Global Catastrophes: The Most Extreme Risks

Seth D. Baum and Anthony M. Barrett
Global Catastrophic Risk Institute
http://sethbaum.com * http://tony-barrett.com * http://gcrinstitute.org

In Vicki Bier (Ed.), 2018. *Risk in Extreme Environments: Preparing, Avoiding, Mitigating, and Managing*. New York: Routledge, pages 174-184. This version 2 October 2017.

Abstract

The most extreme risk are those that threaten the entirety of human civilization, known as global catastrophic risks. The very extreme nature of global catastrophes makes them both challenging to analyze and important to address. They are challenging to analyze because they are largely unprecedented and because they involve the entire global human system. They are important to address because they threaten everyone around the world and future generations. Global catastrophic risks also pose some deep dilemmas. One dilemma occurs when actions to reduce global catastrophic risk could harm society in other ways, as in the case of geoengineering to reduce catastrophic climate change risk. Another dilemma occurs when reducing one global catastrophic risk could increase another, as in the case of nuclear power reducing climate change risk while increasing risks from nuclear weapons. The complex, interrelated nature of global catastrophic risk suggests a research agenda in which the full space of risks are assessed in an integrated fashion in consideration of the deep dilemmas and other challenges they pose. Such an agenda can help identify the best ways to manage these most extreme risks and keep human civilization safe.

1. Introduction

The most extreme type of risk is the risk of a global catastrophe causing permanent worldwide destruction to human civilization. In the most extreme cases, human extinction could occur. Global catastrophic risk (GCR) is thus risk of events of the highest magnitude of consequences, and the risks merit serious attention even if the probabilities of such events are low. Indeed, a growing chorus of scholars rates GCR reduction as among the most important priorities for society today. Unfortunately, many analysts also estimate frighteningly high probabilities of global catastrophe, with one even stating "I think the odds are no better than fifty-fifty that our present civilization on Earth will survive to the end of the present century" (Rees 2003:8).

Regardless of what the probabilities are, it is clear that humanity today faces a variety of serious GCRs. To an extent, humanity always has faced GCRs, in the form of supervolcano eruptions, impacts from large asteroids and comets, and remnants of stellar explosions. Events like these have contributed to several mass extinction events across Earth's history. The Toba volcano eruption about 70,000 years ago may have come close to bringing the human species to a premature end. However, scholars of GCR generally believe that today's greatest risks derive from human activity. These GCRs include war with biological or nuclear weapons, extreme climate change and other environmental threats, and misuse or accidents involving powerful emerging technologies like artificial intelligence and synthetic biology. These GCRs threaten far greater destruction than was seen in the World Wars, the 1918 flu, the Black Death plague, or other major catastrophes of recent memory.

The high stakes and urgent threats of GCR demand careful analysis of the risks and the opportunities for addressing them. However, several factors make GCR difficult to analyze. One factor is the unprecedented nature of global catastrophes. Many of the catastrophes have never occurred in any form, and of course no previous global catastrophe has ever destroyed modern human civilization. The lack of precedent means that analysts cannot rely on historical data as much as they can for smaller, more frequent events. Another factor is the complexity of GCRs, involving global economic, political, and industrial systems, which present difficult analytical decisions about which details to include. Finally, the high stakes of GCR pose difficult dilemmas about the extent to which GCR reduction should be prioritized relative to other issues.

In this paper we present an overview of contemporary GCR scholarship and related issues for risk analysis and risk management. We focus less on the risks themselves, each of which merits its own dedicated treatment. Other references are recommended for the risks, perhaps the best of which are the relevant chapters of Bostrom and Ćirković (2008). Instead, our focus here is on overarching themes of importance to the breadth of the GCRs. The following section defines GCR in more detail and explains why many researchers consider it to be so important. Next, some of the analytical challenges that GCR poses and the techniques that have been developed to meet these challenges are explained. There follows a discussion of some dilemmas that arise when GCR reduction would require great sacrifice or would interfere with each other. Finally, conclusions are drawn.

2. What Is GCR And Why Is It Important?

Taken literally, a global catastrophe can be any event that is in some way catastrophic across the globe. This suggests a rather low threshold for what counts as a global catastrophe. An event causing just one death on each continent (say, from a jet-setting assassin) could rate as a global catastrophe, because surely these deaths would be catastrophic for the deceased and their loved ones. However, in common usage, a global catastrophe would be catastrophic for a significant portion of the globe. Minimum thresholds have variously been set around ten thousand to ten million deaths or $10 billion to $10 trillion in damages (Bostrom and Ćirković 2008), or death of one quarter of the human population (Atkinson 1999; Hempsell 2004). Others have emphasized catastrophes that cause long-term declines in the trajectory of human civilization (Beckstead 2013), that human civilization does not recover from (Maher and Baum 2013), that drastically reduce humanity's potential for future achievements (Bostrom 2002, using the term "existential risk"), or that result in human extinction (Matheny 2007; Posner 2004).

A common theme across all these treatments of GCR is that some catastrophes are vastly more important than others. Carl Sagan was perhaps the first to recognize this, in his commentary on nuclear winter (Sagan 1983). Without nuclear winter, a global nuclear war might kill several hundred million people. This is obviously a major catastrophe, but humanity would presumably carry on. However, with nuclear winter, per Sagan, humanity could go extinct. The loss would be not just an additional four billion or so deaths, but the loss of all future generations. To paraphrase Sagan, the loss would be *billions and billions* of lives, or even more. Sagan estimated 500 trillion lives, assuming humanity would continue for ten million more years, which he cited as typical for a successful species.

Sagan's 500 trillion number may even be an underestimate. The analysis here takes an adventurous turn, hinging on the evolution of the human species and the long-term fate of the universe. On these long time scales, the descendants of contemporary humans may no longer be recognizably "human". The issue then is whether the descendants are still worth caring about, whatever they are. If they are, then it begs the question of how many of them there will be. Barring major global catastrophe, Earth will remain habitable for about one billion more years until the Sun gets too warm and large. The rest of the Solar System, Milky Way galaxy, universe, and (if it exists) the multiverse will remain habitable for a lot longer than that (Adams and Laughlin 1997), should our descendants gain the capacity to migrate there. An open question in astronomy is whether it is possible for the descendants of humanity to continue living for an infinite length of time or instead merely an astronomically large but finite length of time (see e.g. Ćirković 2002; Kaku 2005). Either way, the stakes with global catastrophes could be much larger than the loss of 500 trillion lives.

Debates about the infinite vs. the merely astronomical are of theoretical interest (Ng 1991; Bossert et al. 2007), but they have limited practical significance. This can be seen when evaluating GCRs from a standard risk-equals-probability-times-magnitude framework. Using Sagan's 500 trillion lives estimate, it follows that reducing the probability of global catastrophe by a mere one-in-500-trillion chance is of the same significance as saving one human life. Phrased differently, society should try 500 trillion times harder to prevent a global catastrophe than it should to save a person's life. Or, preventing one million deaths is equivalent to a one-in-500-million reduction in the probability of global catastrophe. This suggests society should make extremely large investment in GCR reduction, at the expense of virtually all other objectives.

Judge and legal scholar Richard Posner made a similar point in monetary terms (Posner 2004). Posner used $50,000 as the value of a statistical human life (VSL) and 12 billion humans as the total loss of life (double the 2004 world population); he describes both figures as significant underestimates. Multiplying them gives $600 trillion as an underestimate of the value of preventing global catastrophe. For comparison, the United States government typically uses a VSL of around one to ten million dollars (Robinson 2007). Multiplying a $10 million VSL with 500 trillion lives gives $$5 \times 10^{21}$ as the value of preventing global catastrophe. But even using "just" $600 trillion, society should be willing to spend at least that much to prevent a global catastrophe, which converts to being willing to spend at least $1 million for a one-in-500-million reduction in the probability of global catastrophe. Thus while reasonable disagreement exists on how large of a VSL to use and how much to count future generations, even low-end positions suggest vast resource allocations should be redirected to reducing GCR. This conclusion is only strengthened when considering the astronomical size of the stakes, but the same point holds either way. The bottom line is that, as long as something along the lines of the standard risk-equals-probability-times-magnitude framework is being used, then even tiny GCR reductions merit significant effort. This point holds especially strongly for risks of catastrophes that would cause permanent harm to global human civilization.

The discussion thus far has assumed that all human lives are valued equally. This assumption is not universally held. People often value some people more than others, favoring themselves, their family and friends, their compatriots, their generation, or others whom they identify with. Great debates rage on across moral philosophy, economics, and other fields about how much people should value others who are distant in space, time, or social relation, as well as the unborn members of future generations. This debate is crucial for all valuations of risk, including GCR. Indeed, if each of us only cares about our immediate selves, then global catastrophes may not be especially important, and we probably have better things to do with our time than worry about them.

While everyone has the right to their own views and feelings, we find that the strongest arguments are for the widely held position that all human lives should be valued equally. This position is succinctly stated in the United States Declaration of Independence, updated in the 1848 Declaration of Sentiments: "We hold these truths to be self-evident: that all men and women are created equal". Philosophers speak of an agent-neutral, objective "view from nowhere" (Nagel 1986) or a "veil of ignorance" (Rawls 1971) in which each person considers what is best for society irrespective of which member of society they happen to be. Such a perspective suggests valuing everyone equally, regardless of who they are or where or when they live. This in turn suggests a very high value for reducing GCR, or a high degree of priority for GCR reduction efforts.

3. Challenges To Analyzing GCR

Given the goal of reducing GCR, one must know what the risks are and how they can be reduced. This requires diving into the details of the risks themselves—details that we largely skip in this paper—but it also requires attention to a few analytical challenges.

The first challenge is the largely unprecedented nature of global catastrophes. Simply put, modern human civilization has never before ended. There have been several recent global catastrophes of some significance, the World Wars and the 1918 flu among them, but these clearly did not knock civilization out. Earlier catastrophes, including the prehistoric mass extinction events, the Toba volcanic eruption, and even the Black Death plague, all occurred before modern civilization existed. The GCR analyst is thus left to study risks of events that are in some way untested or unproven. But the lack of historical precedent does not necessarily imply a lack of ongoing risk. Indeed, the biggest mistake of naïve GCR analysis is to posit that, because no global catastrophe has previously occurred, therefore none will occur. This mistake comes in at least three forms.

The first and most obviously false form is to claim that unprecedented events never occur. In our world of social and technological innovation, it is easy to see that this claim is false. But accounting for it in risk analysis still requires some care. One approach is to use what is known in probability theory as zero-failure data (Hanley 1983; Bailey 1997; Quigley and Revie 2011). Suppose that no catastrophe has occurred over n prior time periods—for example, there has been no nuclear war in the 65 years since two countries have had nuclear weapons. (The second country to build nuclear weapons was the Soviet Union, in 1949.) It can thus be said that there have been zero failures of nuclear deterrence in 65 cases. An approximate upper bound can then be estimated for the probability p of nuclear deterrence failure, i.e. the probability of nuclear war, occurring within an upcoming year. Specifically, p lies within the interval $[0, u]$ with $(1 - \alpha)$ confidence, where $u = 1 - \alpha^{(1/n)}$ gives the upper limit of the confidence interval. Thus for 95% confidence ($\alpha = 0.05$), $u = 1\text{-}0.05^{(1/65)} = 0.05$, meaning that there is a 95% chance that the probability of nuclear war within an upcoming year is somewhere between 0 and 0.05. Note that this calculation assumes (perhaps erroneously) that the 65 non-failures are independent random trials and that p is approximately constant over time, but it nonetheless provides a starting point for estimating the probability of unprecedented events. Barrett et al. (2013) uses a similar approach as part of a validation check of a broader risk analysis of U.S.-Russia nuclear war.

The second form of the mistake is to posit that the ongoing existence of human civilization proves that global catastrophes will not occur. It is true that civilization's continued existence despite some past threats should provide some comfort, but it should only provide *some* comfort. Consider this: if a global catastrophe had previously occurred, nobody would still be around to ponder the matter (at least for catastrophes causing human extinction). The fact of being able to observe one's continued survival is contingent upon having survived. While it is easy to see that this is a mistake, it is harder to correct for it. Again, it requires careful application of probability theory, correcting for what is known as an observation selection effect (Bostrom 2002b, Ćirković et al. 2010). The basic idea is to build the existence of the observer into probability estimates for catastrophes that would eliminate future observers. The result is probability estimates unbiased by the observer's existence, with global catastrophe probability estimates typically revised upwards.

The third form of the mistake is to posit that, because humanity has survived previous catastrophes, or risks of catastrophes, therefore it will survive future ones. This mistake is especially pervasive in discussions of nuclear war. People sometimes observe that no nuclear war has ever occurred and cite this as evidence to conclude that therefore nuclear deterrence and the fear of mutually assured destruction will indefinitely continue to keep the world safe (for discussion see Sagan and Waltz 2013). But there have been several near misses, from the 1962 Cuban missile crisis to the 1995 Norwegian rocket incident, and there is no guarantee that nuclear war will be avoided into the distant future. Similarly, just because no pandemic has ever killed the majority of people (Black Death killed about 22%), or just because early predictions about the rise of artificial intelligence proved false (they expected human-level AI within decades that have long since come and gone; see Crevier 1993; McCorduck 2004), it does not necessarily follow that no pandemics would be so lethal, or that AI cannot reach the lofty heights of the early predictions.

Careful risk analysis can correct for the third form by looking at the full sequences of events that would lead to particular global catastrophes. For example, nuclear weapons in the United States are launched following a sequence of decisions by increasingly high ranking officials, ultimately including the President. This decision sequence can be built into a risk model, with model parameters estimated from historical data on how often each step in the decision sequence has been reached (Barrett et al. 2013). The more often near misses have occurred, and the nearer the misses were, the higher the probability of an eventual "hit" in the form of a nuclear war. The same analytic structure can be applied to other GCRs.

But for many aspects of GCRs, as with other low-probability risks, there is not enough historical or other empirical data to fully characterize the risk. A good example of this is the risk from AI. The concern is that AI with human-level or super-human intelligence could outsmart humanity, assume control of the planet, and inadvertently cause global catastrophe while pursuing whatever objectives it was initially programmed for (Omohundro 2008, Yudkowsky 2008). While there is reason to take this risk seriously—and indeed many do—assessing the risk cannot rely exclusively on empirical data, because no AI like this has ever existed. Characterizing AI risk thus requires expert judgment to supplement whatever empirical data is available (Baum et al. 2011). And while experts, like everyone else, are prone to make mistakes in making predictions and estimating the nature of the world, careful elicitation of expert judgment can reduce these mistakes and improve the overall risk analysis.

That said, for GCR analysis it can be especially important to remember the possibility of experts being wrong. Indeed, for very low probability GCRs, this possibility can dominate the analysis, even when experts have wide consensus and high confidence in their conclusions, and even when the conclusions have significant theoretical and empirical basis (Ord et al. 2010). It can be similarly important to remember the possibility that experts with outlier opinions can be right (Ćirković 2012). Ordinarily, these possibilities would not merit significant attention, but the high stakes of GCR means that even remote possibilities can warrant at least some scrutiny.

A different type of analytical challenge comes from the global nature of GCRs, which makes them especially complex risks to analyze. GCRs are driven variously by the biggest geopolitical rivalries (in the case of biological or nuclear war), advanced research and development and the advantages it can confer (in the case of emerging technologies), or the entire global industrial economy (in the case of environmental collapses). Likewise, the impacts of global catastrophes depend on the resilience of global human civilization to major systemic shocks, potentially including major losses of civil infrastructure, manufacturing, agriculture, trade, and other basic institutions that enable the existence and comforts of modern civilization (Maher and Baum 2013). Assessing the extent of GCR requires accounting for the complexities all these disparate factors plus many others.

Of course it is not possible to include every detail in any risk analysis, and certainly not for global risks. One must always focus on the most important parts. Here it is helpful to recall the high stakes associated with the most severe global catastrophes: the ones that would cause permanent harm to human civilization. While these catastrophes can be highly multifaceted, with a wide variety of impacts, the one impact that stands out as particularly important is that permanent harm. Other impacts, and the causes of those impacts, are simply less important. A GCR analysis can focus on the permanent harm and its causes.

Climate change is an excellent example of a highly complex GCR. Climate change is caused mainly by a wide range of industrial, agricultural, deforestation, and transportation activities, which in turn are connected to a large portion of the activities that people worldwide do on a daily basis. The impacts of climate change are equally vast, affecting meteorological, ecological, and human systems worldwide, causing everything from increased storm surge to increased incidence of malaria. Each of these various impacts is important to certain people and certain ecosystems, but most of them will not make or break humanity's success as a civilization. Instead, the GCR analyst can look directly at worst-case global catastrophe scenarios, such as the possibility of temperatures exceeding the thermal limits of mammals across much of the world, in which case mammals in those regions not in air conditioning will overheat and die (Shewood and Huber 2010). Thus a focus on GCR can make the overall analysis easier.

A subtler complexity, which GCR scholarship is only just starting to address, is the systemic nature of certain GCRs. Most GCR scholarship treats each risk as distinct: there could be a nuclear war, *or* there could be catastrophic climate change, *or* there could be an AI catastrophe, and so on. But these risks do not exist in isolation. They may be caused by the same phenomena, such as a quest for economic growth (causing both climate change, via industrial activity, and AI, via commercial technology development). Or they may cause each other, such as in the concern that climate change will lead to increased violent conflict (Gemenne et al. 2014). They may also have interacting impacts, such as if a war or other catastrophe hits a population already weakened by climate change. These various interactions suggest a systems approach to GCR analysis (Baum et al. 2013; Hanson 2008; Tonn and MacGregor 2009), just as interactions among other risks suggest a systems approach to risk analysis in general (Haimes 2012; Park et al. 2013).

Systemic effects further suggests that global catastrophe could be caused by relatively small events whose impacts cascade into a global catastrophe. Similar systemic effects can already be seen across a variety of contexts, such as the 2003 Italy power outage caused by trees hitting two power lines in Switzerland, with effects cascading across the whole system (Buldyrev et al. 2010). Just as these systems proved fragile to certain small disturbances, perhaps the global human civilization could too. Characterizing these global systemic risks can give a clearer understanding of the totality of the GCRs that human civilization now faces, and can also help identify some important opportunities to reduce or otherwise manage the risks.

4. Some GCR Dilemmas

Unfortunately, managing GCR is not always as simple as analyzing the risks and identifying the risk management options. Some GCR management options pose deep dilemmas that are not easily resolved, even given full information about the risks. The bottom line is that no matter how successful GCR analysis gets, society still faces some difficult decisions.

One dilemma arises from the very high stakes of GCR, or rather the very high magnitude of damages associated with permanent harm to human civilization. The high magnitude suggests that GCR reduction efforts should be prioritized over many other possible courses of action. Sometimes prioritizing GCR reduction efforts is not a significant concern, when the efforts would not be difficult or when they would be worth doing anyway, such as reducing climate change risk by improving energy efficiency. However, sometimes GCR reductions would come at a significant cost. In these cases society may think twice about whether the GCR reductions are worth it, even if the GCR reductions arguably should take priority given the high magnitude of global catastrophe damages.

An example of such a dilemma can be found for climate change and other environmental risks. Because these risks are driven by such a large portion of human activity, reducing the risks could mean curtailing quite a lot of these activities. Society may need to, among other things, build new solar and wind power supplies, redesign cities for pedestrians and mass transit, restructure subsidies for the agriculture and energy sectors, and accept a lower rate of economic growth. Individuals may need to, among other things, change their diets, modes of transportations, places of residence, and accept a simpler lifestyle. Such extensive efforts may pass a cost-benefit test (Stern 2007), especially after accounting for the possibility of global catastrophe (Weitzman 2009), but many people may still not want to do them. Indeed, despite the increasingly stark picture painted by climate change research, the issue still does not rank highly on the public agenda (Pew 2014). Should aggressive effort nonetheless be taken to reduce greenhouse gas emissions and protect the environment? This is a difficult dilemma.

A similar dilemma arises for one proposed solution to climate change: geoengineering. Geoengineering is the intentional manipulation of the global Earth system (Caldeira et al. 2013). A prominent form of geoengineering would inject aerosol particles into the stratosphere in order to block incoming sunlight, thereby reducing temperatures at the surface. While this stratospheric geoengineering could not perfectly compensate for the climatic changes from greenhouse gas emissions, it probably could help avoid some of the worst damages, such as the damages of exceeding the thermal limits of mammals.

However, stratospheric geoengineering comes with its own risks. In particular, if society stops injecting particles into the stratosphere, then temperatures rapidly rise back to where they would have been without any geoengineering (Matthews and Caldeira 2007). The rapid temperature increase could be very difficult to adapt to, potentially causing an even larger catastrophe than regular climate change. This creates a dilemma: Should society run the risk of geoengineering catastrophe, or should it instead endure the pains of regular climate change (Baum et al. 2013)? Given how bad climate change could get, this makes for another difficult dilemma.

The high stakes of GCR suggest that society should do whatever would minimize GCR, and accept whatever suffering might follow. This could mean taking aggressive action to protect the environment, or, if that does not happen, suffering through climate change instead of attempting geoengineering. Looking at the analysis on paper, it is easy to recommend minimizing GCR: the numbers simply point heavily in that direction. But in practice, this would not be an easy recommendation to make, asking many people to make such a great sacrifice. Hopefully, clever solutions can be found that will avoid these big dilemmas, but society should be prepared to make difficult choices if need be.

Another type of dilemma occurs when multiple GCRs interact with each other. Sometimes one action can reduce multiple GCRs. However, sometimes an action would reduce one GCR while increasing another. This poses a dilemma between the two GCRs, a risk-risk tradeoff (Graham and Wiener 1995). A good example of this type of dilemma is nuclear power. Nuclear power can help reduce climate change risk by shifting electricity production away from coal. However, nuclear power can also increase nuclear war risk by enabling nuclear weapons proliferation. This dilemma is seen most clearly in ongoing debates about the nuclear program of Iran. While Iran claims that its program is strictly for peaceful electricity and medical purposes, many observers across the international community believe that Iran is using its program to build nuclear weapons.

Given the risks from climate change and nuclear war, should nuclear power be promoted? How much should it be promoted, and in what circumstances? Resolving these dilemmas requires quantifying and comparing the two GCRs to identify when nuclear power would result in a net decrease in GCR. Unfortunately, at this time GCR research has not quantified climate change and nuclear war risk well enough to be able to make the comparison and reach conclusions about nuclear power. Meanwhile, in circumstances in which nuclear power would not create a nuclear weapons proliferation risk, such as for countries that already have nuclear weapons or clearly do not want them, nuclear power probably would bring a net GCR reduction. This conclusion brings up the first type of dilemma – the general sacrifice for GCR reduction – where nuclear power raises concerns about industrial accidents like Chernobyl or Fukushima. Such accidents are “only” local (or regional) catastrophes, but they are nonetheless plenty large enough to weigh heavily in decision making.

5. Conclusions: A Research Agenda

Regardless of whether GCR reduction should be prioritized above everything else, it is clear that GCR is an important issue and that reducing GCR merits some effort. The big research question then is, which efforts? That is, what are the best, most effective, most desirable ways to reduce GCR? Unfortunately, the GCR research community has not yet made significant advances to answering this vital question. A new research agenda is needed for it.

We believe that GCR research is most helpful at guiding GCR reduction efforts when the research covers all the major risks and risk reduction options in a consistent, transparent, and rigorous manner. It should to include all the major risks and risk reduction options in order to identify which ones are most important and most worth pursuing. Analyzing each risk in isolation fails to account for their various systemic interactions and prevents evaluating risk-risk dilemmas like that posed by nuclear power. In contrast, integrating all the risks and risk reduction options into one assessment can help decision makers identify the best options. Similar integrated assessments have been done for a variety of other topics, such as the popular economic assessments of climate change (e.g. Nordhaus 2008). Something along these lines, adapted for the particulars of GCR, would be of great value to GCR reduction decision making.

An integrated assessment of GCR poses its own analytical challenges. The particulars of different GCRs can be quite different from each other. Fitting together e.g. a climate model, an epidemiological model, and a technological development model would not be easy, nor would filling in the important gaps that inevitably appear between the models. Each GCR is full of rich complexity; the full system of GCRs is more complex still. This makes it even more important to focus on the most important aspects of GCR, lest the analysis get bogged down in details.

It is equally important for the analysis to focus on risk reduction options that are consistent with what attentive decision makers are in a position to do. Otherwise the analysis risks irrelevance. While this is true for any analysis, it is especially true for GCR. The global scale of GCR makes it tempting for analysis to ignore details that seem small relative to the risk but are important to decision makers, and also to entertain risk reduction options that perform well in a theoretically ideal world, "if only everyone would do that". Furthermore, the high stakes of GCR makes it tempting for analysis to recommend rather drastic actions that go beyond what most people are willing to do. It is thus that much more important for GCR analysis to remain in close touch with actual decision makers, to ensure that the analysis can help inform actual decisions.

Despite these challenges, we believe that such a research agenda is feasible and can make an important contribution to society's overall efforts to reduce GCR. Indeed, the future of civilization could depend on it.

References

Adams, F.C. and Laughlin G., 1997. A dying universe: the long-term fate and evolution of astrophysical objects. *Reviews of Modern Physics*, 69(2), p.337-72.

Atkinson, A., 1999. *Impact Earth: asteroids, comets and meteors—the growing threat*. London: Virgin.

Barrett, A.M., Baum, S.D. and Hostetler, K.R., 2013. Analyzing and reducing the risks of inadvertent nuclear war between the United States and Russia. *Science and Global Security*, 21(2), p.106-33.

Baum, S.D., Goertzel, B., and Goertzel, T.G., 2011. How long until human-level AI? results from an expert assessment. *Technological Forecasting & Social Change*, 78(1), p.185-95.

Baum, S.D., Maher, T.M. Jr., and Haqq-Misra, J., 2013. Double catastrophe: intermittent stratospheric geoengineering induced by societal collapse. *Environment, Systems, and Decisions*, 33(1), p.168-80.

Bailey, R.T., 1997. Estimation from zero-failure data. *Risk Analysis*, 17(3), p.375-80.

Beckstead, N., 2013. *On the overwhelming importance of shaping the far future*. Doctoral Dissertation, Department of Philosophy, Rutgers University.

Bossert, W., Sprumont, Y., Suzumura, K., 2007. Ordering infinite utility streams. *Journal of Economic Theory*, 135, p.579-89.

Bostrom, N., 2002. Existential risks: analyzing human extinction scenarios and related hazards. *Journal of Evolution and Technology*, 9.

Bostrom, N., 2002b. *Anthropic bias*. New York, Routledge. Bostrom,

Bostrom, N. and Ćirković, M., 2008. *Global Catastrophic Risks*. Oxford: Oxford University Press.

Buldyrev, S.V., Parshani, R., Paul, G., Stanley, H.E., and Havlin, S., 2010. Catastrophic cascade of failures in interdependent networks. *Nature*, 464, 1025-28.

Caldeira, K., Bala, G., and Cao, L., 2013. The science of geoengineering. *Annual Review of Earth and Planetary Sciences*, 41, p.231-56.

Ćirković, M.M., 2002. Cosmological forecast and its practical significance. *Journal of Evolution and Technology*, 7.

Ćirković, M.M., Sandberg, A., and Bostrom, N., 2010. Anthropic shadow: observation selection effects and human extinction risks. *Risk Analysis*, 30(10), p.1495-506.

Crevier, D., 1993. *AI: the tumultuous history of the search for artificial intelligence*. New York: Basic Books.

Gemenne, F., Barnett, J., Adger, W.N., Dabelko, G.D., 2014. Climate and security: evidence, emerging risks, and a new agenda. *Climatic Change*, 123(1), p.1-9.

Graham, J.D., and Wiener, J.B., editors, 1995. *Risk vs. risk: tradeoffs in protecting health and the environment*. Cambrdge, MA: Harvard University Press.

Haimes, Y.Y., 2012. Systems-based guiding principles for risk modeling, planning, assessment, management, and communication. *Risk Analysis*, 32(9), p.1451-67.

Hanley, J. A. and A. Lippman-Hand (1983). "If Nothing Goes Wrong, Is Everything All Right?" Journal of the American Medical Association 249(13): 1743-1745.

Hanson R., 2008. Catastrophe, social collapse, and human extinction. In: N. Bostrom, M.M. Ćirković, eds. *Global catastrophic risks*. Oxford: Oxford University Press. p.363-78.

Hempsell, C.M., 2004. The potential for space intervention in global catastrophes. *Journal of the British Interplanetary Society*, 57, p.14-21.

Kaku, M., 2005. Escape from the universe. *Prospect Magazine*, February 20. www.prospectmagazine.co.uk/magazine/escapefromtheuniverse

Maher, T.M. Jr., and Baum, S.D., 2013. Adaptation to and recovery from global catastrophe. *Sustainability*, 5(4), p.1461-79.

Matheny, J.G., 2007. Reducing the risk of human extinction. *Risk Analysis*, 27(5), p. 1335-44.

Matthews, H.D., and Caldeira, K., 2007. Transient climate–carbon simulations of planetary geoengineering. *Proceedings of the National Academy of Sciences*, 104, p.9949-54.

McCorduck, P., 2004. *Machines who think: 25th anniversary edition*. Natick, MA: A.K. Peters.

Nagel, T., 1986. *The view from nowhere*. Oxford: Oxford University Press.

Ng, Y.-K., 1991. Should we be very cautious or extremely cautious on measures that may involve our destruction? *Social Choice and Welfare*, 8, p.79-88.

Nordhaus, W.D., 2008. *A question of balance: weighing the options on global warming policies*. New Haven: Yale University Press.

Omohundro, S.M., 2008. The basic AI drives. In P. Wang, B. Goertzel and S. Franklin, eds. *Artificial General Intelligence 2008: Proceedings of the First AGI Conference*. Amsterdam: IOS Press. p.483-92.

Ord, T., Hillerbrand, R., and Sandberg, A., 2010. Probing the improbable: methodological challenges for risks with low probabilities and high stakes. *Journal of Risk Research*, 13(2), p.191-205.

Park J., Seager, T.P., Rao, P.S.C., Convertino, M., and Linkov, I., 2013. Integrating risk and resilience approaches to catastrophe management in engineering systems. *Risk Analysis*, 33(3), p.356-67.

Pew, 2014. Thirteen years of the public's top priorities. Washington, DC: The Pew Research Center For The People & The Press, 27 January. http://www.people-press.org/interactives/top-priorities/

Posner, R., 2004. *Catastrophe: risk and response*. Oxford: Oxford University Press.

Quigley, J. and Revie M., 2011. Estimating the probability of rare events: addressing zero failure data. *Risk Analysis*, 31(7), p.1120-32.

Rawls, J., 1971. *A theory of justice*. Cambridge, MA: Harvard University Press.

Rees, M., 2003. *Our final century: will the human race survive the twenty-first century?* Oxford: William Heinemann.

Robinson, L.A., 2007. How US government agencies value mortality risk reductions. *Review of Environmental Economics and Policy*, 1(2), p.283-99.

Sagan, C., 1983. Nuclear war and climatic catastrophe: some policy implications. *Foreign Affairs*, 62(2), p.257-92.

Sagan, S.D., and Waltz, K.N., 2013. *The spread of nuclear weapons: an enduring debate, third ed*. New York: W.W. Norton.

Sherwood, S.C., and Huber, M., 2010. An adaptability limit to climate change due to heat stress. *Proceedings of the National Academy of Sciences*, 107, p.9552-5.

Stern, N.H., 2007. *The economics of climate change: the Stern review*. Cambridge, UK: Cambridge University Press.

Tonn, B., MacGregor, D., 2009. A singular chain of events. *Futures*, 41, p.706-14.

Weitzman, M.L., 2009. Structural uncertainty and the value of statistical life in the economics of catastrophic climate change. *Review of Economics and Statistics*, 91(1), p.1-19.

Yudkowsky, E., 2008. Artificial intelligence as a positive and negative factor in global risk. Global Catastrophic Risks. In: N. Bostrom, M.M. Ćirković, eds. *Global catastrophic risks*. Oxford: Oxford University Press. p.308-45.