

A Conversation with Seth Baum on October 2, 2013

Participants:

- Seth Baum – Executive Director, Global Catastrophic Risk Institute
- Alexander Berger – Senior Research Analyst, GiveWell
- Sean Conley – Research Analyst, GiveWell

Summary

GiveWell spoke with Seth Baum about existential risks and global catastrophic risks (GCRs). Conversation topics included the likelihood of such risks, crosscutting research and action on catastrophes, and organizations working on crosscutting GCR issues.

***Note:** This set of notes was compiled by GiveWell and gives an overview of the major points made by Seth Baum.*

Existential risks and global catastrophic risks

There are two commonly used definitions of "existential risk." The first definition is a threat that could cause the extinction of all humans. The second definition is broader: an event that could greatly diminish the accomplishments of humanity and humanity's descendants. The second definition is the one used by Nick Bostrom of the Future of Humanity Institute at Oxford.

There is also a range of definitions for the term "global catastrophic risk" (GCR). Many people use the term in a similar way to the second definition of existential risk. Some definitions of GCR include events, such as World War II, that cause a lot of damage but do not have a significant long-term effect on humanity's development when considered from a very high level, or "astronomical" perspective.

The likelihood of catastrophe

Different types of catastrophes have different possible timelines. An asteroid strike could happen at any time (although the risk from an asteroid is one of the few risks that can be quantitatively characterized with confidence, and is known to be low), while catastrophic climate change won't happen in the very near future.

Estimates of total probability of catastrophe

Dr. Baum is not aware of credible estimates of the total probability of a GCR occurring within the next decade or two. There have been a number of texts with qualitative sweeps of potential risks without careful probabilistic analysis, such as those by Martin Rees, Nick Bostrom, and Richard Posner. Martin Rees predicted a 50% chance of humanity surviving the next century, but that estimate is sensitive to many poorly characterized details.

Creating probability estimates for GCRs is challenging. Dr. Baum, Antony Barrett, and Kelly Hostetler created a detailed probability assessment of the risk of nuclear war between the

US and Russia. The estimates for war occurring in a given year range over several orders of magnitude, which demonstrates that even a detailed study of GCR probability is limited. There are many poorly known parameters, such as how likely people are to respond to false alarms.

GCR probabilities aren't fixed and can be affected by people's actions. This is a good thing, but is another factor that makes estimation difficult. The estimates may not be independent of the phenomena being estimated, because if estimates are released for some catastrophe, people may take actions that affect the probability of it occurring.

Overall risk over time

Dr. Baum believes that on balance the risk from GCRs will increase over time as technology develops, but holds that belief with low confidence. There are many unknown details, and it is possible that the risks will increase or decrease. Biomedical research might lead to deadly viruses and/or it might produce incredibly effective vaccines or bio-surveillance systems that very effectively prevent the spread of disease. Looking specifically at emerging technologies, the risks seem to be growing on balance.

Another complicating factor when predicting the direction of risk is that smaller catastrophes could draw attention to an issue and help prevent future events of that type.

Nuclear risk

Nuclear risk appears to be falling over time, as the world is moving towards fewer nuclear weapons. More countries are acquiring them, but most of the weapons are still owned by the US and Russia, which have been reducing them over time. The trend seems likely to continue, and the arguments for outright disarmament may strengthen while the arguments for keeping nuclear weapons may weaken.

Environmental risk

Environmental risk is likely to get worse over time. Resource scarcity could be increasingly problematic, although technological advancements might mitigate that. Climate change risk is likely to rise. A recent Intergovernmental Panel on Climate Change report discussing the science of geoengineering is not optimistic: early research indicates that it won't be possible to remove enough carbon dioxide from the atmosphere, and solar radiation management may be able to lower temperatures but is not a perfect offset for greenhouse gases and likely has other impacts we don't yet understand.

Crosscutting GCR research

Crosscutting GCR research deals with strategic questions about work on various risks, such as which areas to focus resources on. There is currently little work on these types of questions. It's not yet clear whether more research on crosscutting questions will be worthwhile, but it is worth trying it to find out.

The effective altruist community has been influenced by GCR research.

The Future of Humanity Institute and Nick Bostrom have been influential by making their work accessible online, responding to media requests, and getting involved in policy.

Catastrophe prevention and recovery

There are a number of precautions that humanity could take now or in the future to protect against a wide range of GCRs. Such precautions include:

- Increasing grain stores
- Building bunkers
- Space colonization
- Encouraging political transparency
- Promoting altruism and global humanitarianism (including not privileging the well-being of people of one's own country over others)
- Promoting goodwill between nations
- Energy conservation (both to reduce climate change and competition among nations)
- Increasing local self-sufficiency

The general consensus is that grain stores or bunkers on Earth are likely to be more cost-effective than increasing the number of people in space, at least over time scales of decades and possibly up to time scales of hundreds of millions of years. Beyond that, space colonization will become necessary.

Preparations will likely be more politically feasible if they can be justified as measures for more common, smaller-scale disasters such as hurricanes and droughts.

Influencing governments

Often to influence government, you need to influence the policymakers: the president and Congress. There may be some scope for influencing particular departments.

Defense Advanced Research Projects Agency (DARPA) plays a very large role in artificial intelligence research, so persuading it to build safety considerations into its robotics programs would likely have a large effect on AI safety overall. There is a reasonable chance of accomplishing this; the Department of Defense is particularly concerned with security and preventing bad accidents, and regularly deals with programs that could cause bad accidents, so it might be interested in preventing artificial intelligence (AI) risk. The caveat is that AI risk is speculative, and the Department may prefer to pursue more tangible goals.

There are also benefits of having "champions" in a department. One of Dr. Baum's colleagues championed GCRs in the Department of Homeland Security for a period before retiring. Champions in companies can be effective as well. For example, dedicated individuals focused on sustainability within a corporation can have an influence on the company's policies over time.

Other organizations

Some organizations explicitly working on crosscutting GCRs issues include:

- The Future of Humanity Institute.
- The Cambridge Centre for the Study of Existential Risk. The founding members of the organization have been using their influence and the Cambridge brand to pull together important people and generate attention for their project.
- The Global Catastrophic Risk Institute.
- The Lifeboat Foundation.
- The Skoll Global Threats Fund.
- The Global Challenges Foundation, a recently founded organization.

GCRI maintains a directory of organizations that work on aspects of GCR issues:

<http://gcrinstitute.org/organization-directory/>

Other people for GiveWell to talk to

- Huw Price – lead faculty member for the Cambridge Centre for the Study of Existential Risk.
- Seán Ó hÉigeartaigh – Academic Project Manager at the Future of Humanity Institute and also involved at the Cambridge Centre for the Study of Existential Risk. He would be a good person to talk to about the funding of both organizations.

All GiveWell conversations are available at <http://www.givewell.org/conversations>