C. Lemer, V. Izard and S. Dehaene. 2004. Exact and Approximate Arithmetic in an Amazonian Indigene Group. *Science* 306(5695): 499-503.
- Roth, G. and U. Dicke. 2005. Evolution of the brain and intelligence. *Trends in Cognitive Sciences* 9(5): 250-257.
- Sabeti, P. C., S. F. Schaffner, B. Fry, J. Lohmueller, P. Varilly, O. Shamovsky, A. Palma, T. S. Mikkelsen, D. Altshuler and E. S. Lander. 2006. Positive natural selection in the human lineage. *Science* 312(5780): 1614-1620.
- Sapolsky, R. M. 2005. The influence of social hierarchy on primate health.[see comment]. *Science* 308(5722): 648-52.
- Sauro, M. D. and R. P. Greenberg. 2005. Endogenous opiates and the placebo effect: a meta-analytic review. *Journal of Psychosomatic Research* 58(2): 115-20.
- Scholey, A. B., S. Harper and D. O. Kennedy. 2001. Cognitive demand and blood glucose. *Physiology & Behavior* 73(4): 585-592.
- Siegel, J. M. 2005. Clues to the functions of mammalian sleep. *Nature* 437(7063): 1264-1271.
- Silberman, S. 2001. The Geek Syndrome. *Wired* 9(12).
- Skaper, S. D. 2003. Poly(ADP-ribosyl)ation enzyme-1 as a target for neuroprotection in acute central nervous system injury. *Curr Drug Targets CNS Neurol Disord* 2(5): 279-91.
- Smith, J. M. 1982. *Evolution and the Theory of Games*, Cambridge University Press.
- Starfield, B. 2000. Is US health really the best in the world? *Jama-Journal of the American Medical Association* 284(4): 483-+.
- Thomas, M. K., D. M. Lloyd-Jones, R. I. Thadhani, A. C. Shaw, D. J. Deraska, B. T. Kitch, E. C. Vamvakas, I. M. Dick, R. L. Prince and J. S. Finkelstein. 1998. Hypovitaminosis D in medical inpatients.[see comment]. *New England Journal of Medicine* 338(12): 777-83.
- Thompson, E. E., H. Kuttub-Boulos, D. Witonsky, L. Yang, B. A. Roe and A. Di Rienzo. 2004. CYP3A variation and the evolution of salt-sensitivity variants. *American Journal of Human Genetics* 75(6): 1059-1069.
- Tishkoff, S. A., F. A. Reed, A. Ranciaro, B. F. Voight, C. C. Babbitt, J. S. Silverman, K. Powell, H. M. Mortensen, J. B. Hirbo, M. Osman, M. Ibrahim, S. A. Omar, G. Lema, T. B. Nyambo, J. Ghoris, S. Bumpstead, J. K. Pritchard, G. A. Wray and P. Deloukas. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics* 39(1): 31-40.
- Trevathan, W. 1987. *Human Birth: An Evolutionary Perspective*. New York, Aldine de Gruyter.
- Trevathan, W. R., E. O. Smith and J. J. McKenna, Eds. 1999. *Evolutionary medicine*. Oxford, Oxford University Press.

- Udry, J. R. 1978. Differential Fertility by Intelligence - Role of Birth Planning. *Social Biology* 25(1): 10-14.
- Vancourt, M. and F. D. Bean. 1985. Intelligence and Fertility in the United-States - 1912-1982. *Intelligence* 9(1): 23-32.
- Venkateswaran, A., S. C. McFarlan, D. Ghosal, K. W. Minton, A. Vasilenko, K. Makarova, L. P. Wackett and M. J. Daly. 2000. Physiologic Determinants of Radiation Resistance in *Deinococcus radiodurans*. *Applied and Environmental Microbiology* 66(66): 2620-2626.
- Vining, D. R., L. Bygren, K. Hattori, S. Nystrom and S. Tamura. 1988. IQ/Fertility Relationships in Japan and Sweden. *Personality and Individual Differences* 9(5): 931-932.
- Voight, B. F., S. Kudaravalli, X. Q. Wen and J. K. Pritchard. 2006. A map of recent positive selection in the human genome (vol 4, pg 154, 2006). *Plos Biology* 4(4): 659-659.
- von Karolyi, C., E. Winner, W. Gray and G. F. Sherman. 2003. Dyslexia linked to talent: Global visual-spatial ability. *Brain and Language* 85(3): 427-431.
- Waterlow, J. C. 1984. Protein turnover with special reference to man. *Q J Exp Physiol* 69(3): 409-38.
- Williams, G. C. 1996/1966. *Adaptation and Natural Selection*. Princeton, NJ, Princeton University Press.
- Williams, G. C. and R. M. Nesse. 1991. The Dawn of Darwinian Medicine. *Quarterly Review of Biology* 66(1): 1-22.
- Winder, R. and J. Borrill. 1998. Fuels for memory: the role of oxygen and glucose in memory enhancement. *Psychopharmacology* 136(4): 349-56.
- Wood, R. D., M. Mitchell, J. Sgouros and T. Lindahl. 2001. Human DNA repair genes. *Science* 291(5507): 1284-9.
- Worden, R. P. 1995. A Speed Limit for Evolution. *Journal of Theoretical Biology* 176(1): 137-152.
- Yazdanbakhsh, M., P. G. Kremsner and R. van Ree. 2002. Immunology - Allergy, parasites, and the hygiene hypothesis. *Science* 296(5567): 490-494.

Annexe 8

Publication Specimen

“Dignity and Enhancement” (Bostrom)

Dignity and Enhancement¹

Nick Bostrom

Oxford Future of Humanity Institute

Faculty of Philosophy & James Martin 21st Century School

Oxford University

[Commissioned for the President's Council on Bioethics, forthcoming (2007)]

www.nickbostrom.com

Does human enhancement threaten our dignity as some prominent commentators have asserted? Or could our dignity perhaps be technologically enhanced? After disentangling several different concepts of dignity, this essay focuses on the idea of dignity as a quality, a kind of excellence admitting of degrees and applicable to entities both within and without the human realm. I argue that dignity in this sense interacts with enhancement in complex ways which bring to light some fundamental issues in value theory, and that the effects of any given enhancement must be evaluated in its appropriate empirical context. Yet it is possible that through enhancement we could become better able to appreciate and secure many forms of dignity that are overlooked or missing under current conditions. I also suggest that in a posthuman world, dignity as a quality could grow in importance as an organizing moral/aesthetic idea.

The Meanings of Dignity and Enhancement

The idea of dignity looms large in the post-war landscape of public ethics. Human dignity has received prominent billing in numerous national and international declarations and

¹ For their comments, I'm grateful to Robin Hanson, Rebecca Roache, Anders Sandberg, Julian Savulescu, and to participants of the James Martin Advanced Research Seminar (20 October 2006, Oxford) and the Enhance Workshop (27 March 2007, Stockholm) where earlier versions of this paper were presented. I am especially indebted to Guy Kahane for discussions and insights, many of which have been incorporated into this paper, and to Rebecca Roache for research assistance. I would also like to thank Tom Merrill for helpful editorial suggestions.

constitutions. Like some successful politicians, the idea of dignity has hit upon a winning formula by combining into one package *gravitas*, a general feel-good quality, and a profound vagueness that enables all constituencies to declare their allegiance without thereby endorsing any particular course of action.

The idea of dignity, however, also has behind it a rich historical and philosophical tradition. For many of the ancients, dignity was a kind of personal excellence that only a few possessed to any significant degree. Marcus Tullius Cicero (106 to 43 BC), a Roman following in the footsteps of the Athenian Stoics, attributed dignity to all men, describing it as both a characteristic (human rationality) and a requirement (to base one's life on this capacity for rationality).² In Medieval Christianity, the dignity of man was based on the belief that God had created man in His image, allowing man to share some aspects of His divine reason and might.³ Theologians thought they saw man's dignity reflected in his upright posture, his free will, his immortal soul, and his location at the center of the universe. This dignity was viewed as an essential characteristic of the human being, possessed by each one of us, independent of social rank and personal excellence.

In the philosophy of Immanuel Kant, the intrinsic dignity of man was decoupled from theological assumptions about a divine heritage of the human species. According to Kant (here partly echoing the Stoics), all persons have dignity, a kind of absolute value that is incomparable to any price or instrumental utility.⁴ Kant held that dignity is not a quantitative notion; we cannot have more or less of it. The ground of the dignity of persons is their capacity for reason and moral agency. In order to respect this dignity, we must always treat another person as an end and never solely as a means. In order to avoid affronting our own dignity, we must also refrain from treating ourselves merely as a tool (such as by groveling to others, or selling ourselves into slavery) and from acting in ways that would undermine our rational agency (such as by using intoxicants, or committing suicide).⁵

² (Wetz 2000), p. 241f.

³ Ibid., 242.

⁴ This grounding of dignity in personhood and rational moral agency leaves out small children and some humans with mental retardation. This might be viewed as major problem (which Kant largely ignored).

⁵ The Stoics claimed that we *ought* to commit suicide if we know that our rational agency is at risk. Kant's dignity-based argument against suicide is more complex but less persuasive.

The term “human dignity” did not feature in any European declarations or constitutions in the 18th and 19th centuries. According to Franz Josef Wetz, it is found for the first time, albeit more or less in passing, in the German constitution drawn up in 1919 by the Weimar National Assembly, and its next appearance is in the corporate-fascist Portuguese constitution of 1933. Only in the aftermath of the Second World War does the concept’s heyday begin. It appears in about four constitutions in the period of 1900-1945 and in more than 37 from 1945-1997.⁶ It is also prominent in the UN Charter of 1945, the General Declaration of Human Rights of 1948, and in numerous later declarations, proclamations, and conventions.

Within applied ethics, the concept of dignity has been particularly salient in medical ethics and bioethics.⁷ It has been used to express the need for informed consent in medical research on human subjects. It has also been invoked (on both sides of the argument) in debates about end-of-life decisions and assisted euthanasia, and in discussions of organ sales and organ donations, assisted reproduction, human-animal chimaeras, pornography, torture, patenting of human genes, and human cloning. Recently, the idea of dignity has also been prominent in discussions of the ethics of human enhancement, where it has mostly been invoked by bioconservative commentators to argue against enhancement.⁸

If we examine the different uses which have been made of the idea of dignity in recent years, we can distinguish several different concepts. Before we can talk intelligibly about “dignity”, we must disambiguate the term. I propose the following taxonomy to regiment our dignity-talk:

- *Dignity as a Quality*: A kind of excellence; being worthy, noble, honorable. Persons vary in the degree to which they have this property. A form of Dignity as a Quality can also be ascribed to non-persons. In humans, Dignity as a Quality may be thought of as a virtue or an ideal, which can be cultivated, fostered,

⁶ From (Shultziner 2003), citing (Iglesias 2001).

⁷ Some think that this salience is undeserved; e.g. (Macklin 2003; Birnbacher 2005). See also (Beyleveld and Brownsword 2001; Ashcroft 2005; Caulfield and Brownsword 2006).

⁸ E.g. (Kass 2002).

respected, admired, promoted, etc. It need not, however, be identified with moral virtue or with excellence in general.⁹

- *Human Dignity (Menschenwürde)*: The ground upon which – according to some philosophers – rests the full moral status of human beings. It is often assumed that at least all normal human persons have the same level of human dignity. There is some disagreement about what precisely human dignity consists in, and this is reflected in disagreements about which individuals have human dignity: Only persons (as Kant maintained)? Or all human individuals with a developed nervous system who are not brain dead? Or fetuses in the womb too? Might some non-human primates also have this kind of dignity?¹⁰

Two other related ideas are:

- *Human Rights*: A set of inalienable rights possessed by all beings that have full moral status. One might hold that human dignity is the ground for full moral status. Human rights can be violated or respected. We might have a strict duty not to violate human rights, and an imperfect duty to promote respect for human rights.
- *(Dignity as) Social Status*: A relational property of individuals, admitting of gradation. Multiple status systems may exist in a given society. Dignity as Social Status is a widely desired prudential good. Our reasons for seeking social status are not distinctly moral, but the standards and conditions which determine the allocation of social status is a topic for ethical critique. Some social status is earned, but traditionally it was also thought that some individuals have a special

⁹ For Aristotle, excellence and virtue went together; his term for this was *kalon*, the noble. Earlier, however, in what we might call “Homeric ethics”, there was not such a close identification between virtue and honor or excellence. (I’m grateful to Guy Kahane for this point.)

¹⁰ These first two meanings are discussed in (Kolnai 1976) p. 259

intrinsic Dignity as Social Status, such as an aristocrat or a Brahmin.¹¹ Even though the Latin root word (*dignitas*) originally referred to a social status commanding respect, it might be best to refer to this property simply as Social Status to forestall confusion, reserving the word “dignity” for other uses.

All of these concepts are relevant to ethics, but in different ways.¹² In this paper, I shall focus on Dignity as a Quality and the ways in which this concept interacts with that of human enhancement.¹³

Before discussing its relations to enhancement, we shall need a richer characterization of Dignity as a Quality. I will draw on the sensitive linguistic and phenomenological analysis provided by Aurel Kolnai.¹⁴

On the idea of Dignity as a Quality of that which is dignified, Kolnai notes:

Dignity means Worth or Worthiness in some “absolute,” autonomized and objectivized, as it were “featural” sense... [Yet it] has *descriptive content*. ... It is, in this respect, on a par with any of the basic moral virtues such as justice, truthfulness, benevolence, chastity, courage, etc., including even integrity or conscientiousness, none of which is synonymous with Moral Goodness or Virtue as such, and each of which, notwithstanding its possible built-in reference to

¹¹ In respect of referring to a property partly acquired and partly inherent, the original concept of Dignity as Social Status might be thought of as intermediary between the concept of Dignity as a Quality and the concept of Human Dignity.

¹² See also (Nordenfelt 2004) for discussion of different types of dignity. Three of his dignity-concepts can be roughly mapped onto Dignity as a Quality, Human Dignity, and Dignity as Social Status. In addition, Nordenfelt also discusses a notion of *Dignity of Identity*, “the dignity we attach to ourselves as integrated and autonomous persons, persons with a history and persons with a future with all our relationships with other human beings” (p. 75). See also Adam Schulman’s introduction to this volume and (Shultziner 2003). One might also use “dignity” to refer to some combination of social status and self-esteem. For example, Jonathan Glover describes how stripping victims of their dignity (in this sense) is a common prelude to even greater atrocities; (Glover 1999).

¹³ For an earlier discussion of mine on the relation between the enhancement and Human Dignity, see (Bostrom 2005).

¹⁴ (Kolnai 1976). The Hungarian-born moral and political philosopher Kolnai (1900-1973) was, according to Karl Popper and Bernard Williams, one of the most original, provocative, and sensitive philosophers of the twentieth century. His writings have suffered some neglect and are not very widely known by philosophers working in the analytic tradition today. His explication of the concept of Dignity as a Quality is especially interesting because it seems to capture an idea that is motivating many contemporary bioconservative critiques of human enhancement.

Morality (and moral evaluation) as such, is susceptible to contentual description.¹⁵

On this understanding, Dignity as a Quality is a thick moral concept: it contains both descriptive and evaluative components, and may not be in any simple way reducible to more basic moral predicates. Dignity as a Quality also has certain aesthetic overtones. The term might have its own unique contribution to make to our normative vocabulary, but it should not be identified with Morality. If possessing Dignity as a Quality is a virtue, it is one out of many. The concept is hardly a promising candidate for the central and pivotal role in an ethical system that the idea of Human Dignity plays in Kantian philosophy and in some international declarations.¹⁶

We can proceed further by describing the appropriate responses to Dignity as a Quality. These seem to incorporate both aesthetic and moral elements. According to Kolnai, the term subtly connotes the idea of verticality, albeit tempered by also connoting a certain idea of reciprocity:

Can we attempt at all to assign, to adumbrate at least, a distinctive response to Dignity (or “the dignified”)? Whatever such a response might be, it must bear a close resemblance to our devoted and admiring appreciation of beauty (its “high” forms at any rate) on the one hand, to our reverent approval of moral goodness (and admiration, say, for heroic virtue) on the other. Dignity commands empathic respect, a reverential mode of response, an “upward-looking” type of the *pro* attitude: a “bowing” gesture if I may so call it.¹⁷

Next, let us consider what features call for such responses. What characteristics are typically dignified? While not claiming to produce an exhaustive list, Kolnai suggests the following:

¹⁵ Ibid., pp. 251f.

¹⁶ The related concept of *kalon*, however, does have such a foundational role in Aristotle’s ethics.

¹⁷ Ibid., p. 252.

First – the qualities of composure, calmness, restraint, reserve, and emotions or passions subdued and securely controlled without being negated or dissolved... Secondly – the qualities of distinctness, delimitation, and distance; of something that conveys the idea of being intangible, invulnerable, inaccessible to destructive or corruptive or subversive interference. ... Thirdly, in consonance therewith, Dignity also tends to connote the features of self-contained serenity, of a certain inward and toned-down but yet translucent and perceptible power of self-assertion... With its firm stance and solid immovability, the dignified quietly defies the world.¹⁸

Finally, regarding the bearers of such dignity, Kolnai remarks:

The predicates... are chiefly applicable to so-called “human beings,” i.e. persons, but, again, not exclusively so: much dignity in this sense seems to me proper to the Cat, and not a little, with however different connotation, to the Bull or the Elephant. ... Is not the austere mountainous plateau of Old Castile a dignified landscape...? And, though man-made, cannot works of art (especially of the “classic,” though not exactly “classicist,” type) have a dignity of their own?¹⁹

The term “enhancement” also needs to be explicated. I shall use the following rough characterization:

- *Enhancement*: An intervention that improves the functioning of some subsystem of an organism beyond its reference state; or that creates an entirely new functioning or subsystem that the organism previously lacked.

The function of a subsystem can be construed as either *natural* (and be identified with the evolutionary role played by this subsystem, if it is an adaptation), or *intentional* (in which case the function is determined by the contribution that the subsystem makes to

¹⁸ Ibid., pp. 253f.

¹⁹ Ibid., p. 254.

the attainment of relevant goals and intentions of the organism). The functioning of a subsystem is “improved” when the subsystem becomes more efficient at performing its function. The “reference state” may usually be taken to be the normal, healthy state of the subsystem, i.e. the level of functioning of the subsystem when it is not “diseased” or “broken” in any specific way. There is some indeterminacy in this definition of the reference state. It could refer to the state which is normal for some particular individual when she is not subject to any specific disease or injury. This could either be age-relative or indexed to the prime of life. Alternatively, the reference state could be defined as the “species-typical” level of functioning.

When we say “enhancement”, unless we further specify these and other indeterminacies, we do not express any very precise thought. In what follows, however, not much will hinge on exactly how one may choose to fill in this sketch of a definition of enhancement.

Greater Capacities

We can now begin our exploration of the relations between dignity and enhancement. If we recall the features that Kolnai suggests are associated with Dignity as a Quality – composure, distinctness, being inaccessible to destructive or corruptive or subversive interference, self-contained serenity, etc. – it would appear that these could be promoted by certain enhancements. Consider, for example, enhancements in executive function and self-control, concentration, or of our ability to cope with stressful situations; further, consider enhancements of mental energy that would make us more capable of independent initiative and that would reduce our reliance on external stimuli such as television; consider perhaps also enhancement of our ability to withstand mild pains and discomforts, and to more effectively self-regulate our consumption of food, exercise, and sleep. All these enhancements could heighten our Dignity as a Quality in fairly direct and obvious ways.

Other enhancements might reduce our Dignity as a Quality. For instance, a greatly increased capacity for empathy and compassion might (given the state of this world) diminish our composure and our self-contained serenity, leading to a reduction of our Dignity as a Quality. Some enhancements that boost motivation, drive, or emotional

responsiveness might likewise have the effect of destabilizing a dignified inner equilibrium. Enhancements that increase our ability rapidly to adapt to changing circumstances could make us more susceptible to “destructive or corruptive or subversive interference” and undermine our ability to stand firm and quietly defy the world.

Some enhancements, therefore, would increase our Dignity as a Quality, while others would threaten to reduce it. However, whether a particular enhancement – such as a strongly amplified sensitivity to others’ suffering – would in fact diminish our dignity depends on the context, and in particular on the character of the enhanced individual. A greatly elevated capacity for compassion is consistent with an outstanding degree of Dignity as a Quality, provided that the compassionate person has other mental attributes, such as a firm sense of purpose and robust self-esteem, that help contain the sympathetic perturbations of the mind and channel them into effective compassionate action. The life of Jesus, as described in the Bible, exemplifies this possibility.

Even if some enhancement reduced our Dignity as a Quality, it would not follow that the enhanced person would suffer a net loss of virtue. For while Dignity as a Quality might be a virtue, it is not the only virtue. Thus, some loss of Dignity as a Quality could be compensated for by a gain in other virtues. One could resist this conclusion if one believed that Dignity as a Quality is the only virtue rather than one among many. This is hardly a plausible view given the Kolnai-inspired understanding of Dignity as a Quality used in this paper.²⁰ Alternatively, one might hold that a certain threshold of Dignity as a Quality is necessary in order to be able to possess any other virtues. But even if that were so, it would not follow that any enhancement that reduced our Dignity as a Quality would result in a net loss of virtue, for the enhancement need not reduce our Dignity as a Quality below the alleged threshold.

The Act of Enhancement

Our Dignity as a Quality would in fact be greater if some of our capacities were greater than they are. Yet one might hold that *the act of enhancing* our capacities would in itself lower our Dignity as a Quality. One might also hold that *capacities obtained by means of some artificial enhancement* would fail to contribute, or would not contribute as much, to

²⁰ By contrast, e.g. to the Aristotelian concept of *Kalon*.

our Dignity as a Quality as the same capacities would have done had they been obtained by “natural” means.

For example, the ability to maintain composure under stressful conditions might contribute to our Dignity as a Quality if this capacity is the manifestation of our native temperament. The capacity might contribute even more to our Dignity as a Quality if it is the fruit of spiritual growth, as the result of long but successful psychological journey that has enabled us to transcend the trivial stressors that plague everyday existence. But if our composure is brought about by our swallowing of a Paxil, would it still reflect as favorably on our Dignity as a Quality?²¹

It would appear that our maintaining composure under stress will only fully count towards our Dignity as a Quality if we are able to view it as an authentic response, a genuine reflection of our autonomous self. In the case of the person who maintains composure only because she has taken Paxil, it might be unclear whether the composure is really a manifestation of her personality or merely of an extraneous influence. The extent to which her Paxil-persona can be regarded as her true persona would depend on a variety of factors.²² The more permanently available the anxiolytic is to her, the more consistent she is in using it in the appropriate circumstances, the more the choice of taking it is her own, and the more this choice represents her deepest wishes and is accompanied by a constellation of attitudes, beliefs, and values on which the availing herself of this drug is part of her self-image, the more we may incline to viewing the Paxil-persona as her true self, and her off-Paxil persona as an aberration.

If we compare some person who was born with a calm temperament to a one who has acquired the ability to remain calm as a result of psychological and spiritual growth, we might at first be tempted to think that the calmness is more fully a feature of the former. Perhaps the composure of a person born with a calm temperament is more stable, long-lasting, and robust than that of a person whose composure results from learning and

²¹ For this example to work properly, we should assume that the psychological states resulting are the same in each case. Suppose one thinks that there is a special dignity in *feeling* stressed out yet managing to *act* cool through an exertion of self-control and strength of character. Then the thought experiment requires that we *either* assume that the feeling of stress would be absent in all three cases (native temperament, psychological growth, Paxil), *or else* assume that (again in each of the cases) the feeling of stress would be present and the subject would succeed in acting cool thanks to her self-control (which might again have come about in either of the three ways).

²² Cf. (Kramer 1993).

experience. However, one could argue that the latter person's Dignity as a Quality is, *ceteris paribus*, the greater (i.e. even setting aside that this person would likely have acquired many other attributes contributing to his Dignity as a Quality during the course of his psychological trek). The reasoning would be that a capacity or an attribute that has become ours because of our own choices, our own thinking, and our own experiences, is in some sense more authentically ours even than a capacity or attribute given to us prenatally.

This line of reasoning also suggests that a trait acquired through the deliberate employment of some enhancement technology could be more authentically ours than a trait that we possessed from birth or that developed in us independently of our own agency. Could it be that not only the person who has acquired a trait through personal growth and experience, but also one who has acquired it by choosing to make use of some enhancement technology, may possess that trait more authentically than the person who just happens to have the trait by default? Holding other things constant – such as the permanency of the trait, and its degree of integration and harmonization with other traits possessed by the person – this would indeed seem to be the case.

This claim is consistent with the belief that coming to possess a positive trait as a result of personal growth and experience would make an *extra* contribution to our Dignity as a Quality, perhaps the dignity of effort and of the overcoming of weaknesses and obstacles. The comparison here is between traits, capacities, or potentials that we are given from birth and ones that we could develop if we are given access to enhancement technologies.²³

A precedent for the view that our self-shaping can contribute to our Dignity as a Quality can be found in Pico della Mirandola's *Oration on the Dignity of Man* (1486):

We have given you, O Adam, no visage proper to yourself, nor endowment properly your own, in order that whatever place, whatever form, whatever gifts you may, with premeditation, select, these same you may have and possess

²³ The claim I make here is thus also consistent with the view put forward by Leon Kass that the "naturalness" of the means matters. Kass argues that in ordinary efforts at self-improvement we have a kind of direct experience or "understanding in human terms" of the relation between the means and their effects, one that is lacking in the case of technological enhancements (Kass 2003).

through your own judgment and decision. The nature of all other creatures is defined and restricted within laws which We have laid down; you, by contrast, impeded by no such restrictions, may, by your own free will, to whose custody We have assigned you, trace for yourself the lineaments of your own nature. ... We have made you a creature neither of heaven nor of earth, neither mortal nor immortal, in order that you may, as the free and proud shaper of your own being, fashion yourself in the form you may prefer.

While *Mirandola* does not distinguish between different forms of dignity, it seems that he is suggesting both that our Human Dignity consists in our capacity for self-shaping, and also that we gain in Dignity as a Quality through the exercise of this capacity.

It is thus possible to argue that the act of voluntary, deliberate enhancement *adds* to the dignity of the resulting trait, compared to possessing the same trait by mere default.

The Enhancer's Attitude

At this point we must introduce a significant qualification. Other things equal, defiance seems more dignified than compliance and adaptation. As Kolnai notes, “pliability, unresisting adaptability and unreserved self-adjustment are prototypal opposites of Dignity”. Elaborating:

It might be argued that the feature sometimes described as the “meretricious” embodies the culmination of Un-Dignity. ... What characterizes the meretricious attitude is the intimate unity of abstract self-seeking and qualitative self-effacement. The meretricious type of person is, ideally speaking, at once boundlessly devoted to the thriving of his own life and indifferent to its contents. He wallows in his dependence on his environment – in sharp contrast to the dignity of a man’s setting bounds to the impact of its forces and undergoing their influence in a distant and filtered fashion – and places himself at the disposal of alien wants and interests without organically (which implies, selectively) espousing any of them. ... [He] escapes the tensions of alienation by precipitate

fusion and headlong surrender, and evades self-transcendence by the flitting mobility of a weightless self.²⁴

So on the one hand, the “self-made” man or woman might gain in Dignity as a Quality from being the author (or co-author) of his or her own character and situation. Yet on the other hand, it is also possible that such a person instead gains in Un-Dignity from their self-remolding. The possibility of such Un-Dignity, or loss of Dignity as a Quality, is an important concern among some critics of human enhancement. Leon Kass puts it uncompromisingly:

[The] final technical conquest of his own nature would almost certainly leave mankind utterly enfeebled. This form of mastery would be identical with utter dehumanization. Read Huxley’s *Brave New World*, read C. S. Lewis’s *Abolition of Man*, read Nietzsche’s account of the last man, and then read the newspapers. Homogenization, mediocrity, pacification, drug-induced contentment, debasement of taste, souls without loves and longings – these are the inevitable results of making the essence of human nature the last project of technical mastery. In his moment of triumph, Promethean man will become a contented cow.²⁵

The worry underlying this passage is, I think, the fear of a total loss of Dignity as a Quality, and its replacement with positive Un-Dignity.

We should distinguish two different ways in which this could result. The more obvious one is if, in selecting our enhancements, we select ones that transform us into undignified people. The point here is that these people would be undignified no matter how they came about, whether as a result of enhancement or through any other process. I have already discussed this issue, concluding that some enhancements would increase our Dignity as a Quality, other enhancements would risk reducing it, and also that whether a particular enhancement would be a benefit all-things-considered cannot usually be decided by looking only at how it would affect our dignity.

²⁴ (Kolnai 1976), pp. 265f.

²⁵ (Kass 2002), p. 48.

A more subtle source of Un-Dignity is one that emanates from the very activity of enhancement. In this latter case, the end state is not necessarily in itself undignified, but the process of refashioning ourselves which brings us there reduces our Dignity as a Quality. I argued above that a dignified trait resulting from deliberate enhancement can in favorable circumstance contribute more to our Dignity as a Quality than the same trait would if it had happened to be ours by default. Yet I think it should also be acknowledged that in *unfavorable* conditions, the act of self-transformation could be undignified and may indeed express the “meretricious” attitude described by Kolnai.

When is the activity of self-transformation dignity-increasing and when is it dignity-reducing? The Kolnai quote suggests an answer. When self-transformation is motivated by a combination of “abstract self-seeking and qualitative self-effacement”, when it is driven by alien wants and interests that have not been organically and selectively endorsed by the individual being enhanced, when it represents a surrender to mere convenience rather than the autonomous realization of a content-full personal ideal, then the act of enhancement is not dignified and may be positively undignified – in exactly the same way as other actions resulting from similar motivations may fail to express or contribute to our Dignity as a Quality.²⁶

Let us use an example. Suppose that somebody takes a cognition enhancing drug out of mere thoughtless conformity to fashion or under the influence of a slick advertising campaign. There is then nothing particularly dignified about this act of enhancement. There might even be something undignified about it inasmuch as a person who has Dignity as a Quality would be expected to exert more autonomous discretion about which substances she puts in her body, especially ones that are designed to affect her mental faculties. It might still be the case that the person after having taken the cognitive enhancer will gain in Dignity as a Quality. Perhaps the greater power and clarity of her thinking will enable her henceforth better to resist manipulative advertisements and to be more selective in her embrace of fads and fashions. Nonetheless, in itself, the enhancement act may be Undignified and may take away something from her Dignity as

²⁶ The act of enhancement could also be undignified under some other conditions. For example, one might think that if an intervention involves immoral conduct, or if it involves the use of “tainted means” (such as medical procedures developed using information obtained in cruel experiments), this would tend to make the intervention undignified. Again, however, this problem is not specific to enhancement-related acts.

a Quality. The problem is that her motivation for undergoing the enhancement is inappropriate. Her attitude and the behavior that springs from it are Un-Dignified.

Here we would be remiss if we did not point out the symmetric possibility that *refraining* from making use of an opportunity for enhancement can be Un-Dignified in exactly the same way and for the same reasons as it can be Un-Dignified to make use of one. A person who rejects a major opportunity to improve her capacities out of thoughtless conformity to fashion, prejudice, or lazy indifference to the benefits to self and others that would result, would thereby reduce her Dignity as a Quality. Rejection and acceptance of enhancement are alike in this respect: both can reflect an attitude problem.

Emotion Modification as a Special Hazard?

“Enhancements” of drives, emotions, mood, and personality might pose special threats to dignity, tempting us to escape “the tensions of alienation by precipitate fusion and headlong surrender”. An individual could opt to refashion herself to be content with reality as she finds it rather than standing firm in proud opposition. Such a choice could itself express a meretricious attitude. Worse, the transformation could result in a personality that has lost a great portion of whatever Dignity as a Quality it may have possessed before.

One can conceive of modifications of our affective responses that would level our aspirations, stymie our capacity for emotional and spiritual growth, and surrender our ability to rebel against unworthy life conditions or the shortcomings of our own characters. Such interventions would pose an acute threat to our Dignity as a Quality. The fictional drug “soma” in *Brave New World* is depicted as having just such effects. The drug seems to dissolve the contours of human living and striving, reducing the characters in Huxley’s novel to contented, indeterminate citizen-blobs that are almost prototypical of Un-Dignity.

Another prototypical image of Un-Dignity, one from the realm of science, is that of the “wire-headed” rat which has had electrodes inserted into its brain’s reward areas.²⁷ The model a self-stimulating rat, which will relentlessly press its lever – foregoing

²⁷ (Routtenberg and Lindy 1965).

opportunities for mating, rest, even food and drink – until it either collapses from fatigue or dies, is not exactly one that commands a “reverential mode of response” or an “upward-looking type of the *pro* attitude”. If we picture a human being in place of the rat, we would have to say that it is one Un-Dignified human, or at any rate a human engaged in a very Un-Dignified activity.²⁸

Would life in such an Un-Dignified state (assuming for the sake of argument that the pleasure was indefinitely sustainable and ignoring any wider effects on society) be preferable to life as we know it? Clearly, this depends on the quality of the life that we know. Given a sufficiently bleak alternative, intracranial electrical stimulation certainly seems much preferable; for example, for patients who are slowly dying in unbearable cancer pains and for whom other methods of palliation are ineffective.²⁹ It is even possible that for such patients, wire-heading and similar interventions increase their Dignity as a Quality (not to mention other components of their well-being).³⁰ Some estimable English doctors were once in the habit of administering to cancer patients in their last throes an elixir known as the Brompton cocktail, a mixture of cocaine, heroin and alcohol:

Drawing life to a close with a transcendently orgasmic bang, and not a pathetic and god-forsaken whimper, can turn dying into the culmination of one’s existence rather than its present messy and protracted anti-climax... One is conceived in pleasure. One may reasonably hope to die in it.³¹

Bowing out in such a manner would not only be a lot more fun, it seems, but also more *dignified* than the alternative.

But suppose that the comparison case is not unbearable agony but a typical situation from an average person’s life. Then becoming like a wire-headed rat, obsessively pressing a lever to the exclusion of all other activities and concerns, would

²⁸ The Stoics generalized this point, maintaining that “sensual pleasure is quite unworthy of the dignity of man and that we ought to despise it and cast it from us” (Cicero 1913), book 1, chapter 30. The virtue and dignity of asceticism, and the converse sinfulness and debasement of flesh-pleasing, have also been recurring themes in some religious traditions.

²⁹ It is used for this purpose in humans; (Kumar, Toth et al. 1997).

³⁰ For a discussion of the relations between dignity and suffering, see (Pullman 2002).

³¹ (Pearce 2001).

surely entail a catastrophic loss of Dignity as a Quality. Whether or not such a life would nevertheless be preferable to an ordinary human life (again assuming it to be sustainable and ignoring the wider consequences) – depends on fundamental issues in value theory. According to hedonism such a life would be preferable. If the pleasure would be great enough, it might also be preferable according to some other accounts of well-being. On many other value theories, of course, such a wire-headed life would be far inferior to the typical human life. These axiological questions are outside the scope of this essay.³²

Let us refocus on Dignity as a Quality. A life like one of a wire-headed rat would be radically deprived of Dignity as a Quality compared to a typical human life. But the wire-heading scenario is not necessarily representative – even as a caricature – of what a life with some form of emotional enhancement would be like. Some hedonic enhancements would not transform us into passive, complacent, loveless, and longing-less blobs. On the contrary, they could increase our zest for life, infuse us with energy and initiative, and heighten our capacity for love, desire, and ambition. There are different forms of pleasurable states of mind – some that are passive, relaxed, and comfortable, and others that are active, excited, enthusiastic, and joyfully thrilling. The wire-headed rat is potentially a highly misleading model of what even a simply hedonically enhanced life could be like. And emotional enhancement could take many forms other than elevation of subjective well-being or pleasure.

If we imagine somebody whose zest for and enjoyment of life has been enhanced beyond the current average human level, by means of some pharmaceutical or other intervention, it is not obvious that we must think of this as being associated with any loss of Dignity as a Quality. A state of mania is not dignified, but a controlled passion for life and what it has to offer is compatible with a high degree of Dignity as a Quality. It seems to me that such a state of being could easily be decidedly more dignified than the ho-hum affective outlook of a typical day in the average person's life.

Perhaps it would be slightly preferable, from the point of view of Dignity as a Quality, if the better mood resulted from a naturally smiling temperament or if it had

³² To assume that Dignity as a Quality has any intrinsic value would already be to renounce strict hedonism. However, even if one denies that Dignity as a Quality has intrinsic value, one might still think that it has other kinds of significance – for example, it might have instrumental value, or it might have value insofar as somebody desires it, or the concept of Dignity as a Quality might express or summarize certain common concerns.

been attained by means of some kind of psychological self-overcoming. But if some help had to be sought from a safe and efficacious pill, I do not see that it would make a vast difference in terms of how much Dignity as a Quality could be invested in the resulting state of mind.

One important factor in the Dignity as a Quality of our emotions is the extent to which they are appropriate responses to aspects of the world. Many emotions have an evaluative element, and one might think that for such an emotion to have Dignity as a Quality it must be a response to a situation or a phenomenon that we recognize as deserving the evaluation contained in the emotion. For example, anger might be dignified only on occasions where there is something to be angry about and the anger is directed at that object in recognition of its offensiveness. This criterion could in principle be satisfied not only by emotions arising spontaneously from our native temperament but also by emotions encouraged by some affective enhancement. Some affective enhancements could expand our evaluative range and create background conditions that would enable us to respond to values with regard to which we might otherwise be blind or apathetic. Moreover, even if some situations objectively call for certain emotional responses, there might be some indeterminacy such that any response within a range could count as objectively appropriate. This is especially plausible when we consider baseline mood or subjective well-being. Some people are naturally downbeat and glum; others are brimming with cheer and good humor. Is it really the case that one of these sentiments is objectively appropriate to the world? If so, which one? Those who are sad may say the former; those who are happy, the latter. I doubt that there is a fact of the matter.

It appears to me that the main threat to Dignity as a Quality from emotional enhancement would come not from the use of mood-brighteners to improve positive affect in everyday life, but from two other directions. One of these is the socio-cultural dimension, which I shall discuss in the next section. The other is the potential use of emotional “enhancements” by individuals to clip the wings of their own souls. This would be the result if we used emotional enhancers in ways that would cause us to become so “well-adjusted” and psychologically adaptable that we lost hold of our ideals, our loves and hates, or our capacity to respond spontaneously with the full register of human emotion to the exigencies of life.

Critics of enhancement are wont to dwell on how it could erode dignity. They often omit to point out how enhancement could help raise our dignity. But let us pause and ask ourselves just how much Dignity as a Quality a person has who spends four or five hours every day watching television? Whose passions are limited to a subset of eating, drinking, shopping, gratifying their sexual needs, watching sport, and sleeping? Who has never had an original idea, never willingly deviated from the path of least resistance, and never devoted himself seriously to any pursuit or occupation that was not handed him on the platter of cultural expectations? Perhaps, with regard to Dignity as a Quality, there is more distance to rise than to fall.

Socio-Culturally Mediated Effects

In addition to their direct effects on the treated individuals, enhancements might have indirect effects on culture and society. Such socio-cultural changes will in turn affect individuals, influencing in particular how much Dignity as a Quality they are likely to develop and display in their lives. Education, media, cultural norms, and the general social and physical matrix of our lives can either foster or stymie our potential to develop and live with Dignity as a Quality.

Western consumerist culture does not seem particularly hospitable to Dignity as a Quality. Various spiritual traditions, honor cultures, Romanticism, or even the Medieval chivalric code of ethics seem to have been more conducive to Dignity as a Quality, although some elements of contemporary culture – in particular, individualism – could in principle be important building blocks of a dignified personality. Perhaps there is a kind of elitism or aristocratic sensibility inherent in the cultivation of Dignity as a Quality that does not sit easily with the mass culture and egalitarian pretensions of modernity. Perhaps, too, there is some tension between the current emphasis on instrumentalist thinking and scientific rationality, on the one hand, and the (dignified) reliance on stable personal standards and ideals on the other. The perfect Bayesian rationalist, who has no convictions but only a fluid network of revisable beliefs, whose probability she feels

compelled to update according to a fixed kinematics whenever new evidence impinges on her senses, has arguably surrendered some of her autonomy to an algorithm.³³

How would the widespread use and social acceptance of enhancement technologies affect the conditions for the development of individual Dignity as a Quality? The question cannot be answered a priori. Unfortunately, nor can it currently be answered a posteriori other than in the most speculative fashion. We lack both the theory and the data that would be required to make any firm predictions about such matters. Social and cultural changes are difficult to forecast, especially over long time spans during which the technological bases of human civilizations will undergo profound transformations. Any answer we give today is apt to reveal more about our own hopes, fears, and prejudices than about what is likely to happen in the future.

When Leon Kass asserts that homogenization, mediocrity, pacification, drug-induced contentment, debasement of taste, and souls without loves and longings are the inevitable results of making human nature a project of technical mastery, he is not, as far as I can glean from his writings, basing this conviction on any corroborated social science model, or indeed on any kind of theory, data set, or well-developed argument. A more agnostic stance would better match the available evidence. We can, I think, conceive of scenarios in which Kass' forebodings come true, and of other scenarios in which the opposite happens. Until somebody develops better arguments, we shall be ignorant as to which it will be. Insofar as both scenarios are within reach, we might have most reason to work to realize one in which enhancement options do become available and are used in ways which increase our Dignity as a Quality along with other more important values.

The Dignity of Civilizations

Dignity as a Quality can be attributed to entities other than persons, including populations, societies, cultures, and civilizations. Some of the adverse consequences of enhancement that Kass predicts would pertain specifically to such collectives.

“Homogeneity” is not a property of an individual; it is a characteristic of a group of

³³ I say this as a fan of the Bayesian way. Another view would be that we do not have any coherent notion of autonomy that is distinct from responding to one's reasons, in which case the perfect Bayesian rationalist might – at at least in her epistemic performance) the epitome of dignity. That view would be more congruent with many earlier writers on dignity, including Kant.

individuals. It is not so clear, however, what Dignity as a Quality consists in when predicated to a collective. Being farther from the prototype application of the idea of dignity, such attributions of Dignity as a Quality to collectives may rely on value judgments to a greater extent than is the case when we apply it to individuals, where the descriptive components of the concept carry more of the weight.

For example, many moderns regard various forms of *equality* as important for a social order to have Dignity as a Quality. We may hold that there is something undignified about a social order which is marked by rigid status hierarchies and in which people are treated very unequally because of circumstances of birth and other factors outside their control. Many of us think that there is something decisively Undignified about a society in which beggars sit on the sidewalk and watch limousines drive by, or in which the conspicuous consumption of the children of the rich contrasts too sharply with the squalor and deprivation of the children of the poor.

An observer from different era might see things differently. For instance, an English aristocrat from the 17th century, placed in a time machine and brought forward into contemporary Western society, might be shocked at what would see. While he would, perhaps, be favorably impressed by our modern comforts and conveniences, our enormous economic wealth, our medical techniques and so forth, he might also be appalled at the loss of Dignity as a Quality that has accompanied these improvements. He steps out of the time machine and beholds vulgarized society, swarming with indecency and moral decay. He looks around and shudders as he sees how the rich social architecture of his own time, in which everybody, from the King down to the lowliest servant, knew their rank and status, and in which people were tied together in an intricate tapestry of duties, obligations, privileges, and patronage – how this magnificently ordered social cathedral has been flattened and replaced by an endless suburban sprawl, a *homogenized* society where the spires of nobility have been demolished, where the bonds of loyalty have been largely dissolved, the family pared down to its barest nucleus, the roles of lord and subject collapsed in that of consumer, the Majesty of the Crown usurped by a multinational horde of Burger Kings.

Whether or not our imaginary observer would judge that on balance the changes had been for the better, he would most likely feel that they had been accompanied by a

tragic loss and that part of this loss would be a loss of Dignity as a Quality, for individuals but especially for society. Moreover, this loss of societal Dignity would reside in some of the same changes that many of us would regard as gains in societal Dignity as a Quality.

We spark up a conversation with our time-traveling visitor and attempt to convince him that his view about Dignity as a Quality is incorrect. He attempts to convince us that it is our view that is defective. The disagreement, it seems, would be about value judgments and, to some extent, about aesthetic judgments. It is uncertain whether either side would succeed in persuading the other.

We could imagine other such transtemporal journeys, perhaps bringing a person from ancient Athens into the Middle Ages, or from the Middle Ages into the Enlightenment Era, or from the time when all humans were hunter-gatherers into any one of these later periods. Or we could imagine these journeys in the reverse, sending a person back in time. While each of these time travelers would likely recognize certain *individuals* in all the societies as having Dignity as a Quality, they might well find all the *societies* they were visiting seriously lacking in Dignity as a Quality. Even if we restrict ourselves to the present time, most of us probably find it easier to identify Un-Dignity in societies that are very different from our own, even though we have been taught that we ought not to be so prejudiced against of foreign cultures.

The point I wish make with these observations is that if you or I were shown a crystal ball revealing human society as it will be a few centuries from today, it is likely that the society we would see would appear to us as being in important respects Undignified compared to our own. This would seem to be the default expectation even apart from any technological enhancements which might by then have entered into common use. And therein lies one of those fine ironies of history. One generation conceives a beautiful design and lays the ground stones of a better tomorrow. Then they die, and the next generation decides to erect a different structure on the foundation that was build, a structure that is more beautiful in their eyes but which would have been hideous to their predecessors. The original architects are no longer there to complain, but if the dead could see they would turn in their graves. *O tempora, o mores*, cry the old, and the bones of our ancestors rattle their emphatic consent!

It is possible to have take a more optimistic view of the possibilities of secular change in the societal and cultural realms. One might believe that the history of humankind shows signs of moral progress, a slow and fluctuating trend towards more justice and less cruelty. Even if one does not detect such a trend in history, one might still hope that the future will be bring more unambiguous amelioration of the human condition. But there are many variables other than Dignity as a Quality that influence our evaluation of possible cultures and societies (such as the extent to which Human Dignity is respected to name but one). It may be that we have to content ourselves with hoping for improvements in these other variables, recognizing that Dignity as a Quality, when ascribed to forms of social organization rather than individuals, is too indeterminate a concept – and possibly too culture-relative – for even an optimist to feel confident that future society or future culture will appear highly dignified by current lights.

I will therefore not discuss by what means one might attempt to increase the Dignity as a Quality of present or future society, except to note that enhancement could possibly play a role. For example, if homogenization is antithetical to a society having Dignity as a Quality, then enhancements that strengthen individuals' ability to resist group pressure and that encourage creativity and originality, maybe even a degree of eccentricity, could help not only individuals to attain more Dignity as a Quality but also society, thanks to the cultural diversification that such individuals would create.

A Relational Component?

Let us return to the Dignity as a Quality of individuals. One might attribute Dignity as Quality to an individual not only because of her intrinsic characteristics but – arguably – also because of her relational properties. For example, one might think that the oldest tree has a Dignity as a Quality that it would not possess if there were another tree that was older, or that the last Mohican had a special Dignity as a Quality denied to the penultimate Mohican.

We humans like to pride ourselves on being the smartest and most advanced species on the planet. Perhaps this position gives us a kind of Dignity as a Quality, one which could be shared by all humans, including mediocrities and even those who fall below some non-human animals in terms of cognitive ability? We would have this

special Dignity as a Quality through our belonging to a species whose membership has included such luminaries as Michelangelo and Einstein. We might then worry that we would risk losing this special dignity if, through the application of radical enhancement technologies, we created another species (or intelligent machines) that surpassed human genius in all dimensions? Becoming a member of the second-most advanced species on the planet (supposing one were not among the radically enhanced) sounds like a demotion.

We need to be careful here not to conflate Dignity as a Quality with other concepts, such as social rank or status. With the birth of cognitively superior posthumans, the rank of humans would suffer (at least if rank were determined by cognitive capacity). It does not follow that our Dignity as a Quality would have been reduced; that is a separate question. Perhaps we should hold, rather, that our Dignity as a Quality would have been increased, on grounds of our membership in another collective – the Club of Tellurian Life. This club, while less exclusive than the old Club of Humanity, would boast some extremely illustrious members after the human species had been eclipsed by its posthuman descendants.

There might nevertheless be a loss of Dignity as a Quality for individual human beings. Those individuals who were previously at the top of their fields would no longer occupy such a distinguished position. If there is a special Dignity as a Quality (as opposed to merely social status) in having a distinguished position, then this dignity would be transferred to the new occupants of the pinnacles of excellence.

We cannot here explore all the possible ways in which relational properties could be affected by human enhancement, so I will draw attention to just one other relational property, that of uniqueness. Reproductive cloning is not a prototypal enhancement, but we can use it to raise a question.³⁴ Does a person's uniqueness contribute something to her Dignity as a Quality? If so, one might object to human cloning on grounds that it would result in a progeny who – other things equal – would have less Dignity as a Quality than a sexually conceived child. Of course, we should not commit the error of genetic essentialism or genetic determinism; but neither should we make the opposite

³⁴ One could argue that reproductive cloning would be an enhancement of our reproductive capacities, giving us the ability to reproduce in a way that was previously impossible.

error of thinking that genes don't matter. People who have the same genes tend to be more similar to one another than people who are not genetically identical. In this context, "uniqueness" is a matter of degree, so a set of clones of an average person would tend to be "less unique" than most people.³⁵

Naturally occurring identical twins would be as genetically similar as a pair of clones. (Natural identical twins also tend to share the same womb and rearing environment, which clones would not necessarily do.) Since we do not think that natural twins are victims of a significant misfortune, we can conclude that *either* the loss of one's degree of uniqueness resulting from the existence of another individual who is genetically identical to oneself does not entail a significant loss of Dignity as a Quality, *or* losing some of one's Dignity as a Quality is not a significant misfortune (or both).

One might still worry about more extreme cases. Consider the possibility of not just a few clones being created of an individual, but many millions. Or more radically, consider the possibility of the creation of millions of copies of an individual who would all be much more similar to one another than monozygotic twins are.³⁶ In these imaginary cases, it seems more plausible that a significant loss of Dignity as a Quality would occur among the copied individuals. Perhaps this would be a *pro tanto* reason against the realization of such scenarios.

Dignity Outside the Human World: Quiet Values

Dignity as a Quality is not necessarily confined to human beings and collectives of human beings.

The redwoods, once seen, leave a mark or create a vision that stays with you always. No one has ever successfully painted or photographed a redwood tree. The feeling they produce is not transferable. From them comes silence and awe. It's not only their unbelievable stature, nor the color which seems to shift and vary under your eyes, no, they are not like any trees we know, they are

³⁵ Unless, perhaps, cloning were so rare that being a clone would itself mark one out as a highly unusual and "unique" kind of person.

³⁶ Human "uploading" is one possible future technology that might lead to such a scenario; (Moravec 1988). Another would be the creation of many copies of the same sentient artificial intelligence.

ambassadors from another time. They have the mystery of ferns that disappeared a million years ago into the coal of the carboniferous era. ... The vainest, most slap-happy and irreverent of men, in the presence of redwoods, goes under a spell of wonder and respect. ... One feels the need to bow to unquestioned sovereigns.³⁷

It is easy to emphasize with the response that John Steinbeck describes, and it fits quite well with Kolnai's account of the characteristic response to dignity.

Another example:

[One] of my colleagues [recounts a story] about once taking his young son to a circus in town, and discovering a lone protestor outside the tent silently holding aloft a sign that read "REMEMBER THE DIGNITY OF THE ELEPHANTS." It hit him like a lightning bolt, he said. The protester's point is surely an intelligible one, though we could debate whether it is genuinely reason enough to avoid all types of circuses.³⁸

We need a name for the property that we feel we are responding to in examples like the above, and "Dignity as Quality" fits the bill. We might also apply this concept to certain actions, activities, and achievements, perhaps to certain human relationships, and to many other things, which I shall not explore here.

The Dignity as a Quality that we attribute to non-humans (or more accurately, to non-persons) is of a different type from that which we attribute to human beings. One way to characterize the difference is by using a distinction introduced by Stephen Darwall.³⁹ Darwall describes two different kinds of attitude both of which are referred to by the term "respect". The first kind he calls *recognition respect*. This attitude consists in giving appropriate consideration or recognition to some feature of its object in deliberating about what to do, and it can have any number of different sorts of things as

³⁷ (Steinbeck 1962), p. 168f.

³⁸ (Duncan 2006), p. 5.

³⁹ (Darwall 1977). What follows is a simplified description of Darwall's account which skirts over some of its finer points.

its object. The other kind, which he calls *appraisal respect*, consists in an attitude of positive appraisal of a person either as a person or as engaged in some particular pursuit. The appropriate ground for appraisal respect is that a person has manifested positive characteristics or excellences which we attribute to his character, especially those which belong to him as a moral agent.

For example, when we say that Human Dignity must be respected, we presumably mean that it must be given recognition respect. We owe this respect to all people equally, independently of their moral character or any special excellences that they might have or lack. By contrast, when say that we should respect Gandhi for his magnanimity, we are probably referring to appraisal respect (although his magnanimity should also in certain contexts be given recognition respect). Similarly, if someone has a high degree of Dignity as a Quality (perhaps Gandhi again), this also calls for appraisal respect.

The kind of Dignity as a Quality that we attribute to non-agents does not call for appraisal respect, since only agents have moral character. Thus we can distinguish between Dignity as a Quality in the narrow sense, as a property possessed only by (some) agents, and which calls for appraisal respect; and Dignity as a Quality in a wider sense, which could be possessed by any number of types of object, and which calls for recognition respect only. We do not have to literally *admire* or *give credit to* the redwoods for having grown so tall and having lived so long; but we can still recognize them as possessing certain features that we should take into account in deliberating about what we do to them. In particular, if we are truly impressed by their Dignity as a Quality (in the wide sense), then we ought to show our recognition respect for their dignity – perhaps by not harvesting them down for their timber, or by refraining from urinating on them.

Dignity as a Quality, in this wide sense, is ubiquitous. What is limited, I would suggest, is not the supply but our ability to appreciate it. Even inanimate objects can possess it. For a mundane example, consider the long, slow, sad decline of a snowman melting in the backyard. Would not an ideally sensitive observer recognize a certain Dignity as a Quality in the good Snowman, Esq.?

The ethical fades here into the aesthetical (and perhaps into the sentimental), and it is not clear that there exists any sharp line of demarcation. But however we draw our

conceptual boundaries, our normative discourse would be impoverished if it could not lend expression to and genuinely take into account what is at stake in cases like these. Perhaps we could coin the category of *quiet values* to encompass not only Dignity as a Quality in this extended sense, but also other small, subtle, or non-domineering values. We may contrast these quiet values with a category of *loud values*, which would be more starkly prudential or moral, and which tend to dominate the quiet values in any direct comparison. The category of loud values might include things like alleviation of suffering, justice, equality, freedom, fairness, respect for Human Dignity, health and survival, and so forth.⁴⁰

It is not necessarily a fault of applied ethics, insofar as it aims to influence regulation and public policy, that it tends to focus exclusively on loud values. If on one side of the scales we put celebrating the Dignity as a Quality of Mr. Snowman, and on the other we put providing a poverty-stricken child with a vaccination, the latter will always weigh more heavily.

Nevertheless, there may be a broader significance to the quiet values. While individually weak, in aggregate they are formidable. They are the dark matter of value theory (or, for all ye business consultants among my readers, *the long tail of axiology*). Fail to uphold a quiet value on one occasion, and nothing noticeable is lost. But extirpate or disregard all the quiet values all the time, and the world turns into a sterile, desolate, impoverished place. The quiet values add the luminescence, the rich texture of meaning, the wonder and awe, and much of the beauty and nobility of human action. In major part, this contribution is aesthetic, and the realization of this kind of value might depend crucially on our subjective conscious responses. Yet, at least in the idea of Dignity as a Quality, which is our focal concern here, the moral and the aesthetic blend into one another, and the possibility of responding to the realm of quiet values (or helping it into existence through acts of creative imagination and feeling) can have moral implications.

⁴⁰ It is, of course, a substantive normative question in which of these categories to place a value. For example, Nietzsche might have held Dignity as a Quality to be a loud value, and he might have thought that equality was no value at all. One big question, even if one does not share Nietzsche's view, is how we ought to treat Dignity as a Quality from an impartial standpoint. Is it better to have a few supremely dignified persons surrounded by many with little dignity, or better to have a modicum of dignity widely spread?

The Eschatology of Dignity

Kolnai describes a certain mode of utopian thinking as inimical to Dignity as a Quality:

[Some people believe] that by the ensuring through a collective agency of everybody's "Human Dignity" (including a sense of individual self-assertion and self-fulfillment) everyone will also acquire Dignity as a Quality or, what comes to the same thing, the concept of "Dignity as a Quality" will lose its point – a view prefigured by the first great apostle of Progress, Condorcet, who confidently foresaw a rationally and scientifically redrawn world in which there would be no opportunity for the exercise of heroic virtue nor any sense in revering it. ... The core of Un-Dignity, as I would try to put it succinctly, is constituted by an attitude of refusal to recognize, experience, and bear with, the tension between Value and Reality; between what things ought to be, should be, had better be or are desired to be and what things are, can be and are allowed to be.⁴¹

This raises the question of whether there would be any role left to play for Dignity as a Quality if the world, thanks to various political, medical, economical, and technological advances, reached a level of perfection far beyond its present troubled state. The question becomes perhaps especially acute if we suppose that the transhumanist aspiration to overcome some of our basic biological limitations were to be realized. Might the tension between Value and Reality then be relaxed in such a way that Dignity as a Quality would become meaningless or otiose?

Let us make a leap into an imaginary future posthuman world, in which technology has reached its logical limits. The superintelligent inhabitants of this world are *autopotent*, meaning that they have complete power over and operational understanding of themselves, so that they are able to remold themselves at will and assume any internal state they choose. An autopotent being could, for example, easily transform itself into the shape of a woman, a man, or a tree. Such a being could also easily enter any subjective state it wants to be in, such as state of pleasure or indignation,

⁴¹ Ibid., p. 262. Kolnai stresses that the "core of Un-Dignity" does *not* include "either submission to the existing order of things and the virtue of patience, or a sustained endeavor for reform, improvement and assuagement."

or a state of experiencing the visual and tactile sensations of a dolphin swimming in the sea. We can also assume that these posthumans have thorough control over their environment, so that they can make molecularly exact copies of objects and implement any physical design for which they have conceived of a detailed blueprint. They could make a forest of redwood trees disappear, and then recreate an exactly similar forest somewhere else; and they could populate it with dinosaurs or dragons – they would have the same kind of control of physical reality as programmers and designers today have over virtual reality, but with the ability to imagine and create much more detailed (e.g. biologically realistic) structures. We might say that the autopotent superintelligences are living in a “plastic world” because they can easily remold their environment exactly as they see fit.

Now, it might be that in any technological utopia which we have any real chance of creating, all individuals will remain constrained in important ways. In addition to the challenges of the physical frontiers, which might at this stage be receding into deep space as the posthuman civilization expands beyond its native planet, there are the challenges created by the existence of other posthumans, that is, the challenges of the social realm. Resources even in Plastic World would soon become scarce if population growth is exponential, but aside from material constraints, individual agents would face the constraints imposed on them by the choices and actions of other agents. Insofar as our goals are irreducibly social – for example to be loved, respected, given special attention or admiration, or to be allowed to spend time or to form exclusive bonds with the people we choose, or to have a say in what other people do – we would still be limited in our ability to achieve our goals. Thus, a being in Plastic World may be very far from omnipotent. Nevertheless, we may suppose that a large portion of the constraints we currently face have been lifted and that both our internal states and the world around us have become much more malleable to our wishes and desires.

In Plastic World, many of the moral imperatives with which we are currently struggling are easily satisfiable. As the loud values fall silent, the quiet values become

more audible.⁴² With most externally imposed constraints eliminated by technological progress, the constraints *which we choose to impose on ourselves* become paramount.

In this setting, Dignity as a Quality could be an organizing idea. While inanimate objects cannot possess Human Dignity, they can be endowed with a kind of Dignity as a Quality. The autopotent inhabitants of Plastic World could choose to cultivate their sensibility for Dignity as a Quality and the other quiet values. By choosing to recognize these values and to treat the world accordingly, they would be accepting some constraints on their actions. It is by accepting such constraints that they could build, or rather *cultivate* their Plastic World into something that has greater value than a daydream. It is also by accepting such constraints – perhaps only by doing so – that it would be possible for them to preserve their own Dignity as a Quality. This dignity would not consist in resisting or defying the world. Rather, theirs would be a dignity of the strong, consisting in self-restraint and the positive nurturance of both internal and external values.

References

- Ashcroft, R. (2005). "Making sense of dignity." *Journal of Medical Ethics* 31: 679-682.
- Beyleveld, D. and R. Brownsword (2001). *Human dignity in bioethics and biolaw*. Oxford, Oxford University Press.
- Birnbacher, D. (2005). "Human cloning and human dignity." *Reproductive BioMedicine Online* 10(supplement 1): 50-55.
- Bostrom, N. (2005). "In Defence of Posthuman Dignity." *Bioethics* 19(3): 202-214.
- Caulfield, T. and R. Brownsword (2006). "Human dignity: a guide to policy making in the biotechnology era?" *Nature Reviews Genetics* 7: 72-76.
- Cicero, M. T. (1913). *Cicero De officiis*. Trans. Walter Miller. Cambridge, Harvard University Press.
- Darwall, S. L. (1977). "Two Kinds of Respect." *Ethics* 88(1): 36-49.
- Duncan, C. (2006). "Respect for Dignity: A Defense." *Manuscript* 10/06 draft.
- <http://www.ithaca.edu/faculty/cduncan/respect.doc>

⁴² This is not to say that the quiet values would actually be heard or heeded if and when the loud values fall silent. Whether that would happen is difficult to predict. But an *ideal* moral agent would begin to pay more attention to the quiet values in such circumstances and would let them play a greater role in guiding her conduct.

- Glover, J. (1999). *Humanity: a moral history of the twentieth century*. London, J. Cape.
- Iglesias, T. (2001). "Bedrock Truths and the Dignity of the Individual." *Logos* 4(1): 114-134
- Kass, L. (2002). *Life, liberty, and the defense of dignity: the challenge for bioethics*. San Francisco, Encounter Books.
- Kass, L. (2003). "Ageless Bodies, Happy Souls: Biotechnology and the Pursuit of Perfection." *The New Atlantis* 1: 9-28.
- Kolnai, A. (1976). "Dignity." *Philosophy* 51(197): 251-271.
- Kramer, P. D. (1993). *Listening to Prozac*. New York, N.Y., U.S.A., Viking.
- Kumar, K., C. Toth, et al. (1997). "Deep Brain Stimulation for Intractable Pain: A 15-Year Experience." *Neurosurgery* 40(4): 736-746.
- Macklin, R. (2003). "Dignity is a useless concept." *British Medical Journal* 327: 1419-1420.
- Moravec, H. P. (1988). *Mind children: the future of robot and human intelligence*. Cambridge, Mass., Harvard University Press.
- Nordenfelt, L. (2004). "The Varieties of Dignity." *Health Care Analysis* 12(2): 69-81.
- Pearce, D. (2001). "When Is It Best To Take Crack Cocaine?" *LA Weekly* July 6-12.
- Pullman, D. (2002). "Human dignity and the ethics and aesthetics of pain and suffering." *Theoretical Medicine* 23: 75-94.
- Routtenberg, A. and J. Lindy (1965). "Effects of the availability of rewarding septal and hypothalamic stimulation on bar pressing for food under conditions of deprivation." *Journal of Comparative and Physiological Psychology* 60(2): 158-161.
- Shultziner, D. (2003). "Human dignity - functions and meanings." *Global Jurist Topics* 3(3): 1-21.
- Steinbeck, J. (1962). *Travels with Charley; in search of America*. New York, Viking Press.
- Wetz, F. J. (2000). *The Dignity of Man. Anatomy Art - Fascination Beneath the Surface*. G. v. Hagens and A. Whalley. Heidelberg, Institute for Plastination, Heidelberg.