Introduction: Confronting Future Catastrophic Threats to Humanity

Seth D. Baum and Bruce E. Tonn

Futures 72: 1-3. This version 14 October 2015.

http://sethbaum.com * http://gcrinstitute.org

This is the introductory editorial to the Futures special issue "Confronting Future Catastrophic Threats to Humanity".

Humanity faces a range of threats to its viability as a civilization and its very survival. These catastrophic threats include natural disasters such as supervolcano eruptions and large asteroid collisions as well as disasters caused by human activity such as nuclear war and global warming. The threats are diverse, but their would-be result is the same: the collapse of global human civilization or even human extinction.

These diverse threats are increasingly studied as one integrated field, using terms such as existential risk [1], global catastrophic risk [2], global megacrisis [3], ultimate harm [4], survival research [5], and simply catastrophe [6]. Contributions to this field come from diverse intellectual disciplines, including economics [7], philosophy [1], astronomy [8], and risk analysis [9], in addition to the disciplines of each specific risk. Futures studies has been especially active, with futures journals (including this one) hosting several special issues and symposia [3,5,10] and publishing numerous stand-alone articles. The present special issue continues this interdisciplinary, futures-oriented tradition.

The theme of this special issue is confronting future catastrophic threats to humanity. Our motivation for this more practical bent is twofold. First, it is our observation that the bulk of the catastrophic threats literature has thus far focused mainly on philosophical aspects, in particular the moral significance of catastrophic threats and challenges to quantifying their probability, as well as empirical aspects, in particular the nature and size of the various threats. How to confront the threats has gone relatively understudied as an integrated field of research, though there is significant research on how to confront specific threats such as global warming and nuclear war. Second, confronting the threats is arguably the most important part. The threats are not merely academic—they actually do threaten humanity, and so for the sake of humanity they should be confronted. The philosophical and empirical aspects are undoubtedly important, but ultimately people need to know what to do to keep humanity intact.

With this theme in hand, we distributed an open call for papers, inviting submissions from all disciplines and perspectives on the topic. Submissions were evaluated strictly on intellectual merit provided that they fit with the theme; no attempt was made to curate a specific coverage of topics or selection of authors. As a result, the included papers span an eclectic mix of topics, from nuclear war to quantum computing. The range of authors includes several early-career scholars who we hope will continue to make insightful contributions to the catastrophic threats and futures studies fields throughout their careers.

The special issue begins with a paper by Randle and Eckersley on public perceptions of future catastrophic threats. The paper presents a survey given to citizens of Australia, Canada, the United Kingdom, and the United States, making it perhaps the most extensive catastrophic threats survey yet published. The results indicate some significant

recognition of catastrophic threats, with 39% of survey participants giving a 30% or greater probability that "humans will be wiped out in the next 100 years". The results also show wide belief in the need for fundamental change, with 78% agreeing with the statement "we need to transform our worldview and way of life if we are to create a better future for the world." While survey responses do not necessarily translate into action, these responses do suggest that there could be significant public support for efforts to confront catastrophic threats to humanity.

The next three papers all cover threats involving computer technologies. Majot and Yampolskiy present what may be the first-ever analysis of quantum computing as a catastrophic threat. It is an issue of encryption: "Those governments and organizations with quantum technology would be in the god-like position of rendering some current methods of encryption completely obsolete". Able to thwart encryption, they could conduct massive spying on citizens, corporations, and countries, potentially enabling catastrophic totalitarianism and economic chaos. Quantum-safe encryption exists but needs to be better developed and more widely deployed to avoid these problems.

Ćirković connects three seemingly unrelated issues: the threat from advanced artificial intelligence (AI), the potential for global government ("singleton"), and the possibility that humanity exists within a computer simulation. Ćirković posits that running simulations of humanity requires an AI more advanced than would be needed to cause a catastrophe to the non-simulated, biological humanity. Furthermore, avoiding AI catastrophe may require a singleton, to steer all AI development away from dangerous directions. But a singleton could also steer AIs away from humanity-simulations. This makes it less likely that humanity is now in a simulation, and therefore, "if we have reasons to rationally believe that the probability of our living in a simulation is low, we should assign more resources to reducing the chances for, say, catastrophic asteroid impact or a global nuclear war."

Brundage reviews the recent book *Superintelligence*, also on the threat from advanced AI. Brundage is broadly supportive of the book's claim that advanced AI is a major threat, finding that the book "rightfully earns its place on the bookshelf of anyone seriously interested in the future of humanity". One criticism made concerns the book's call for "differential technological development", in which safe AI technologies are favored over dangerous ones. Brundage states that the book "provides little insight into which technologies to differentially accelerate/decelerate or how to do so" while noting that such interventions are complex and easier said than done.

The subsequent three papers present means of helping humanity survive catastrophes. Two papers provide complementary perspectives on the use of refuges to help small human populations survive. Beckstead argues that such refuges would not be a cost-effective intervention, because some refuges and isolated populations already exist, and because refuges may not be particularly helpful for a range of catastrophes. Some catastrophes would leave too many survivors for refuge populations to be important; others would kill refuge populations. For some intermediate catastrophes, refuges might help "by keeping part of the population isolated, or maintaining coordination among enough people with food, material resources, and the right skills", but Beckstead suggests that other interventions are likely more cost-effective.

Baum, Denkenberger, and Haqq-Misra examine potential designs for isolated refuges. They develop the concept of surface-independence, meaning refuges that can maintain

self-sufficiency without interaction with Earth's surface. They propose that surface-independence "would maximize prospects for refuge success", making it "the gold standard of refuge excellence". Surface-independent refuges could be underground, under water, or in outer space. Three design issues are considered: how to provide food for refuge inhabitants, how to keep underground refuges at a comfortable temperature, and how to handle the high cost of extraterrestrial refuges. Each issue has potentially viable solutions, suggesting that surface-independent refuges may be feasible, though without comment on how cost-effective they would be relative to other interventions.

Denkenberger and Pearce consider catastrophes that destroy global agriculture, such as crop pathogens or nuclear winter. They propose that all of humanity could be kept alive for at least five years on food grown from energy from existing vegetation and fossil fuels. Members of other species could also be fed to protect biodiversity. For example, existing trees could feed mushrooms, and natural gas could be fed to certain edible bacteria. These means of food production would need to be scaled up during the catastrophe, so Denkenberger and Pearce analyze rates of ramping up their various food solutions. The analysis finds that "it is technically feasible to feed all humans and save a considerable number of species [which would also be threatened by the catastrophe] in all but the most extreme catastrophes". This result suggests a possible means of confronting a wide class of catastrophic threats.

The two following papers examine solutions for climate-related threats. Baum surveys a range of options for confronting the threat of nuclear winter. The options are in three groups: (1) reducing the probability of nuclear war, such as by diplomacy to resolve international disputes; (2) reducing the severity of nuclear winter, such as by disarmament to reduce the number of nuclear weapons used; and (3) increasing the resilience of humanity to nuclear winter, such as through refuges or novel food production means as discussed elsewhere in this special issue. A core argument is that the threat can be confronted via a wide range of actions by a wide range of actors, and indeed "perhaps everyone in the world is capable of contributing productively".

Haqq-Misra explores the merits of geoengineering to create larger ice caps. The aim here would be to induce a colder and more stable global climate, which could in turn make human civilization more stable. Ice cap geoengineering would be a long-term project, taking millennia of dedicated effort to achieve. As with near-term geoengineering, ice cap geoengineering raises ethical issues regarding what qualifies as a desirable climate. A central issue is how future generations are valued: "If civilization is to survive for millions, rather than thousands, of years into the future, then some degree of climate manipulation to stabilize a cooler climate may be worth consideration".

Finally, Baum presents a practical perspective on the ethics of catastrophic risk. The standard ethical argument for confronting catastrophic threats to humanity is based on the far-future benefits of confronting the threats. Those who do not support this argument may still act on the threats due to (1) near-future benefits from confronting near future threats and (2) ancillary "co-benefits" of the actions unrelated to the threats, especially by "mainstreaming" actions on the threats into existing activities. Baum surveys the threats, finding that "probably a large majority" of the total threat can be confronted with actions that appeal to "what people already care about", and furthermore that these actions "will often be the best to promote, achieving the largest GCR reduction relative to effort spent".

The collection of papers in this special issue shows the wide range of practical steps that can be taken to confront catastrophic threats to humanity. No special skill is needed for some actions, whereas others require advanced knowledge or special social position. Reading these papers makes it clear that the scholarly community is just scratching the surface at identifying and characterizing actions to confront the threats. As a result, only crude generalizations can be made about which actions may be most effective. Clarifying this should be a high priority in order for humanity to have an intelligent response to the catastrophic threats that it faces. The Journal welcomes further papers on the theme of existential threats and related subjects.

References

[1] Bostrom, N., 2013. Existential risk prevention as a global priority. Global Policy 4(1), 15–31.

[2] Bostrom, N., Ćirković, M. (Eds.), 2008. Global Catastrophic Risks. Oxford: Oxford University Press.

[3] Halal, W., & Marien, M. (2011). Global megacrisis: A survey of four scenarios on a pessimism-optimism axis. Journal of Futures Studies 16(2), 65–84.

[4] Persson, I., Savulescu, J., 2012. Unfit for the Future: The Need for Moral Enhancement. Oxford: Oxford University Press.

[5] Seidel, P. (Ed.), 2003. Survival research [special issue]. World Futures, 59, 125-318.

[6] Posner, R., 2004. Catastrophe: Risk and Response. Oxford: Oxford University Press.

[7] Wagner, G., Weitzman, M.L., 2015. Climate Shock. Princeton: Princeton University Press.

[8] Rees, M., 2003. Our Final Century: Will the Human Race Survive the Twenty-first Century? Oxford: William Heinemann.

[9] Matheny, J.G., 2007. Reducing the risk of human extinction. Risk Analysis vol. 27, no. 5, pages 1335-1344.

[10] Tonn, B., MacGregor, D. (Eds.), 2009. Human extinction [special issue]. Futures 41(10), 673-774.