

The Future of Humanity

NICK BOSTROM

The future of humanity has traditionally been a theological topic. All the major religions have teachings about the ultimate destiny of humanity or the end of the world. Eschatological themes have also been explored by philosophers, including Hegel, Kant and Marx. Science fiction authors, too, have had plenty to say on the subject. Very often, the future has served as a projection screen for our hopes and fears, for entertaining drama, morality tales, and reflections of tendencies in contemporary society. Only rarely is humanity's future taken seriously as a subject matter on which it is important to try to have factually correct beliefs.

Most important differences between ourselves and our forebears are ultimately related to technology. In the early days of our species, technological progress was slow. Tens of thousands of years would pass without much accumulation. Only within the last couple of hundred years could a person expect to experience significant technological change within her lifetime. Inventor and writer Ray Kurzweil argues that technological development is still accelerating. On the basis of exponential trends in a number of high-tech areas, he predicts a technological "singularity" before the middle of this century (Kurzweil 2005).

Technology in a wide sense (including not only gadgets but also methods, techniques and institution design principles) is the fundamental cause of long-term economic growth. Economic growth is what has enabled the world population to increase to over 6 billion people; up from the 4 million or so that inhabited the planet when humans lived as hunter-gatherers. Economic growth has also enabled cities and labor specialization, and hence indirectly all the phenomena made possible by high-density population centers with skilled laborers – including, significantly, a much faster pace of innovation.

Pessimists about the future often focus on the environmental problems facing the growing world population. They worry that our current wasteful and polluting ways are unsustainable and threatening to human civilization. Paul Ehrlich's *Population Bomb* (1968) and the Club of Rome's report *Limits to Growth* (1972), which sold 30 million copies, predicted economic collapse and mass starvation by the 1980s or 1990s as a result of population growth and resource depletion (Ehrlich 1968, Meadows and Club of Rome 1972). The basic idea of population growth as the nemesis of human welfare goes back to the English demographer and political economist Thomas Robert Malthus (1766–1834). Malthus argued that the lower classes could never permanently be lifted

out of poverty, because as their condition improved they would have more surviving children and more mouths to feed. Over time, population would outgrow food supply, starvation would occur, and the majority of men would again be reduced to subsistence-level incomes (Malthus 1798).

In the long run, average income can only increase if economic growth is faster than population growth. Long-term economic growth is determined by technological progress. The predictions of Malthus and his latter-day followers failed because economic growth has been faster, and population growth slower, than they expected. Malthus would have been surprised to find that fertility has declined dramatically in high-income countries. Global population growth is currently just over 1 percent, while global economic growth over the last three decades has averaged about 3 percent per year (US Census Bureau 2007; Maddison and Organization for Economic Co-operation and Development. Development Centre 2003: 257–63).

The human species is not in an evolutionary equilibrium. Our current reproductive instincts and child-rearing preferences are not fitness-maximizing. It takes many generations for biological evolution to reshape our behavioral tendencies. If the present fitness landscape remained unchanged for a long time, we should expect *Homo sapiens* to evolve new dispositions that promote fitness under modern conditions – such as an aversion to contraceptives, a strong desire for big families, and perhaps a disinclination to fitness-reducing choices such as extended education. Memetic evolution might produce these results faster. Some groups, such as the Hutterites, an Anabaptist sect, have been growing despite high defection rates because of their extremely high fertility rate – an average Hutterite woman gives birth to nine children (Lang and Gohlen 1985). The Hutterites oppose birth control and see high fertility as a sign of divine blessing. Both in biological and memetic terms, human evolution is still occurring – probably at an unusually fast pace since our habitat has changed so much in recent times. If the human socio-economic habitat were magically frozen in its present state, Malthus would eventually be vindicated.

Not all pessimists focus on environmental problems or Malthusian scenarios. Many other catastrophe scenarios have been proposed. Of these, one can distinguish an especially severe subset: *existential risks* (Bostrom 2002b). An existential disaster is one which would either cause the extinction of Earth-originating intelligent life or permanently and drastically curtail its potential. Such an event would completely and irreversibly destroy humanity's future.

Existential risks have not received as much scholarly attention as they deserve. In recent years, there have been three serious books and one major paper on this topic. John Leslie, a philosopher, puts the probability of humanity failing to survive the next five centuries at 30 percent, partly based on the controversial “Doomsday argument” (Leslie 1996, Bostrom 2002a). Sir Martin Rees, an astronomer and president of Britain's Royal Society, is even more pessimistic, thinking the odds that we shall survive the twenty-first century are no better than 50 percent (Rees 2003). Richard Posner, an eminent American legal scholar, offers no numerical estimate but rates the risk “significant” (Posner 2004). Nick Bostrom, in the paper that introduced the concept of existential risk, maintained that assigning a probability of less than 25 percent to existential disaster in this century would be misguided (Bostrom 2002b). It is possible that a publication bias is responsible for these alarming opinions. Presumably, people who think

the threats are severe are more likely to write books on the topic. It is nevertheless unsettling that those who have done research in this area seemingly agree that there is a serious risk that humanity's journey will come to a premature end.

The greatest existential risks arise from human activity. Our species has survived volcano eruptions, meteor impacts and other natural hazards for tens of thousands of years. It seems unlikely that any of these old risks should exterminate us in the near future. By contrast, human civilization is introducing many novel phenomena into the world, ranging from nuclear weapons to designer pathogens to high-energy particle colliders. The most severe existential risks of this century derive from expected technological developments. Advances in biotechnology might make it possible to design new viruses that combine the easy contagion and mutability of influenza with the lethality of HIV. Molecular nanotechnology might make it possible to create weapons systems that dwarf both thermonuclear bombs and biowarfare agents in destructiveness (Drexler 1985). Superintelligent machines might be built, and their actions could determine the future of humanity – and whether we shall have one (Yudkowsky 2007). Many of the existential risks that now seem to be among the greatest were conceptualized only in recent decades, and there might well be others that we have not yet become aware of.

The same technologies that pose these risks will also enable us to reduce some risks. Biotechnology can help us develop better diagnostics, vaccines and anti-viral drugs. Molecular nanotechnology could offer even stronger prophylactics (Freitas 1999). Superintelligent machines would be the last invention that human beings need to make, since a superintelligence would by definition be far more effective than human brains in all intellectual endeavors, including strategic thinking, scientific analysis and technological creativity (Bostrom 1998). In addition to creating and mitigating risks, these powerful technological capabilities would also affect the human condition in many other ways.

Supposing we avoid existential disasters, what then might become of humanity? Looking back, developments such as language, agriculture and perhaps the industrial revolution may be said to have fundamentally changed the human condition. There are at least a thousand times more of us now; and with current world average life expectancy at sixty-seven years we live perhaps three times longer than our Pleistocene predecessors. The mental life of human beings has been transformed by developments such as language, literacy, urbanization, division of labor, industrialization, science, communications, transport, and media technology. What developments can we foresee that would alter the human condition at least as profoundly as these past transitions?

One view is that there will be no fundamental change. Many people appear to hold an implicitly static conception of the human condition. On such a conception, there will surely be changes in politics, culture and gadgetry, but the basic parameters of life and human nature will remain unchanged.

The static view, however, is implausible. It would imply that we have recently arrived at the final human condition, even at a time when things seem to be changing faster than ever. It would also imply a radical break with several long-established trends. If the world economy continues to grow at the same pace as in the last half century, then by 2050 the world will be seven times richer than it is today. World

population is predicted to increase to just over 9 billion in 2050, so average wealth would also increase dramatically (United Nations Population Division 2004). Extrapolating further, by 2100 the world would be almost fifty times richer than today. A single modest-sized country might then have as much wealth as the entire present world.

Over the course of human history, the doubling time of the world economy has been drastically reduced on several occasions, such as in the agricultural transition and the industrial revolution. Should another such transition occur in this century, the world economy might be orders of several magnitudes larger by the end of the century (Hanson 2000).

Another reason for assigning a low probability to the static view is that we can foresee that various specific technological advances will give humans important new capacities. Virtual reality environments will constitute an expanding fraction of our experience. The capability of recording, surveillance, biometrics and data-mining technologies will increase dramatically, making it possible to keep track of what is going on in physical reality to an unprecedented extent (Brin 1998). Nanotechnology will have wide-ranging consequences for manufacturing, medicine and computing. New institutions such as prediction markets might improve the capability of human groups to forecast future developments (Hanson 1995). The impacts of these and other technological developments on the character of human lives are difficult to predict, but that they will have such impacts seems a safe bet.

History shows a long-term trend toward increasing scales of integration of human society: from tribes, to villages, to city states, to kingdoms, nations, empires; and, more recently, regional organizations such as the European Union, and some very partial and limited forms of global governance (Wright 1999). One possibility is that humanity will eventually emerge as a singleton, a world-order where at the highest level there is only one independent agent (Bostrom 2007). A singleton could overcome international coordination problems that now plague our species, such as wars, arms races, and free-rider behavior resulting in underproduction of global public goods (Kaul 1999). It might also increase some risks. In the past, if one country or culture adopted policies that stopped growth, development would continue in other countries which would eventually attain such advantages that they could either invade the laggard country or force it to reform. In a singleton, there would be no outside competitor. Perhaps new technologies for surveillance and law enforcement could also make it immune to internal revolt. Even the direction of evolution could be controlled by a singleton (Bostrom 2005).

Among the most important potential developments are ones that would enable us to alter our biology directly through technological means. Such interventions could affect us more profoundly than modification of beliefs, habits, culture and education. If we learn to control the biochemical processes of human senescence, healthy lifespan could be radically prolonged. A person with the age-specific mortality of a 20-year-old would have a life expectancy of about a thousand years. The ancient but hitherto mostly futile quest for happiness could meet with success if we develop safe and effective methods of controlling the brain circuitry responsible for subjective wellbeing (Pearce 2004). Drugs and other neurotechnologies could make it increasingly feasible for users to shape themselves into the kind of people they want to be – their personality, emotional character, mental energy, romantic attachments and moral character.

Cognitive enhancements might deepen our intellectual lives (Bostrom and Ord 2006, Bostrom and Sandberg 2007).

Those who believe that such developments will not occur should consider whether their skepticism is really about ultimate feasibility or merely about timescales. Some of these technologies will be difficult to develop. Does that give us reason to think that they will never be developed? Not even in fifty years? Two hundred years? Ten thousand years? If we avoid existential catastrophe, humanity could have a long future, and it would seem myopic to assume that human nature will not eventually be technologically transformed into some kind of “posthuman” nature (Bostrom 2003b).

If and when artificial intelligence advances to the point where it matches the human mind in general reasoning abilities, superintelligence is likely to follow swiftly from further improvements in software and hardware (Vinge 1993, Bostrom 1998). The creation of superintelligent machines would be the most momentous event in the history of our species. Humanity’s remoter future might be dominated by artificial minds, our “mind children” (Moravec 1988).

It could be possible for biological human beings to become non-biological by “uploading” their minds to computers. Uploading could be done by gradually replacing parts of the brain with prosthetic chips, or (more likely) by creating a detailed three-dimensional map of the neuronal network in a particular brain and emulating this computational structure on a powerful computer. A human upload could have an indefinitely long lifespan as it would not be subject to biological senescence. Periodic backup copies could be created for security. Speed-up of thought processes would result from implementing the upload on a faster computer, so an upload might, for instance, experience a year of subjective time over the course of one hour. Uploads could live in virtual reality or they could use a robotic body to interact with the physical world. Since uploads could create an unlimited number of copies of themselves, a Malthusian situation could quickly arise unless reproduction were limited (Bostrom 2005, Hanson 1994).

It could also be possible to create vast numbers of conscious computer-simulated people with experiences similar to those typical of an early-twenty-first-century human, raising the possibility that we ourselves might now be inhabiting a computer simulation created by a posthuman civilization. Important coherence constraints on tenable views about the future prospects of our species have recently been derived from this consideration. The so-called Simulation argument purports to show that *either* nearly all human-level civilizations go extinct before becoming posthuman, *or* there is a strong convergence among posthuman civilizations so that almost none of them is interested in creating this kind of ancestor simulation, *or* we are almost certainly living in a computer simulation (Bostrom 2003a).

With machine intelligence and other technologies such as advanced nanotechnology, space colonization should become economical. Such technology would enable us to construct “von Neumann probes” – machines with the capability of traveling to a planet, building a manufacturing base there, and launching multiple new probes to colonize other stars and planets (Tipler 1981). A space colonization race could ensue (Hanson 1998). Over time, the resources of the entire accessible universe might be turned into some kind of infrastructure, perhaps an optimal computing substrate (“computronium”). Viewed from the outside, this process might take a very simple and

predictable form – a sphere of technological structure, centered on its Earthly origin, expanding uniformly in all directions at some significant fraction of the speed of light (Moravec 1999). What happens on the “inside” of this structure – what kinds of lives and experiences (if any) it would sustain – would depend on initial conditions and the dynamics shaping its temporal evolution. It is conceivable, therefore, that the choices we make in this century could have extensive consequences.

References and Further Reading

- Bostrom, N. (1998). “How Long before Superintelligence?,” *International Journal of Futures Studies*, 2.
- Bostrom, N. (2002a). *Anthropic Bias: Observation Selection Effects in Science and Philosophy* (New York: Routledge).
- Bostrom, N. (2002b). “Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards,” *Journal of Evolution and Technology*, 9.
- Bostrom, N. (2003a). “Are You Living in a Computer Simulation?,” *Philosophical Quarterly*, 53 (211): 243–55.
- Bostrom, N. (2003b). *The Transhumanist FAQ: v 2.1*. World Transhumanist Association. Available from <http://transhumanism.org/index.php/WTA/faq/>
- Bostrom, N. (2005). “The Future of Human Evolution,” in C. Tandy (ed.), *Death and Anti-Death* (Palo Alto, Calif.: Ria University Press).
- Bostrom, N. (2007). “What Is a Singleton?,” *Linguistic and Philosophical Investigations*, forthcoming.
- Bostrom, N. and Ord, T. (2006). “The Reversal Test: Eliminating Status Quo Bias in Bioethics,” *Ethics*, 116 (4): 656–80.
- Bostrom, N. and Sandberg, A. (2007). “Cognitive Enhancement: Methods, Ethics, Regulatory Challenges,” *Science and Engineering Ethics*, forthcoming.
- Brin, D. (1998). *The Transparent Society* (Reading, Mass.: Addison-Wesley).
- Drexler, K. E. (1985). *Engines of Creation: The Coming Era of Nanotechnology* (London: Fourth Estate).
- Ehrlich, P. R. (1968). *The Population Bomb* (New York: Ballantine Books).
- Freitas, R. A. (1999). *Nanomedicine* (Austin, Tex.: Landes Bioscience).
- Hanson, R. (1994). “What if Uploads Come First: The Crack of a Future Dawn,” *Extropy*, 6 (2).
- Hanson, R. (1995). “Could Gambling Save Science? Encouraging an Honest Consensus,” *Social Epistemology*, 9 (1): 3–33.
- Hanson, R. (1998). “Burning the Cosmic Commons: Evolutionary Strategies for Interstellar Colonization.” Available from <http://hanson.gmu.edu/filluniv.pdf>
- Hanson, R. (2000). “Long-term Growth as a Sequence of Exponential Modes.” Working manuscript.
- Kaul, I. (1999). *Global Public Goods* (Oxford: Oxford University Press).
- Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcend Biology* (New York: Viking).
- Lang, H. and Gohlen, R. (1985). “Completed Fertility of the Hutterites: A Revision,” *Current Anthropology*, 26 (3): 395.
- Leslie, J. (1996). *The End of the World: The Science and Ethics of Human Extinction* (London: Routledge).
- Maddison, A. and Organisation for Economic Co-operation and Development, Development Centre (2003). *The World Economy: Historical Statistics, Development Centre Studies* (Paris: Development Centre of the Organisation for Economic Co-operation and Development).
- Malthus, T. R. (1798). *An Essay on the Principle of Population* (London: J. Johnson).
- Meadows, D. H. and Club of Rome (1972). *The Limits to Growth: A Report for the Club of Rome’s Project on the Predicament of Mankind* (New York: Universe Books).

- Moravec, H. (1999). *Robot: Mere Machine to Transcendent Mind* (New York: Oxford University Press).
- Moravec, H. P. (1988). *Mind Children: The Future of Robot and Human Intelligence* (Cambridge, Mass.: Harvard University Press).
- Pearce, D. (2004). *The Hedonistic Imperative*. Available from <http://www.hedweb.com/hedab.htm>
- Posner, R. (2004). *Catastrophe* (Oxford: Oxford University Press).
- Rees, M. (2003). *Our Final Hour: A Scientist's Warning. How Terror, Error, and Environmental Disaster Threaten Humankind's Future in This Century – on Earth and Beyond* (New York: Basic Books).
- Tipler, F. (1981). "Extraterrestrial Beings Do Not Exist," *Quarterly Journal of the Royal Astronomical Society*, 21 (267).
- United Nations Population Division (2004). "World Population Prospects: The 2004 Revision," Population Database.
- US Census Bureau (2007). *Total Midyear Population for the World: 1950–2006*. Available from <http://www.census.gov/ipc/www/worldpop.html>
- Vinge, V. (1993). "The Coming Technological Singularity," *Whole Earth Review*, Winter issue.
- Wright, R. (1999). *Nonzero: The Logic of Human Destiny* (New York: Pantheon Books).
- Yudkowsky, E. (2007). "Artificial Intelligence as a Positive and Negative Factor in Global Risk," in N. Bostrom and M. Cirkovic (eds). *Global Catastrophic Risks* (Oxford: Oxford University Press).