The Challenge of Analyzing Global Catastrophic Risks

Seth Baum
Global Catastrophic Risk Institute
http://sethbaum.com * http://gcrinstitute.org

Published in *Decision Analysis Today* 38(1): 20-24 (2019). This version 27 July 2020.

Global catastrophic risks are risks of the highest severity, regardless of their probability. Exact definitions vary, but they all point to catastrophes more severe than anything in recent human history. I tend to define global catastrophe as the collapse of modern global civilization. Some people in the field focus specifically on risks to human extinction, often using the term existential risk.

The concept of global catastrophe has a long history in theology, such as in the notion of apocalypse. The modern scientific study of global catastrophe can be traced to 1940s research on the threat from asteroids and a Manhattan Project study of the possibility of nuclear detonations igniting the atmosphere. Today, the field covers threats of natural origin including asteroids, comets, and volcanoes, as well as anthropogenic threats such as nuclear war and global warming. There is considerable interest in threats from future technologies, in particular biotechnology and artificial intelligence.

Much of the interest in global catastrophic risk comes from big-picture thinking about human civilization and its role in the world and the universe. Indeed, people drawn to global catastrophic risk often have backgrounds in philosophy and cosmology. The essential idea is that a global catastrophe could be a discontinuity in the course of human history. Noting the prospect for humans or our long-term descendants to expand into outer space, the stakes may be literally astronomical.

Similarly, the case made for the importance of global catastrophic risk typically centers on a moral concern for future generations. A global catastrophe could substantially diminish future generations, or eliminate them entirely in the event of human extinction. This extreme severity could make global catastrophic risk an unusually large class of risk, even if the probability is low. However, there has been little work comparing the long-term consequences of global catastrophes to the long-term consequences of smaller events. This is an important open area of research.

The Analytical Challenge

The prospect of comparing long-term consequences speaks to a wider theme: the global catastrophic risks are quite challenging to analyze. While long-term consequences are difficult to analyze for any risk, even the short-term consequences of global catastrophe are opaque. How resilient is modern global civilization to extreme catastrophes? If civilization collapses, what happens next? These questions do not have clear answers, but they are vital for evaluating the severity of global catastrophe.

The probability of global catastrophe is also hard to pin down. There is an inherent lack of data: modern global civilization has never previously been destroyed. Compounding the challenge is a phenomenon known as the observation selection effect: if a global catastrophe had previously occurred, we might not be alive to observe that fact. Rigorous estimates of the probability of global catastrophe must account for the lack of data and the observation selection effect, both of which make the analysis more difficult than it is for many other risks.

For some risk management decisions, it is not essential to rigorously quantify the global catastrophic risks or be able to compare them to other possible outcomes. To take one very simple example, it is probably good for us to turn the light off when we leave the room. Doing so could ever so slightly reduce the risk from global warming and have other benefits such as saving money on energy. The "cost" of turning the light off is usually fairly trivial—just the tiny effort it takes to flip the switch. One does not need a rigorous decision analysis accounting for long-term consequences and observation selection effects to reach the conclusion to flip the switch.

For other decisions, quantitative decision analysis is more important. For example, how much should society focus on global catastrophic risk relative to other issues? This question is essential for decisions on the allocation of scarce resources such as policymaker attention and philanthropic funding. Also, what should be done when some action poses a tradeoff between a global catastrophic risk and something else, or between multiple global catastrophic risks? These tradeoffs are quite common and worth illustrating with a few examples.

First, consider the decision of whether states should disarm nuclear weapons. This is currently a major point of contention for the international community. Disarmament advocates often emphasize the outsized severity of nuclear detonations, while disarmament opponents often emphasize the advantages of nuclear deterrence. Essentially, nuclear weapons could increase the severity of major war and decrease its probability. Nuclear disarmament could have the reverse effect. For nuclear disarmament, would the decrease in severity be large enough to offset any increase in probability?

Second, consider the decision to build nuclear power plants. Nuclear power can reduce the risk from global warming by shifting electricity production away from fossil fuel. It can also increase the risk of nuclear war by facilitating proliferation—this is seen especially clearly in the international concern over the nuclear program of Iran. Nuclear power can have other effects as well, including benefits for local air pollution, the risk of meltdown, and the displeasure of anti-nuclear citizens. Accounting for all these factors, under what circumstances would nuclear power be of net benefit to society?

Third, consider the decision to restrict the development of advanced artificial intelligence. Forward-thinking scholars of AI consider the possibility that AI systems may eventually be able to outsmart humans across a wide range of domains, similar to how they can now outsmart us in a few specific ones like chess and Go (and simpler domains like multiplication). There may even be some nonzero probability of runaway AI scenarios in which humans lose control, with potentially catastrophic consequences. However, restricting the development of AI could deprive society of many benefits that come from this technology, potentially including reductions in other global catastrophic risks. Therefore, in what ways, if at all, should the development of AI be restricted?

Each of these three decisions could be of profound consequence for society, yet none of them are easy to evaluate. They involve tradeoffs between multiple global catastrophic risks, or between global catastrophic risks and other important values. Evaluating these tradeoffs can require a quantitative analysis of the global catastrophic risks themselves and the relative importance of global catastrophic risk to other values. Such analysis is difficult, but it is important to pursue in order to inform these and other decisions.

Quantitative Analysis of Global Catastrophic Risk

There is a nascent effort to apply rigorous risk and decision analysis to quantitatively analyze the global catastrophic risks. Leading this effort are the Garrick Institute for the Risk Sciences at UCLA (https://www.risksciences.ucla.edu) and my own group, the Global Catastrophic Risk Institute (http://gcrinstitute.org). Our aim is help society understand the global catastrophic risks and make better decisions on them.

Central to this effort is the synthesis of relevant information. While there is no data on past global catastrophes, there is nonetheless a lot of information that can help evaluate the global catastrophic risks and the opportunities to reduce them. There are historical near-miss events that may have gone partway to global catastrophe, such as the Cuban missile crisis. There is information on the mechanisms underlying global catastrophes, such as the physics of asteroid-Earth collision. There are indicators of potential changes in risks, such as the decline in Russia-US relations since the recent Ukraine crisis. And there are the informed opinions of subject matter experts. All of these sources of information can be used to quantify global catastrophic risk, even in the absence of historical event data.

My group has been especially active in quantifying the risk of nuclear war. As with many risk analyses, we proceed by building models of how nuclear war could occur and what its consequences could be, by synthesizing available information of relevance to model parameters, and by crafting probability distributions to depict the uncertainty about model parameters.

Our model of the probability of nuclear war (available at https://ssrn.com/abstract=3137081) distinguishes between two primary types of nuclear war. First, there is “intentional” nuclear war, in which state leadership makes the decision to initiate nuclear war. This can come either from escalating a conventional (non-nuclear) war or from a non-war crisis, such as the Cuban missile crisis. Second, there is “inadvertent” nuclear war, in which state leadership mistakes some event as a nuclear attack and launches nuclear weapons in what it believes is retaliation but is in fact the first strike. The other event could be a non-war nuclear detonation, such as a nuclear terrorist attack, or a false alarm of nuclear detonation, such as a warning system malfunction.

To inform the quantification of the probability model parameters, we compiled a dataset of 60 historical near-miss incidents that may have threatened to turn into nuclear war. (These incidents are data points in the class of nuclear war near-misses, not the class of nuclear wars, though there is one nuclear war data point, World War II.) Post-WWII incidents span from the 1946 Azerbaijan crisis, in which US President Harry Truman allegedly threatened the Soviet Union with nuclear attack when it initially refused to leave Iranian Azerbaijan after WWII, to the 2018 Hawaii missile alert, in which the Hawaii Emergency Management Agency accidentally sent out a mass text message warning of an incoming ballistic missile.

A next step for this work is to assess these near-miss incidents in terms of how close they came to nuclear war. It is clear that some, such as the Cuban missile crisis, came closer than others, such as the Hawaii missile alert. For the Hawaii missile alert to have resulted in nuclear war, someone with nuclear launch authority would have needed to receive and act on the warning within the 38 minutes before a correction message was sent. There is no indication that anything along these lines came close to occurring.

Quantifying how close these incidents came to nuclear war is a greater challenge. The historical record is ambiguous, and scholars of nuclear war disagree on how close the incidents came. This disagreement resembles disagreements on the interpretation of other types of near-miss events documented by analysts such as Robin Dillon and Catherine Tinsley. Their work can be a valuable resource for the study of nuclear war near-misses and similar events of relevance to global catastrophic risk. More generally, the base of knowledge built up over the decades in risk and decision analysis has much to offer to the analysis of global catastrophic risk.

Global Catastrophic Risk Decisions

The difficulty of analyzing global catastrophic risk is only one challenge to effective decision-making about global catastrophic risk. The other major challenge is ensuring that careful analysis is used by the decision-makers.

In many cases, risk and decision analyses are commissioned by decision-makers who want the analysis to inform their decisions. In these cases, the findings of the analysis do not necessarily weigh heavily into the decision, but there is at least a clear path from analysis to decision. However, in global catastrophic risk, this is generally not the case. Risk management institutions are typically charged with focusing on risks that are local-scale and higher probability, while decision-makers on important aspects of global catastrophic risk (for example, political leadership of nuclear-armed states) tend not to approach their decisions in risk and decision analytic terms.

Meanwhile, analysis of global catastrophic risk is often motivated by an intrinsic concern about global catastrophic risk held by the analysts. That is certainly the case for myself: in reflecting upon my own moral values, I find the reduction of global catastrophic risk to be important, and my work proceeds accordingly. Though perhaps admirable, this situation lacks a built-in audience for the risk and decision analysis.

In response to this situation, my colleagues and I have developed two approaches. One is to promote the idea that global catastrophic risk reduction is important. While it is not our place to tell others what they should care about, we can at least explain why we have reached the conclusion that global catastrophic risk reduction is important, in particular the outsized severity of global catastrophe and the availability of opportunities to reduce the risk. In some cases, this "direct" approach is enough to generate interest.

The other approach seeks opportunities to reduce global catastrophic risk that are consistent with whatever other values the decision-makers hold. This approach is "indirect" because it does not directly emphasize the importance of global catastrophic risk. It is inspired by the concept of "mainstreaming" developed by the natural hazards community. To mainstream is to take a less prominent issue, such as natural hazards or global catastrophic risk, and embed it into a more prominent (i.e., mainstream) issue, such as economic development. Doing so makes it easier for decision-makers to factor those other issues in by not forcing them to restructure their agenda. In general, people do prefer to reduce their vulnerability to natural hazards and to reduce global catastrophic risk. Often, it's just a matter of finding opportunities that make sense from their perspective.

To that end, we find that global catastrophic risk decision analysis should include social science of decision-makers alongside technical analysis of the risks and decision options. The social science includes understanding specific people who make the decisions as well as the social and institutional contexts in which the decisions are made. Including the social science does add considerably to the complexity of what is already a challenging set of risks to analyze. However, in the interest of actually reducing the risks, we believe it is vital.