HUMAN EXTINCTION RISKS IN THE COSMOLOGICAL AND ASTROBIOLOGICAL CONTEXTS

Milan M. Ćirković
Astronomical Observatory Belgrade
Volgina 7, 11160 Belgrade
Serbia and Montenegro
e-mail: mcirkovic@aob.aob.bg.ac.yu

Abstract. We review the subject of human extinction (in its modern form), with particular emphasis on the natural sub-category of existential risks. Enormous breakthroughs made in recent decades in understanding of our terrestrial and cosmic environments shed new light on this old issue. In addition, our improved understanding of extinction of other species, and the successes of the nascent discipline of astrobiology create a mandate to elucidate the necessary conditions for survival of complex living and/or intelligent systems. A range of topics impacted by this "astrobiological revolution" encompasses such diverse fields as anthropic reasoning, complexity theory, philosophy of mind, or search for extraterrestrial intelligence (SETI). Therefore, we shall attempt to put the issue of human extinction into a wider context of a general astrobiological picture of patterns of life/complex biospheres/intelligence in the Galaxy. For instance, it seems possible to define a secularly evolving *risk function* facing any complex metazoan lifeforms throughout the Galaxy. This multidisciplinary approach offers a credible hope that in the very close future of humanity all natural hazards will be well-understood and effective policies of reducing or eliminating them conceived and successfully implemented. This will, in turn, open way for a new issues dealing with the interaction of sentient beings with its astrophysical environment on truly cosmological scales, issues presciently speculated upon by great thinkers such as H. G. Wells, J. B. S. Haldane or Olaf Stapledon.

Keywords: existential risks, anthropic principle, astrobiology, physical eschatology

1. Introduction: ERs and taxonomies

- *Existential risks* are those where an adverse outcome would either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential (Bostrom 2001).
- It is useful to classify ERs, following Bostrom, according to their relationship with the concept of *posthumanity*, vaguely understood as the future society of technologically enhanced humanity, with much greater intellectual and physical capacities, including life-span. (A sort of Baconian *New Atlantis*, achievable through rational means, posthumanity is to be understood—*for purposes of ER analysis!* – mainly as a placeholder for the total creative potential of us and our descendants.) This taxonomy, rendered colorfully in terms of T. S. Eliot (and John Earman!) is the following:
  - **Bangs**: Earth-originating intelligent life goes extinct in relatively sudden disaster (e.g., an impact of 100 km body on Earth).
  - **Crunches**: The potential of humankind to develop into posthumanity is permanently thwarted (e.g., resource depletion).

- **Shrieks**: Some form of posthumanity is attained, but it is an extremely narrow band of what is possible and desirable (e.g., totalitarian world government is established by first posthuman individuals).
- **Whimpers**: A posthuman civ arises, but evolves in a direction that leads gradually but irrevocably to the complete disappearance of the things we value (e.g., various "slow degeneration" scenarios, like in Schroeder's (2002) *Permanence*).

- However, different taxonomies are possible. The obvious one is into natural and man-made ERs, with the hybrid category gathering those which are not clearly distinguishable:
  - **Natural**, e.g.: cosmic impacts (of asteroidal/cometary bodies), supervolcanism, new natural diseases, supernovae & GRBs.
  - **Human-made**, e.g.: global nuclear war, bioaccidents, nanotechnology risks, artificial intelligence (AI) misuse.
  - **Hybrid**, e.g.: various ecological risks (in particular runaway greenhouse effect), conflict with intelligent extraterrestrial beings, shutdown of the simulation we're hypothetically living in.
- *This is an old theme with a new twist!*

2. So, what's new here?

- Some of the threats have never been even thought of before (e.g., nanotechnology risks).
- Others we understand much better now (e.g., impact hazards).
- Belated appreciation of the anthropic reasoning in general, and the Doomsday Argument in particular.
- The rise of new multidisciplinary fields such as *astrobiology* and *physical eschatology*.
- Increased public awareness (e.g., SL-9; global warming debates).
- The rise of new social and cultural movements (e.g., transhumanism, ecological movements, cyberpunk).
- Overwhelming phenomenon of globalization.
- *All of these (and others, no doubt) are partially overlapping!* Study of ERs is a truly multidisciplinary field, requiring participation of a wide spectrum of specialists.

INTERLUDE I - INADEQUACY OF STANDARD RISK ANALYSIS

Standard risk analysis is painfully inadequate to cope with ERs. Reputedly, during the Manhattan Project, Arthur Compton claimed that an experiment is worth performing if one can show that the risk of global disaster (still thought possible to follow froom nuclear testing at the time) is less than five parts in a million. He did not explain how he reached the figure. How small probability is acceptable if the adverse outcome of doing X is the destruction of humanity and all its values? If, as Pascal among others taught us, the *expectation value* is a true measure of worthiness of a bet, what measure do you assign to present and/or future humanity? Do we count future generations or not? All these issues are mostly untackled on an academic level thus far.

3. Cosmological context

- Anthropic principle(s) as universal selection effect(s); the Doomsday Argument.

- Ongoing cosmological "revolution": no Big Crunch, accelerated expansion, multiverse.
- Signal failure of uniformitarianism in cosmology.

INTERLUDE I - DOOMSDAY ARGUMENT IN A NUTSHELL

One of the most intriguing side issues in discussing the future of humanity is the so-called Doomsday Argument, which was conceived (but not published) by the astrophysicist Brandon Carter in the early 1980s, and it has been first exposed in print by John Leslie in 1989 and in a *Nature* article by Richard Gott (1993). The most comprehensive discussion of the issues involved is Leslie's monograph of 1996, *The End of the World*. The core idea can be expressed through the following urn-ball experiment. Place two large urns in front of you, one of which you know contains ten balls, the other a million, but you do not know which is which. The balls in each urn are numbered 1, 2, 3, 4 ... Now take one ball at random from the left urn; it shows the number 7. This clearly is a strong indication that the left urn contains only ten balls. If the odds originally were fifty-fifty (identically-looking urns), an application of Bayes' theorem gives the posterior probability that the left urn is the one with only ten balls as P_{post} (n=10) = 0.99999. Now consider the case where instead of two urns you have two possible models of humanity, and instead of balls you have human individuals, ranked according to birth order. One model suggests that the human race will soon become extinct (or at least that the number of individuals will be greatly reduced), and as a consequence the total number of humans that ever will have existed is about 100 billion. The other model indicates that humans will colonize other planets, spread through the Galaxy, and continue to exist for many future millennia; we consequently can take the number of humans in this model to be of the order of, say, 10^{18}. As a matter of fact, you happen to find that your rank is about sixty billion. According to Carter and Leslie, we should reason in the same way as we did with the urn balls. That you should have a rank of sixty billion is much more likely if only 100 billion humans ever will have lived than if the number was 10^{18}. Therefore, by Bayes' theorem, you should update your beliefs about mankind's prospects and realize that an impending doomsday is much more probable than you thought previously.

4. Astrobiological context

- Gould's "tiers of time"; general conditions for the existence of the complex metazoan biospheres.
- The "Rare Earth" hypothesis.
- Mass extinctions in the geological past: clues to at least some of ERs.
- Astrobiology as a paradigm of the new multidisciplinary synthesis of sciences offers important *methodological* lessons for studying ERs.

5. Natural hazards: "Armageddon"

- Impact hazard
- Radiation hazard (SNe and GRBs)
- Supervolcanism
- Natural climate change (Milankovich's cycles and all that)
- Naturally occuring diseases
- *Something unforeseen...*

INTERLUDE II – EXISTENTIAL RISK THAT ALREADY OCCURED?
One of the most interesting ideas in recent literature joins human genetics and geo-science to theorize that a supervolcanic eruption of the Toba caldera in Indonesia, about 75,000 years ago has already caused drastic diminishing of the number of then-living humans, mostly because of the "volcanic winter" effects and subsequent lack of food (see S. H. Ambrose, "Late Pleistocene human population bottlenecks, volcanic winter, and differentiation of modern humans," *J. Hum. Evol.* **34**, 623 [1998]) . This is perhaps—and luckily enough!—the only case of an ER realized during the humankind's tenure on Earth; quite possibly, we escaped a "Bang"-like extinction by a hairbreadth. Further research on this is eagerly awaited...

6. Man-made hazards: "Dr. Strangelove"

- Nuclear warfare
- Abuse (intentional or not) of biotechnology
- Abuse (intentional or not) of nanotechnology
- Abuse (intentional or not) of artificial intelligence (AI)
- Vacuum decay, killer strangelets, mini-black holes, and other QFT apocalypses
- *Something unforeseen...*

INTERLUDE II – A LURKING THREAT
The paperback edition of Leslie's *The End of the World* has a rather menacing cover illustration partially hidden behind huge letter of the title title. A note on the back cover reads only cryptically: »Cover photograph: courtesy of NRAO/NSF« without any trace of explanation, as if the content of the image—a vaguely spherical nebulosity of bright green, yellow and red filaments—is obviously clear. Of course, as an astrophysicist by trade, I have recognized the target object: a symmetric (young) supernova remnant, imaged in radio waves. However, it seems highly doubtful to me that an average reader of the book (or even a philosopher!) could easily recognized it, or connect it with some of the dangers to humanity's survival discussed in the book itself.

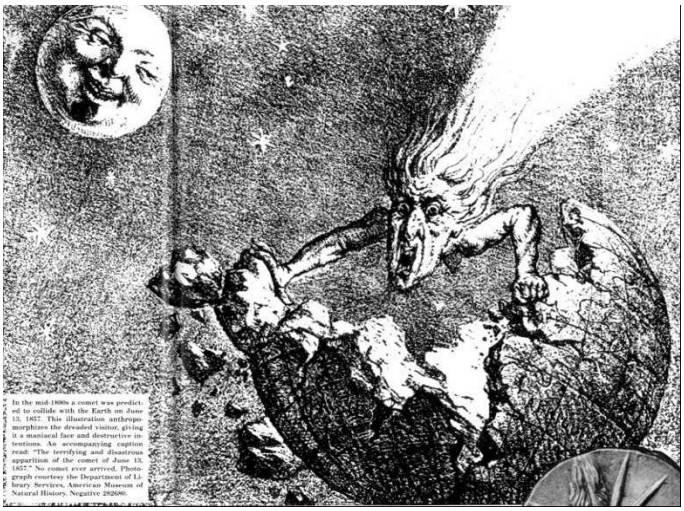
This is ironical, since the threat posed to Earth and humanity (and living beings on inhabited planets in general) by cosmic explosions generally, remains one of the least investigated and most underestimated natural hazards. Even if the timescales for this particular ER are very long in comparison the most of the other hazards, it still might have caused some of the extinction of species in the past of the terrestrial biosphere. In fact, it was this suggestion by great German paleontologist, Otto Schindewolf, which was published in 1962, under the indicative title "Neokatastrophismus?", which was one of the first stabs at the Lyellian uniformitarian dogma. And the recent massive interest in the evil relatives of "normal" supernovae- Gamma-Ray Bursts, whose progenitors are sometimes dubbed *hypernovae*—has already brought renewed suggestions along this line (e.g., Melott et al. http://xxx.lanl.gov/abs/astro-ph/0309415). Moreover, Annis (1999) has suggested a model in which such colossal explosions are the main regulators of intelligent life in the Galaxy, thus explaining the (in)famous Fermi paradox.

7. Hybrid hazards: "Matrix", "Independence Day," and all that

- Runaway greenhouse/other climate calamities – link to geological/ palaeontological controversies.
- The issue of extraterrestrial intelligence; similarities and differences with the issue of AI.
- Simulation argument; the shutdown threat.

8. Instead of conclusions

- ERs in the context of Fermi's question: the Great Filter at work?
- Countering ERs with an effective *policy*!
- Academic research of ERs is necessary in the first place, tightly followed by educational, media, and other actions leading, finally, to policy implementation.



Comets can be dangerous! A XIX century French caricature explaining the etimology of the everyday word **dis**(=evil)**aster**(=star).

RECOMMENDED READING
(quite subjective choice!)

The essential literature on ERs is:

1. Leslie, J. (1996) *The End of the World: The Ethics and Science of Human Extinction*, (Routledge, London).
2. Bostrom, N. (2001) "Existential Risks" *Journal of Evolution and Technology*, vol. **9** (http://www.jetpress.org/index.html).
3. Rees, M. J. (2003) *Our Final Hour* (Basic Books, New York).

Historical roots can be found in:

4. Haldane, J. B. S. (1923) *DAEDALUS* or Science and the Future (Kegan Paul, Trench, Trubner & Co., London; http://www.santafe.edu/~shalizi/Daedalus.html).

5. Russell, B. (1924) *ICARUS or the Future of Science*, (E. P. Dutton, London, 1924; http://www.santafe.edu/~shalizi/Icarus.html).
6. Haldane, J. B. S. (1927) "The Last Judgment" in *Possible Worlds and Other Essays* (Chatto & Windus, London).
7. Eddington, A. S. (1931) "The End of the World: from the Standpoint of Mathematical Physics," *Nature* **127**, 447-453.
8. Whittaker, E. T. (1942) *The Beginning and the End of the World*, (Oxford University Press, Oxford).
9. Schindewolf, O. (1962) "Neokatastrophismus?" *Deutsch Geologische Gesellschaft Zeitschrift Jahrgang* **114**, 430-445.
10. Gould, S. J. (1987) *Time's Arrow, Time's Cycle* (Harvard University Press, Cambridge).

Nascent discipline of physical eschatology:

11. Davies, P. C. W. (1973) "The Thermal Future of the Universe," *Monthly Notices of the Royal Astronomical Society* **161**, 1-5.
12. Barrow, J. D. and Tipler, F. J. (1978) "Eternity is unstable," *Nature* **276**, 453-459.
13. Dyson, F. (1979) "Time without end: Physics and biology in an open universe," *Reviews of Modern Physics* **51**, 447-460.
14. Adams, F. C. and Laughlin, G. (1997) "A dying universe: the long-term fate and evolution of astrophysical objects,", *Reviews of Modern Physics* **69**, 337-372.
15. Krauss, L. M. and Turner, M. S. (1999) "Geometry and Destiny," *General Relativity and Gravitation* **31**, 1453-1459.

Anthropic principle(s) and observation selection effects generally:

16. Bostrom, N. (2002) *Anthropic Bias: Observation Selection Effects in Science and Philosophy* (Routledge, New York).
17. Ćirković, M. M. and Bostrom, N. (2000) "Cosmological Constant and the Final Anthropic Hypothesis," *Astrophysics and Space Science* **274**, 675-687.
18. Carter, B. (1983) "The anthropic principle and its implications for biological evolution," *Philos. Trans. R. Soc. London A* **310**, 347-363. [This one to be read especially critically!]

Doomsday Argument:

19. Leslie, J. (1989) "Risking the World's End," *Bulletin of the Canadian Nuclear Society* **21** (May 1989), 10-15. [The very first exposition of DA in print.]
20. Leslie, J. (1990) "Is the end of the world nigh?" *Philosophical Quarterly* **40**, 65-72.
21. Gott, J. R. (1993) "Implications of the Copernican principle for our future prospects," *Nature* **363**, 315-319.
22. Bostrom, N. (2001) "The Doomsday Argument, Adam & Eve, UN^{++} and Quantum Joe," *Synthese* **127**, 359-387.
23. Olum, K. D. "The doomsday argument and the number of possible observers," *Philosophical Quarterly* **52**, 164-184 (2002).

24. Darling, D. (2001) *Life Everywhere* (Basic Books, New York).
25. Raup, D. M. (1991) *Extinction: Bad Genes or Bad Luck?* (W. W. Norton, New York).
26. Ward, P. D. and Brownlee, D. (2000) *Rare Earth: Why Complex Life Is Uncommon in the Universe* (Springer, New York).
27. Ward, P. D. and Brownlee, D. (2002) *The Life and Death of Planet Earth: How the New Science of Astrobiology Charts the Ultimate Fate of Our World* (Henry Holt and Company, New York).
28. Cockell, C. (2002), "Astrobiology—a new opportunity for interdisciplinary thinking," *Space Policy* **18**, 263-266.
29. Webb, S. (2002) *Where is Everybody? Fifty Solutions to the Fermi's Paradox* (Copernicus, New York).
30. Brin, G. D. (1983) "The 'Great Silence': the Controversy Concerning Extraterrestrial Intelligence," *Q. Jl. R. astr. Soc.*, **24**, 283-309.
31. Annis, J. (1999) "An Astrophysical Explanation for the Great Silence," *J. Brit. Interplan. Soc.*, **52**, 19-22.
32. Scalo, J. and Wheeler, J. C. (2002) "Astrophysical and astrobiological implications of gamma-ray burst properties," *Astrophys. J.* **566**, 723-737.

A tiny bit of relevant SF:

33. Egan, G. (1997) *Diaspora* (Orion/Millennium, London). [A haunting description of the consequences of a close γ-ray burst for Earth.]
34. Schroeder, K. (2002) *Permanence* (Tor Books, New York). [An explication of the "whimper" type scenario.]
35. Lem, S. (1984), *His Master's Voice* (Harvest Books, Fort Washington). [A classic, excellent for many SETI-related issues, among others showing how meme-spreading can lead to the "classical" global nuclear warfare.]
36. Vinge, V. (2000) *A Fire upon the Deep* (Millenium, London). [Originally written in 1991, this space-opera finely envisions all sorts of possibilities following the use and abuse of AI.]